Distribution Agreement

Varying-coefficient Regression Analyses for Semi-competing Risks Data

By
Ruosha Li
Doctor of Philosophy

Biostatistics

Limin Peng, Ph.D.
Advisor

Amita K. Manatunga, Ph.D.
Committee Member

Eugene Huang, Ph.D.
Committee Member

Michael G. Benatar, M.D., Ph.D.
Committee Member

Accepted:

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

Date

Varying-coefficient Regression Analyses for Semi-competing Risks Data

By

Ruosha Li

B.S., Peking University, 2006

M.S., Emory University, 2010

Advisor: Limin Peng, 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

2011

Abstract

Biomedical studies for chronic diseases often involve multiple event times. In this dissertation, we focus on a scenario where one terminating event can dependently censor a nonterminating event, but not vice versa. Such data structure was termed as semi-competing risks data by Fine et al. (2001). We concern regression analyses of semi-competing risks data under the modeling frameworks that accommodate varying covariate effects, including quantile regression (Koenker and Bassett, 1978) and temporal regression (Fine et al., 2004). The two modeling frameworks are gaining increased popularity in survival analysis for their flexibilities and ease of interpretations.

In the first project, we propose quantile regression methods for left-truncated semi-competing risks data. The project is motivated by the Denmark diabetes registry study, where the nonterminating event time of interest, time to diabetic nephropathy (DN), is subject to the dependent censoring by time to death. Biological interests are centered on regression analysis of time to DN, without removing the effect of death. A notable complication in this dataset is the administrative left truncation to death, which greatly complicates the analysis. We propose inference procedures for the conditional quantiles of the cumulative incidence function of DN, by appropriately handling left-truncation via the technique of inverse probability of censoring weighting. We show that the proposed estimator has nice asymptotic properties including uniform consistency and weak convergence. We illustrate the practical utility of the proposed method via simulation studies and an application to the Denmark diabetes registry data.

In the second project, we study quantile regression on the marginal distribution of the nonterminating event. The project is motivated by the AIDS Clinical Trial Group (ACTG) 364 study, where a study endpoint, time to first virologic failure, is subject to censoring by patients dropouts. We develop a quantile regression method which focuses on the marginal conditional quantiles of the study endpoint, while providing information on the association between the study endpoint and the patient dropout. The proposed estimating equations well utilize the special semi-competing risks data structure, and can be solved by an efficient iterative algorithm. We derive the asymptotic properties of the resulting estimator, including uniform consistency and weak convergence. Simulation studies demonstrate the proposed method performs well with moderate sample size. We applied the proposed methods to the ACTG 364 study for analyses of the virologic endpoint.

In the third project, we study the same data structure as that in the first project from a different perspective. Specifically, we develop temporal regression methods for the cumulative incidence function of DN in the Denmark diabetes registry study, to evaluate the temporal relationship between covariates and DN progression. We propose estimation and inference procedures for the time-varying regression coefficients. Furthermore, some preliminary simulation results show that the proposed methods perform well with realistic sample sizes.

Varying-coefficient Regression Analyses for Semi-competing Risks Data

By

Ruosha Li
B.S., Peking University, 2006
M.S., Emory University, 2010

Advisor: Limin Peng, Ph.D.

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
2011

Acknowledgement

First and foremost, I would like to express my sincere gratitude to my advisor, Dr. Limin Peng, for her guidance, inspiration and encouragement over the past three years. Her expertise and timely help have got me through many hard times. She has not only provided me with well-rounded research experience, but has also taught me how to become a determined and detailed researcher. I feel very fortunate to have her as my mentor and advisor.

I would like to thank my committee members, Dr. Amita Manatunga, Dr. Eugene Huang and Dr. Michael Benatar, for their time and effort in reviewing my dissertation. Their suggestions and comments led to substantial improvement in this work.

I am indebted to Dr. John Hanfelt, Dr. Amita Manatunga, and Dr. Ying Guo for providing me with precious research opportunities. Working with them has been especially exciting and rewarding. I would like to thank Dr. Robert Lyles and Dr. Michael Benatar for the great collaboration experience, and Kirk Easley for his guidance on statistical consulting. I would like to extend my thanks to the department staff, who have helped me take care of many things. I am also grateful to my dear friends at Emory for making the past few years so enjoyable and memorable.

Finally, I would like to thank my parents for their faith in me and their unwavering support, without which this work would not have been possible.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background

Biomedical studies for chronic disease often involve multiple event times. An event is classified as terminating if its occurrence prevents subsequent observations of other events, and non-terminating otherwise. In this dissertation, we focus on a bivariate structure where a terminating event can censor a non-terminating event, but not vice versa. This data structure differs from the traditional competing risks data where the events are mutually exclusive, and was termed as semi-competing risks data by Fine et al. (2001). Here and in the sequel, we use T_1 and T_2 to denote time to the non-terminating event and time to the terminating event, respectively.

Semi-competing risks problem are frequently encountered in clinical studies in the following two scenarios. The first scenario involves time to morbidity and time to mortality (Peng et al., 2008). An example is the Denmark diabetes registry study (Andersen et al., 1993) where time to diabetic nephropathy is subject to dependent censoring by time to death. On the other hand, death is still observable if diabetic nephropathy occurred first. The second scenario involves time to a study endpoint and time to patients' dropout. This is the case of AIDS Clinical Trial Group (ACTG) 364 study, where the primary endpoint is the first virologic failure. However, many patients withdrew before the end of the study for disease related reasons, which prevented subsequent follow-up and observation of the virologic endpoint. In the sequel, we refer to these two scenarios as the mortality-morbidity scenario and the endpoint-dropout scenario respectively.

Statistical challenges for semi-competing risks problems lie in the analysis of T_1, which must account for the dependent censoring by the terminating event, while the analysis of T_2 can typically be performed using traditional techniques concerning independently censored univariate survival data. To analyze T_1, one may choose from two types of approaches according to the nature of the problem (Peng et al., 2008). If the problem falls in the paradigm of time to morbidity and time to mortality, it is preferable to focus on the *crude quantities*, which do not hypothesize the removal of death for inference on morbidity. Commonly used *crude quantities* include the cause-specific hazard $\lambda_1(t) = \lim_{h\to 0}\Pr(t \le T_1 < t+h, T_1 < T_2 | T_1 \ge t, T_2 \ge t)/h$ and the cumulative incidence function $F_1(t) = \Pr(T_1 \le t, T_1 < T_2)$. When T_1 and T_2 correspond to time to a study endpoint and time to dropout respectively, it is more sensible to treat T_2 as dependent censoring time and use the *net quantities* for statistical inferences and scientific implications. The *net quantities*, like the marginal distribution function $F_{T_1}(t) = \Pr(T_1 \le t)$, hypothesize a setting where patients' dropout does not exist. They are more relevant to the underlying biology of interest, while appropriately accounting for complications that occurred to the observational process.

In this dissertation, we study three research problems on regression analyses of semi-competing risks data. We focus on regression models that accommodate nonconstant covariate effects (varying-coefficient models), as the constant effect assumption posed by most existing literatures may be violated in many practical situations. For example, our analyses of the Denmark diabetes registry data in Section 2.2.2 suggest a varying pattern in the effect of diabetic onset age on time to diabetic nephropathy. Neglection of such varying effect patterns may cause biased inferences on the quantities of interest. Moreover, varying effect patterns themselves are often scientifically meaningful and clinically important.

In the first two projects, we concern quantile regression for modeling of the semi-competing risks data. Quantile regression is a regression technique introduced by Koenker and Bassett (1978), and has emerged as an important alternative to accelerated failure time (AFT) model and Cox proportional hazards model in survival analysis. Such a modeling strategy generally has advantages in model interpretation and flexibility. However, there has been limited research on quantile regression with semi-competing risks data. In the first project, we concern quantile regression on the *crude quantities* of T_1 under the mortality-morbidity scenario. Specifically, we focus on a situation with an additional complication of left truncation to T_2, as in the Denmark diabetes registry study. In the second project, we concern quantile regression modeling of T_1 under the endpoint-dropout scenario, where the *net quantities* are of primary interest.

In the third project, we concern temporal regression for modeling the *crude quantities* of T_1. We focus on the same data structure as that in the first project, which involves left truncation to the terminating event. By accommodating time-varying regression coefficients, temporal regression can offer important insights into the temporal relationship between the response and covariates (Fine et al., 2004).

In the rest of this Chapter, we first present two motivating examples, the Denmark diabetes registry study and the ACTG 364 study. Reviews of relevant existing literature are provided after each example. Following that, we give brief introductions of quantile regression and temporal regression. Finally, we present an outline of this dissertation.

1.2 First Motivating Example

1.2.1 Denmark Diabetes Registry Study

The Denmark diabetes registry study (Andersen et al., 1993) is a prospective cohort study on insulin-dependent diabetes patients referred to the Steno Memorial Hospital, a diabetes specialist hospital in Denmark. From 1933 to 1981, the study enrolled about 2700 patients diagnosed with insulin-dependent diabetes mellitus prior to age 31. One landmark event is diabetic nephropathy (DN), an indicator of kidney failure and a prime indication for dialysis. It is of scientific interest to characterize the of progression of DN after the diagnosis of insulin-dependent diabetes.

At entry, patients' age, age of diabetes diagnosis, gender and the presence of DN were recorded. The patients were followed to death, emigration, or December 31, 1984. The observation on time to DN could be terminated by death, which remained observable when DN occurred first. Letting the time origin be the diagnosis of insulin-dependent diabetes, time to DN and time to death form a bivariate semi-competing risks structure.

A notable complication in this study is the administrative left truncation on death. That is, patients who died before study enrollment were excluded from study. As a result, data were available only if time to death is larger than time to study entry. Note that time to diabetic nephropathy was not subject to left truncation. Some patients had developed DN before enrolled into the study.

1.2.2 Literature Review

In this example, T_1 and T_2 correspond to time to DN and time to death, respectively. Let L be time to study entry, data were observable only when $T_2 > L$. The major challenge for analyzing semi-competing risks data subject to left truncation, as with the untruncated case, lies in the inference on the nonterminating event time T_1. As discussed in Peng et al. (2008), when T_1 and T_2 correspond to time to morbidity and time to mortality, as in the Denmark diabetes registry study, interpretations of net quantities may be controversial because they unrealistically hypothesize a setting where death can be completely eliminated before morbidity. Moreover, crude quantities can be estimated nonparametrically (Prentice et al., 1978) in contrast to net quantities that are cursed by non-identifiability (Tsiatis, 1975). By these considerations, crude quantities are more preferable than net quantities in the analyses of the Denmark diabetes registry and similar studies. The principles discussed above are generally applicable to competing risks scenario and multi-state analysis.

In the one-sample case, how to estimate crude quantities in presence of left truncation has been well studied. For example, one may only use the information on

$T = T_1 \wedge T_2$ and employ the approach by Andersen et al. (1993) or Huang and Wang (1995) for competing risks data to estimate the cumulative cause-specific hazard $\Lambda_1(t) = \int_0^t \lambda_1(s)ds$ and $F_1(t)$. Estimation tailored to semi-competing risks setting was developed by Peng and Fine (2006a), and their estimator avoids information loss due to artificial truncation, which is, removal of data with $T_2 > L$ but $T \leq L$.

It is remarkable that left truncation poses considerable complexity to regression analysis of semi-competing risks data. This may be due to the fact that covariates can have some indirect effects on T_1 through the truncation scheme defined as $T_2 > L$. In the absence of left truncation, semi-competing risks regression based on crude quantities may follow the existing approaches for competing risks data, such as Fine and Gray (1999), Klein and Andersen (2005), among others. Recently, some research efforts have been made to extend these methods to left-truncated competing risks data (Geskus, 2010; Shen, 2011; Zhang et al., 2011). However, adapting these methods to the left-truncated semi-competing risks data considered here would require artificial truncation and thus incur undesirable efficiency loss.

1.3 Second Motivating Example

1.3.1 AIDS Clinical Trial Group 364 study

AIDS Clinical Trial Group (ACTG) 364 study is a randomized, multicenter clinical trial in AIDS patients with plasma human immunodeficiency virus HIV RNA above 500 copies/ml. In this study, 195 subjects were enrolled and randomly assigned to one of the three treatment arms, which are protease inhibitor nelfinavir (NFV), the non-nucleoside efavirenz (EFV), and NFV+EFV (Albrecht, Bosch, Hammer, Liou, Kessler, Para, Eron, Valdez, Dehlinger, and Katzenstein, Albrecht et al.). One study endpoint is the time from study entry to time of the first virologic failure (confirmed HIV RNA above 200 copies/ml). Baseline characteristics include baseline HIV RNA, and whether 3TC is a new nucleoside reverse transcriptase inhibitor (NRTI) for the patient. It is of scientific interest to compare the progression of virologic failure among the three treatment arms, adjusting for baseline covariates.

During the 2 years follow up period, 55 patients dropped out from the assigned study for disease-related reasons, such as toxicity, complications, excessively high viral load, etc. There was also administrative censoring because of limited period of follow-up. At the end of the study, 101 patients were observed to experience virologic failure. Among the 94 patients whose failure time was not observed, 83 were due to administrative censoring and 11 were due to dropout. Although the dropouts rate prior to virologic failure seems low, 9 out of 11 were in the NFV+EFV arm. This suggests that that the dropouts may be correlated to the treatment arms and thus may be informative for the virologic endpoint.

1.3.2 Literature Review

In this example, T_1 and T_2 correspond to time to first virologic failure and time to dropout, respectively. Denote the administrative censoring time as C. One can naively treat $T_2 \wedge C$ as an non-informative censoring time on T_1, and adopt one of the traditional techniques like the AFT model or the Cox proportional hazards model. If the dropout is informative for the virologic endpoint, then these analyses correspond to the cause-specific hazard function. It is more appealing to focus on *net quantities*, which correspond to a situation where dropouts do not exist, for example, the marginal distribution of time to first virologic failure $F_{T_1}(t) = \Pr(T_1 \leq t)$. Moreover, it is worthwhile to examine the association between T_1 and T_2; that is, how the study endpoint and dropouts are correlated with each other.

When T_2 and T_1 are dependent, the marginal distribution $F_{T_1}(t)$ and the dependence structure are not identifiable under a pure nonparametric setting (Tsiatis, 1975). As a result, additional assumptions are often imposed on the joint distribution of T_1 and T_2 to solve the identifiability issue. One may apply some existing methods for competing risks data (Emoto and Matthews, 1990; Heckman and Honoré, 1989; Link, 1989), after reducing the data to competing risks format by keeping only the time and type of the first event $T = T_1 \wedge T_2$. However, it is not practically desirable to discard the extra information on T_2 in the semi-competing risks setting.

Some methods tailored to semi-competing risks data have been developed under the one-sample setting. For example, Fine et al. (2001) formulated the dependence structure between T_1 and T_2 by positing the Clayton copula (Clayton, 1978) on the observable region, and derived a closed form estimator for $F_{T_1}(t)$. Subsequent work includes Wang (2003), which studied the degree of dependence under a general copula class, and Jiang et al. (2005), which proposed a self-consistent estimator for $F_{T_1}(t)$ under the Clayton model.

In the regression setting, Lin et al. (1996) modeled the marginal covariates effect on both T_1 and T_2 via a bivariate accelerated failure time (AFT) model. They accounted for dependent censoring by employing a novel artificial censoring technique. To avoid excessive artificial censoring which can lead to substantial efficiency loss, Peng and Fine (2006b) developed a new artificial censoring scheme which achieved reduction in artificial censoring rate. Peng and Fine (2007) studied a general class of functional regression models for the marginal survival function of T_1, which includes the Cox proportional hazards model as a special case. Recently, Hsieh et al. (2008) generalized Wang (2003)'s method to semi-competing risks data with covariates. Their approach allowed association parameter to vary in different subgroups, but requires that covariates only take discrete values. To date, very limited work has been done for quantile regression of T_1 subject to dependent censoring, especially in the semi-competing risks set-up.

1.4 Quantile Regression

In this section we briefly introduce the general framework of quantile regression, a regression technique first introduced by Koenker and Bassett (1978). While many traditional regression methods focus on the conditional mean of the response variable given covariates, quantile regression methods offer a mechanism for estimating a range of conditional quantiles to achieve a more comprehensive and robust analysis (Koenker, 2005).

Given a $(p+1) \times 1$ covariate vector $Z = (1, \tilde{Z})^T$, the conditional quantiles of a random variable T is defined as $Q_T(\tau|Z) = \inf\{t : \Pr(T \leq t|Z) \geq \tau\}$. A quantile regression model may assume that

$$Q_T(\tau|Z) = g\{Z^T\beta_0(\tau)\}, \quad 0 < \tau < 1, \tag{1.1}$$

where $g\{\cdot\}$ is a monotone link function, for example, log link or identical link, $\beta_0(\tau)$ is a $(p+1) \times 1$ vector of unknown regression coefficients. The first component of $\beta_0(\tau)$ is the transformed baseline quantile function $g^{-1}\{Q_1(\tau|Z=0)\}$, while the rest components represent the effects of corresponding covariates on $g^{-1}\{Q_1(\tau|Z)\}$ that may change with τ.

It is worth pointing out that quantile regression model (1.1) with $g\{\cdot\} = \log\{\cdot\}$ and $\beta(\tau) = (Q_\varepsilon(\tau), b^T)^T$ reduces to the accelerated failure time (AFT) model

$$\log T = \tilde{Z}^T b + \varepsilon,$$

where b is an unknown vector of regression coefficient and ε is the i.i.d. error term. By allowing $\beta_0(\tau)$ to vary across τ, model (1.1) can accommodate more realistic effect patterns of covariates than location-shifts imposed by many traditional models.

In survival analysis, quantile regression is emerging as an important alternative to

Cox proportional hazards model and AFT model. Research efforts have been devoted to quantile regression for independently censored survival data. Early work by Powell (1984, 1986) adapted the idea of least absolute deviation (LAD), under the assumption that censoring time is always observed. However, this assumption might be too strong, as censoring time is not always observed in most of the survival settings. Ying et al. (1995) proposed a nonparametric procedure for median regression, which requires unconditional independence between censoring time and survival time. Under the usual conditional independence assumption, Yang (1999) considered a median regression approach based on weighted empirical survival and hazard function. However, this approach requires the error structure to be nearly i.i.d. and cannot accommodate heteroscedastic error structures. Without imposing the unconditional independence assumption or stringent constraints on error distributions, Portnoy (2003) employed the principle of self-consistency (Efron, 1967) and developed a grid-based recursively reweighted estimation procedure. The resulting estimate reduces to the Kaplan-Meier estimator in the one sample case, and its asymptotic properties were recently established by Portnoy and Lin (2010). Peng and Huang (2008) developed an alternative grid-based approach based on the martingale structure, their method is well-defined and facilitates neat asymptotic properties. In the one sample setting the resulting estimator reduces to the Nelson-Aalen estimator. Most recently, Huang (2010) developed a novel grid-free estimation procedure based on quantile calculus. The resulting estimator is asymptotically equivalent to that of Peng and Huang (2008), when the grid size of the latter is of order $o(n^{-1/2})$.

While quantile regression has been well studied for independently censored survival data, there has been limited work on quantile regression for survival settings involving dependent censoring. One relevant work is Peng and Fine (2009), which studied the conditional quantiles of the cumulative incidence function. Their method can be applied to right-censored only competing risks and semi-competing risks data, but is not applicable in presence of left-truncation. On the other hand, no one has studied quantile regression for the *net quantities* of T_1, adjusting for dependent censoring by T_2.

1.5 Temporal Regression

To study the temporal relationship between covariates and response with event time data, many existing literatures have considered hazard-based regression models. For example, research efforts have been devoted to the extension of Cox proportional hazards model by formulating time-varying effects on the hazard function (Zucker and Karr, 1990; Murphy and Sen, 1991; Valsecchi et al., 1996; Martinussen et al., 2002; Martinussen and Scheike, 2002; Winnett and Sasieni, 2003; Cai and Sun, 2003; Tian et al., 2005). These methods generally require smoothing in the estimation step. Alternatively, a distinct type of modeling strategies focuses on the mean of a temporal process associated with the event times. With survival data, for example, one can formulate the temporal covariate effects on the distribution function or survival function of the event time. This type of temporal regression strategies was first proposed by Fine et al. (2004) for multistate event time data under noninformative missing mechanisms. As compared to the hazard-based regression models, the mean-based temporal regression models may be more appealing in several aspects. First, estimation with these models does not require smoothing, thereby enabling easier computation and higher efficiency. Furthermore, the mean-based models often lead to straightforward interpretation and prediction on the temporal process of interest. In this section, we give a brief introduction of temporal regression with survival data along this direction.

With univariate survival data, Peng and Huang (2007) formulated a temporal regression model as

$$S_T(t|Z) = \exp[-\exp\{Z^T\beta_0(t)\}], \tag{1.2}$$

where $S_T(t|Z) = \Pr(T > t|Z)$ is the survival function of a univariate event time T, $Z = (1, \tilde{Z})^T$ is the $(p+1)\times 1$ covariate vector, and $\beta_0(t) = \{\beta_0^{(0)}(t), \beta_0^{(1)}(t), ..., \beta_0^{(p)}(t)\}^T$ is a vector of the unknown time-varying regression coefficient. The first component $\beta_0^{(0)}(t)$ corresponds to the baseline log-transformed cumulative hazard, and the remaining p components correspond to the temporal covariate effects. The model can be viewed as an extension of the Cox proportional hazards model, which corresponds to the situation when $\beta_0(t)$ is constant except in the first component. Peng and Huang (2007) developed smoothing free estimation and inference procedures, by constructing a forgoing estimating equation based on the martingale structure.

For analyses of the *net quantities* of T_1 in semi-competing risks settings, Peng and Fine (2007) studied a similar model for the marginal survival function of T_1, specified as

$$S_{T_1}(t|Z) = g\{Z^T\beta_0(t)\},$$

where Z and $\beta_0(t)$ are $(p + 1) \times 1$ vectors of covariates and unknown regression coefficients respectively, and $g(\cdot)$ is a known monotone link function. The model incorporates a class of functional generalized linear models for the survival function. To account for the dependent censoring from T_2, the authors assumed a time-dependent copula for the joint distribution of T_1 and T_2.

In semi-competing risks settings where scientific interest centers on the *crude quantities* of T_1, one may consider a semiparametric temporal regression model on the cumulative incidence $F_1(t|Z) = \Pr(T_1 \le t, T_1 < T_2|Z)$, which assumes that

$$F_1(t|Z) = g\{Z^T\beta_0(t)\}, \tag{1.3}$$

with $g(\cdot)$ being a prespecified link function. As with the univariate case, model (1.3) can be viewed as an extension of the proportional subdistribution hazard model by Fine and Gray (1999) when $g(x) = 1 - \exp\{-\exp(x)\}$. The model was first studied by Scheike et al. (2008) for competing risks data. Scheike et al. (2008) proposed estimation and inference procedures for $\beta_0(t)$, using the technique of inverse probability of censoring weighting (IPCW). However, their approach cannot handle the situations with left truncation to T_2, a common complication in observational studies.

1.6 Outline

In Chapter 2 we propose quantile regression methods based on the conditional quantiles of the cumulative incidence function. We constructed a set of unbiased estimating equations, which makes good use of the data structure and avoids artificial truncation. The proposed methods provide meaningful interpretations as well as the flexibility to accommodate varying covariate effects, and can be easily implemented based on existing statistical software. Asymptotic properties of the resulting estimators are established including uniform consistency and weak convergence. We also provided the analytical form of the influence functions, which facilitate a standard error estimation procedure that does not require resampling. Monte Carlo simulations showed the proposed method have satisfactory finite-sample performance. Finally, an application to Denmark diabetes registry data provides an illustration of our proposals.

In Chapter 3 we develop quantile regression methods for T_1 when it is subject to dependent censoring by T_2. Our methods provide simultaneous inference on the marginal conditional quantiles and the dependence structure between T_1 and T_2. We formulated the dependence structure via a Copula model, and constructed two sets of unbiased estimating equations which fully utilize the observed information. We develop an efficient iterative algorithm to solve the proposed estimating equations. The resulting estimators can be shown to be uniformly consistent and converge weakly to Gaussian. The finite-sample performance of the proposed estimators are evaluated by Monte Carlo simulations. Furthermore, we illustrate the practical utility of the proposed methods via an application to the ACTG 364 study.

In Chapter 4 we study temporal regression methods based on the cumulative incidence of T_1 when there involves left-truncation to the terminating event. We construct estimating equations that offer valid estimates of the temporal coefficients by properly handling left-truncation. We also derive consistent estimators for the asymptotic covariance matrix. Some preliminary simulation studies suggest proper finite sample performances of the proposed estimators.

In Chapter 5 we provide a summary and discuss future work for the dissertation. We present possible extensions of the proposed methods, and directions for future research.

Chapter 2

Quantile Regression for Left-truncated Semi-competing Risks Data

2.1 Regression Procedures

2.1.1 Data and Model

We begin with a formal introduction of data and notation. Let C denote an independent censoring time for both T_1 and T_2, say administrative censoring. Define $X = T_1 \wedge T_2 \wedge C$, $Y = T_2 \wedge C$, $\delta = I(T_1 < Y)$, and $\eta = I(T_2 < C)$, where $\wedge$ denotes the minimum operator and $I(\cdot)$ is the indicator function. Let $\widetilde{\boldsymbol{Z}}$ be $p \times 1$ vector of recorded covariates and $\boldsymbol{Z} = (1, \widetilde{\boldsymbol{Z}}^T)^T$. Observed data include n independent and identically distributed replicates of $(X^*, Y^*, \delta^*, \eta^*, L^*, \boldsymbol{Z}^*)$, denoted by $(X_i^*, Y_i^*, \delta_i^*, \eta_i^*, L_i^*, \boldsymbol{Z}_i^*)_{i=1}^n$, where $(X^*, Y^*, \delta^*, \eta^*, L^*, \boldsymbol{Z}^*)$ follows the conditional distribution of $(X, Y, \delta, \eta, L, \boldsymbol{Z})$ conditional on $L < Y$.

Mimicking scenarios in prevalence studies, where censoring induced by follow-up mechanism is of primary concern and can only occur after sampling time, we restrict the truncation time L be always less than the censoring time C. Such an assumption has been imposed in much previous work on truncated and censored data, for example, Wang (1991) and Asgharian et al. (2002). In addition, we assume that (L, C) is independent of (T_1, T_2) and $\boldsymbol{Z}$. Extensions to cases where (L, C) are dependent on $\boldsymbol{Z}$ are briefly discussed in Section 2.3.

We consider a quantile regression model based on the cumulative incidence conditional quantile function defined as $Q_1(\tau|\boldsymbol{Z}) = \inf\{t : F_1(t|\boldsymbol{Z}) \geq \tau\}$, where $F_1(t|\boldsymbol{Z}) = \Pr(T_1 \leq t, T_1 < T_2|\boldsymbol{Z})$. The concept of cumulative incidence conditional quantile, like its precursor in the one-sample setting introduced by Peng and Fine (2006a), has a straightforward interpretation as the first time given covariate $\boldsymbol{Z}$ by which the probability of the nonterminating event having occurred in presence of the terminating event exceeds τ. Based on this crude quantity, we consider a model which assumes that

$$Q_1(\tau|\boldsymbol{Z}) = g\{\boldsymbol{Z}^T \boldsymbol{\beta}_0(\tau)\}, \quad \tau \in [\tau_L, \tau_U], \tag{2.1}$$

where $g(\cdot)$ is a known monotone continuous link function, for example, log link or identical link, $\boldsymbol{\beta}_0(\tau)$ is a $(p+1)\times 1$ vector of unknown regression coefficients, and $0 < \tau_L \le \tau_U \le 1$. Note that when $\tau_L = \tau_U$, model (2.1) only holds for a single τ, like the median regression. The first component of $\beta_0(\tau)$ is the transformed baseline quantile function $g^{-1}\{Q_1(\tau|\boldsymbol{Z}=0)\}$, while the rest components represent effects of corresponding covariates on $g^{-1}\{Q_1(\tau|\boldsymbol{Z})\}$ that may change with τ. This model was recently studied by Peng and Fine (2009) for competing risks data in the absence of left truncation, a data feature that can greatly complicate the regression analysis as explained earlier.

2.1.2 Estimation of $\boldsymbol{\beta}_0(\tau)$

To estimate $\boldsymbol{\beta}_0(\tau)$ in model (2.1), our basic idea resembles that of inverse probability of censoring weighting (IPCW) (Robins and Rotnitzky, 1992) and is to weigh the observed data in an appropriate way such that the bias induced by truncation and censoring is corrected in the estimation of $\boldsymbol{\beta}_0(\tau)$. The key is about constructing weights for truncated and possibly also censored observations. Let $\alpha(\boldsymbol{z}) = P(L < T_2|\boldsymbol{Z}=\boldsymbol{z})$ represent the truncation probability given $\boldsymbol{Z}=\boldsymbol{z}$, and $G(y) = P(L < y \le C)$. An inverse weight is suggested by the following equality:

$$E\left\{\frac{I(X_i^* \le t, \delta_i^* = 1, \eta_i^* = 1)}{W(Y_i^*, \boldsymbol{Z}_i^*)}\middle|\boldsymbol{Z}_i^*\right\} = F_1(t|\boldsymbol{Z}_i^*), \tag{2.2}$$

where $W(y,\boldsymbol{z}) = G(y)/\alpha(\boldsymbol{z})$. To show (2.2), without loss of generality, we temporarily omit $\boldsymbol{Z}_i^*$s. Let $\Omega(\boldsymbol{\omega};t) = \{(t_1,t_2,l,c):\ t_1 \le t, t_1 < t_2, l < t_2 \le c\}$, where $\boldsymbol{\omega}$ denotes the vector, (t_1,t_2,l,c), and let $\mathrm{P}_{T_1,T_2,L,C}(t_1,t_2,l,c)$ be the joint distribution function of (T_1,T_2,L,C). We first note that $\mathrm{dP}_{T_1^*,T_2^*,L^*,C^*}(t_1,t_2,l,c) = \alpha^{-1}\mathrm{dP}_{T_1,T_2,L,C}(t_1,t_2,l,c)$

in the interior of $\Omega(\boldsymbol{\omega};t)$. Therefore,

$$
\begin{aligned}
& E\left\{\frac{I(X^* \le t, \delta^*=1, \eta^*=1)}{G(Y^*)/\alpha}\right\} = E\left\{\frac{I(T_1^* \le t, T_1^* < T_2^*, L^* < T_2^* \le C^*)}{G(T_2^*)/\alpha}\right\} \\
= & \int_{\Omega(\boldsymbol{\omega};t)} \frac{\alpha}{G(t_2)} \mathrm{dP}_{T_1^*,T_2^*,L^*,C^*}(t_1,t_1,l,c) = \int_{\Omega(\boldsymbol{\omega};t)} \frac{1}{G(t_2)} \mathrm{dP}_{T_1,T_2,L,C}(t_1,t_2,l,c) \\
= & E\left\{\frac{I(T_1 \le t, T_1 < T_2, L < T_2 \le C)}{G(T_2)}\right\} = E\left[E\left\{\frac{I(T_1 \le t, T_1 < T_2, L < T_2 \le C)}{G(T_2)}\middle| T_1, T_2\right\}\right] \\
= & E\left\{I(T_1 \le t, T_1 < T_2) \times \frac{G(T_2)}{G(T_2)}\right\} = EI(T_1 \le t, T_1 < T_2) = F_1(t).
\end{aligned}
$$

In light of (2.2), we propose to estimate $\boldsymbol{\beta}_0(\tau)$ by $\hat{\boldsymbol{\beta}}(\tau)$, which is the solution to the following estimating equation:

$$
\boldsymbol{S}_n(\boldsymbol{b},\tau) \equiv n^{-1/2} \sum_{i=1}^{n} \boldsymbol{Z}_i^* \left[\frac{I\{X_i^* \le g(\boldsymbol{Z}_i^{*T}\boldsymbol{b}), \delta_i^*=1, \eta_i^*=1\}}{\hat{W}(Y_i^*, \boldsymbol{Z}_i^*)} - \tau\right] = 0, \tag{2.3}
$$

where $\hat{W}(y,\boldsymbol{z}) = \hat{G}(y)/\hat{\alpha}(\boldsymbol{z})$. Here $\hat{G}(y)$ and $\hat{\alpha}(\boldsymbol{z})$ are some estimates for $G(y)$ and $\alpha(\boldsymbol{z})$. While there may be multiple choices of $\hat{G}(y)$ and $\hat{\alpha}(\boldsymbol{z})$, here and in the sequel we shall focus on a special case where $\hat{G}(y)$ and $\hat{\alpha}(\boldsymbol{z})$ are constructed as follow.

First, noting that L is only independently right truncated by T_2 under the assumption of $L < C$, one may estimate $F_L(y) \equiv P(L \le y)$ by the Lynden-Bell estimator (Lynden-Bell, 1971; Woodroofe, 1985),

$$
\hat{F}_L(t) = \prod_{s>t} \left\{1 - \frac{\mathrm{d}F_{Ln}^*(s)}{R_n(s)}\right\},
$$

where $F_{Ln}^*(s) = n^{-1}\sum_{i=1}^{n} I(L_i^* \le s)$ and $R_n(s) = \frac{1}{n}\sum_{i=1}^{n} I(L_i^* < s \le Y_i^*)$. Following Chao (1987), we propose to estimate the truncation probability $\alpha(\boldsymbol{z})$ by

$$
\hat{\alpha}(\boldsymbol{z}) = \int_0^{\nu} \hat{S}_{T_2|\boldsymbol{Z}=\boldsymbol{z}}(u)\hat{F}_L(\mathrm{d}u),
$$

where $\hat{S}_{T_2|\boldsymbol{Z}=\boldsymbol{z}}(t)$ is an estimator of $S_{T_2|\boldsymbol{Z}=\boldsymbol{z}}(t) \equiv \Pr(T_2 > t|\boldsymbol{Z}=\boldsymbol{z})$, and ν is an upper bound of Y satisfying regularity condition *C2* in Section 2.1.3. In practice, $\hat{S}_{T_2|\boldsymbol{Z}=\boldsymbol{z}}(t)$ may be obtained by using any existing regression approach for left truncated and right censored data provided that some mild requirements are met. We discuss the detailed requirements on $\hat{S}_{T_2|\boldsymbol{Z}=\boldsymbol{z}}(t)$ in 2.4.1, and show in 2.4.4 that these theoretical requirements are met when the popular Cox proportional hazards model is assumed for T_2.

To estimate $G(y)$, we propose the following estimator,

$$\hat{G}(y) = \frac{1}{n}\sum_{i=1}^{n} \frac{I(L_i^* < y \leq Y_i^*)\hat{\alpha}(\boldsymbol{Z}_i^*)}{\hat{S}_{T_2}(y - |\boldsymbol{Z}_i^*)}.$$

by employing IPCW based on the fact that

$$P(L^* < y \leq Y^*|\boldsymbol{Z}^*) = P(L < y \leq C)P(T_2 \geq y|\boldsymbol{Z}^*)/P(L < T_2|\boldsymbol{Z}^*).$$

In Section 2.4.1, we show that $\hat{G}(y)$ is a uniformly consistent estimator of $G(y)$.

When a proper $\hat{W}(y, \boldsymbol{z})$ is available, equation (2.3) can be easily solved by its monotonicity (Fygenson and Ritov, 1994). Specifically, following similar lines of Peng and Fine (2009, Appendix), we can show that $2n^{1/2}\boldsymbol{S}_n(\boldsymbol{b}, \tau)$ equals the derivative of the following L_1–type convex function with regard to $\boldsymbol{b}$:

$$\begin{aligned} U_n(\boldsymbol{b}, \tau) = \sum_{i=1}^{n} \delta_i^* \eta_i^* |g^{-1}(X_i^*)\hat{W}_i^{-1} - \boldsymbol{b}^T \boldsymbol{Z}_i^* \hat{W}_i^{-1}| + |M_1 - \boldsymbol{b}^T \sum_{l=1}^{n} -(\boldsymbol{Z}_l^* \delta_l^* \eta_l^* \hat{W}_l^{-1})| \\ + |M_1 - \boldsymbol{b}^T \sum_{k=1}^{n} (2\boldsymbol{Z}_k^* \tau)|, \end{aligned}$$

where $\hat{W}_i \equiv \hat{W}(Y_i^*, \boldsymbol{Z}_i^*)$, and M_1 is a sufficiently large positive number that can bound $|\boldsymbol{b}^T \sum_{l=1}^{n} -(\boldsymbol{Z}_l^* \delta_l^* \eta_l^* \hat{W}_l^{-1})|$ and $|\boldsymbol{b}^T \sum_{k=1}^{n} (2\boldsymbol{Z}_k^* \tau)|$. By this fact, $\hat{\boldsymbol{\beta}}(\tau)$ can be equivalently obtained as the minimizer of $U_n(\boldsymbol{b}, \tau)$ using standard software, like the $l1fit()$ function in S-PLUS or the $rq()$ function in the contributed R package *quantreg*.

2.1.3 Asymptotic Results

We derive the asymptotic properties of $\hat{\boldsymbol{\beta}}(\tau)$ under the following regularity conditions:

C1 $\boldsymbol{Z}$ is uniformly bounded, i.e., $\sup_i ||\boldsymbol{Z}_i|| \leq M < \infty$.

C2 (i) There exist $\nu > 0$ such that $P(C = \nu) > 0$ and $P(C > \nu) = 0$; (ii) $a_L \leq a_Y$, $b_L \leq b_Y$; (iii) $\inf_{y \in (a_Y, \nu]} G(y) > 0$; (iv) $\inf_{||\boldsymbol{z}|| \leq M} S_{T_2|\boldsymbol{Z}=\boldsymbol{z}}(\nu) > 0$. Here for a nonnegative random variable Q, we define $a_Q = \inf\{q \geq 0 : P(Q \leq q) > 0\}$, $b_Q = \sup\{q \geq 0 : P(Q \geq q) > 0\}$.

C3 (i) $\boldsymbol{\beta}_0(\tau)$ is Lipschitz continuous for $\tau \in [\tau_L, \tau_U]$; (ii) $f_1(t|\boldsymbol{z})$ is continuous and bounded above uniformly in t and z, where $f_1(t|\boldsymbol{z}) = \mathrm{d}F_1(t|\boldsymbol{z})/\mathrm{d}t$.

C4 For some $\rho_0 > 0$ and $c_0 > 0$, $\inf_{b \in \mathcal{B}(\rho_0)} \mathrm{eigmin} \boldsymbol{A}(\boldsymbol{b}) \geq c_0$. Here $\mathcal{B}(\rho) = \{\boldsymbol{b} \in R^{p+1} : \inf_{\tau \in [\tau_L, \tau_U]} ||\boldsymbol{b} - \boldsymbol{\beta}_0(\tau)|| \leq \rho\}$, $\boldsymbol{A}(\boldsymbol{b}) = E[\boldsymbol{Z}^{\otimes 2} f_1\{g(\boldsymbol{Z}^T\boldsymbol{b})|\boldsymbol{Z}\} g'(\boldsymbol{Z}^T\boldsymbol{b})]$, $g'(\cdot)$ represents the derivative of link function $g(\cdot)$, and $\boldsymbol{u}^{\otimes 2} = \boldsymbol{u}\boldsymbol{u}^T$ for a vector $\boldsymbol{u}$.

We briefly comment on the regularity conditions as follows. First, *C1* requires bounded covariates, which is reasonable in many realistic situations. *C2*(i) imposes a mild assumption on the censoring time. For example, *C2*(i) is often satisfied in clinical settings with administrative censoring. One may adopt a truncated censoring time $C^* = min(C, C_U)$, where C_U is chosen to be slightly less than b_C, to make *C2*(i) a more realistic assumption at a price of small information loss. *C2*(ii) implies that (L, T_2) is observable throughout the whole support and thus ensures the identifiability of $S_{T_2|\boldsymbol{Z}=\boldsymbol{z}}(t)$. By *C2*(iii)–(iv), $G(y) = P(L < y \leq C)$ and $R(y) \equiv \Pr(L < y \leq Y|L < Y)$ are bounded away uniformly from 0, and this is necessary for the consistency of the plug-in weight, $\hat{\alpha}(\boldsymbol{z})/\hat{G}(y)$. *C3* assumes that the cause-1 subdensity is uniformly bounded and smooth to the second order. *C4* requires the asymptotic limit of $U_n(\boldsymbol{b}, \tau)$ be strictly convex in a neighborhood of $\boldsymbol{\beta}_0(\tau)$ for $\tau \in [\tau_L, \tau_U]$, which is critical for the asymptotic identifiability of $\boldsymbol{\beta}_0(\tau)$ and the uniform consistency of $\hat{\boldsymbol{\beta}}(\tau)$. Under these conditions, we have the following theorems:

Theorem 2.1.1. *Under conditions C1-C4,* $\lim_{n\to\infty}\sup_{\tau\in[\tau_L,\tau_U]}||\hat{\boldsymbol{\beta}}(\tau)-\boldsymbol{\beta}_0(\tau)||\xrightarrow{p}0$.

Theorem 2.1.2. *Under conditions C1-C4,* $n^{1/2}\{\hat{\boldsymbol{\beta}}(\tau)-\boldsymbol{\beta}_0(\tau)\}$ *converge weakly to a mean zero Gaussian process for* $\tau\in[\tau_L,\tau_U]$ *with covariance matrix*

$$A\{\boldsymbol{\beta}_0(\tau')\}^{-1}E[\boldsymbol{\zeta}(\tau')\boldsymbol{\zeta}(\tau)]A\{\boldsymbol{\beta}_0(\tau)\}^{-T},$$

where the formal definition of $\boldsymbol{\zeta}(\tau)$ *is provided in Section 2.4.2.*

To establish the asymptotic properties of $\hat{\boldsymbol{\beta}}(\tau)$, we first prove the uniform consistency and weak convergence of the plug-in weight, $\hat{\alpha}(\boldsymbol{z})/\hat{G}(y)$, by using the empirical process technique (Van der Vaart and Wellner, 1996). The uniform consistency of $\hat{\boldsymbol{\beta}}(\tau)$ for $\tau\in[\tau_L,\tau_U]$ is facilitated by the monotonicity of the estimating function. For the weak convergence of $n^{1/2}\{\hat{\boldsymbol{\beta}}(\tau)-\boldsymbol{\beta}_0(\tau)\}$, one challenge lies in the non-smoothness of $\boldsymbol{S}_n(\boldsymbol{b},\tau)$. This problem is handled by similar lines of Alexander (1984) and Lai and Ying (1988). The detailed proofs of Theorems 2.1.1 and 2.1.2, along with the influence functions of $n^{1/2}\{\hat{\boldsymbol{\beta}}(\tau)-\boldsymbol{\beta}_0(\tau)\}$, are provided in Section 2.4.1 and 2.4.2.

2.1.4 Covariance Estimation

In Section 2.4.2, we show that $\boldsymbol{S}_n\{\boldsymbol{\beta}_0(\tau),\tau\}$ converges weakly to a mean zero tight Gaussian Process with covariance matrix $\boldsymbol{\Sigma}(\tau',\tau)=\boldsymbol{E}[\boldsymbol{\zeta}(\tau')\boldsymbol{\zeta}(\tau)]$, which can be consistently estimated by $\hat{\boldsymbol{\Sigma}}(\tau',\tau)=\frac{1}{n}\sum_{i=1}^n\hat{\boldsymbol{\zeta}}_i(\tau')\hat{\boldsymbol{\zeta}}_i(\tau)$. Here $\hat{\boldsymbol{\zeta}}_i(\tau)$ are $\boldsymbol{\zeta}_i(\tau)$ with unknown quantities replaced by their empirical counterparts or consistent estimators. In order to estimate the covariance matrix of $\sqrt{n}\{\hat{\boldsymbol{\beta}}(\tau)-\boldsymbol{\beta}_0(\tau)\}$, one may use the resampling method of Parzen et al. (1994) or Jin et al. (2001). Such an approach is quite straightforward but may be computationally intensive especially with a large sample size. By utilizing the asymptotic linearity of $\boldsymbol{S}_n\{\boldsymbol{\beta}(\tau),\tau\}$ in the vicinity of $\boldsymbol{\beta}_0(\tau)$, we derive a consistent plug-in estimator for the covariance matrix of $\sqrt{n}\{\hat{\boldsymbol{\beta}}(\tau)-\boldsymbol{\beta}_0(\tau)\}$. The covariance estimation procedure mimics the method introduced by Huang (2002) and is described as follows. A brief justification is provided in Section 2.4.3.

1. Use spectral decomposition to find a symmetric matrix $\boldsymbol{E}_n(\tau)$ such that $\boldsymbol{E}_n(\tau)^2 = \hat{\boldsymbol{\Sigma}}(\tau,\tau)$;

2. Calculate $\boldsymbol{D}_n(\tau) = \left[\boldsymbol{S}_n^{-1}\{\boldsymbol{e}_{n,1}(\tau),\tau\} - \hat{\boldsymbol{\beta}}(\tau), ..., \boldsymbol{S}_n^{-1}\{\boldsymbol{e}_{n,p+1}(\tau),\tau\} - \hat{\boldsymbol{\beta}}(\tau)\right]$, where $\boldsymbol{e}_{n,j}(\tau)$ is the jth column of $\boldsymbol{E}_n(\tau)$, and $\boldsymbol{S}_n^{-1}(\boldsymbol{e},\tau)$ denotes the solution to $\boldsymbol{S}_n(\boldsymbol{b},\tau)=\boldsymbol{e}$.

3. A consistent estimate for the asymptotic covariance matrix of $\sqrt{n}\{\hat{\boldsymbol{\beta}}(\tau)-\boldsymbol{\beta}_0(\tau)\}$ is given by

$$n\boldsymbol{D}_n(\tau')\boldsymbol{E}_n^{-1}(\tau')\hat{\boldsymbol{\Sigma}}(\tau',\tau)\boldsymbol{E}_n(\tau)^{-1}\boldsymbol{D}_n(\tau)^T.$$

   In the special case that $\tau'=\tau$, we get a consistent estimate for the asymptotic variance matrix given by $n\boldsymbol{D}_n^{\otimes 2}(\tau)$.

2.1.5 Other Inferences

In this subsection we study inferences which respectively address the following two questions: (1) how to summarize the overall effect of a given covariate and assess its significance? (2) does the effect keep constant over different quantiles?

Let $v^{(j)}$ denote the jth component of a vector $\boldsymbol{v}$. For question (1), we proposed to summarize the quantile effects of $Z^{(j)}$ by a trimmed mean effect defined as $\Phi_{1,j}\{\boldsymbol{\beta}_0(\tau)\} = \frac{1}{\tau_L-\tau_U}\int_{\tau_L}^{\tau_U}\beta_0^{(j)}(\tau)\mathrm{d}\tau,\ j=1,\ldots,p+1$. However, caution needs be exercised in the interpretation of $\Phi_{1,j}\{\hat{\boldsymbol{\beta}}(\tau)\}$ when a cross-over effect pattern is present. As a natural estimator of $\Phi_{1,j}\{\boldsymbol{\beta}_0(\tau)\}$, $\Phi_{1,j}\{\hat{\boldsymbol{\beta}}(\tau)\}$, preserves the consistency and asymptotic normality of $\hat{\boldsymbol{\beta}}(\tau)$ because $\Phi_{1,j}(\cdot)$ is compactly differentiable (Andersen et al., 1993). Furthermore, it can be shown that $\sqrt{n}[\boldsymbol{\Phi}_1\{\hat{\boldsymbol{\beta}}(\tau)\}-$

$\boldsymbol{\Phi}_1\{\boldsymbol{\beta}_0(\tau)\}]$ is asymptotically equivalent to $n^{-1/2}\sum_{i=1}^n \frac{1}{\tau_L-\tau_U}\int_{\tau_L}^{\tau_U} \boldsymbol{A}\{\boldsymbol{\beta}_0(\tau)\}^{-1}\boldsymbol{\zeta}_i(\tau)\mathrm{d}\tau$, where $\boldsymbol{\Phi}_1(\boldsymbol{u}) = \{\Phi_{1,1}(\boldsymbol{u}),\ldots,\Phi_{1,p+1}(\boldsymbol{u})\}^T$ and $\boldsymbol{A}\{\boldsymbol{\beta}_0(\tau)\}$ is defined in regularity condition *C4*. In the justification for the proposed covariance estimation presented in 2.4.3, we show that $\sqrt{n}\boldsymbol{D}_n(\tau)\boldsymbol{E}_n(\tau)^{-1}$ is a consistent estimate for $\boldsymbol{A}\{\boldsymbol{\beta}_0(\tau)\}^{-1}$. Therefore the covariance of $\sqrt{n}\boldsymbol{\Phi}_1\{\hat{\boldsymbol{\beta}}(\tau)\}$ can be consistently estimated by

$$n^{-1}\sum_{i=1}^n\left\{\frac{1}{\tau_L-\tau_U}\int_{\tau_L}^{\tau_U}\sqrt{n}\boldsymbol{D}_n(\tau)\boldsymbol{E}_n(\tau)^{-1}\hat{\boldsymbol{\zeta}}_i(\tau)\mathrm{d}\tau\right\}^{\otimes 2},$$

This result renders a Wald-type test for the null hypothesis $H_{01}:\ \Phi_{1,j}\{\boldsymbol{\beta}_0(\tau)\} = 0,\ \ \tau\in[\tau_L,\ \tau_U]$. Testing H_{01} can provide an overall evaluation of the significance of $Z^{(j)}$'s quantile effects $(j = 2,\ldots,p+1)$.

We propose to formulate (2) as a hypothesis testing problem with the null hypothesis, $H_{02}:\beta_0^{(j)}(\tau) = c_0, \tau\in[\tau_L,\tau_U]$, where c_0 is an unspecified constant and $j = 2,\ldots,p+1$. If H_{02} holds, then $\int_{\tau_L}^{\tau_U}\Xi(\tau)\beta_0^{(j)}(\tau)\mathrm{d}\tau - \Phi_{1,j}\{\boldsymbol{\beta}_0(\tau)\} = 0$ must hold for any weight function $\Xi(\tau)$ that satisfies $\int_{\tau_L}^{\tau_U}\Xi(\tau)\mathrm{d}\tau = 1$. This fact motivates a test statistic constructed based on $\Phi_{2,j}(\hat{\boldsymbol{\beta}}) = \int_{\tau_L}^{\tau_U}\Xi(\tau)\hat{\beta}^{(j)}(\tau)\mathrm{d}\tau - \Phi_{1,j}(\hat{\boldsymbol{\beta}})$ $(j = 1,\ldots,p+1)$. Define $\boldsymbol{\Phi}_2(\boldsymbol{u}) = \{\Phi_{2,1}(\boldsymbol{u}),\ldots,\Phi_{2,p+1}(\boldsymbol{u})\}^T$. Using similar arguments to those for covariance estimation, we can show that the covariance of $\sqrt{n}\boldsymbol{\Phi}_2(\hat{\boldsymbol{\beta}})$ can be consistently estimated by

$$n^{-1}\sum_{i=1}^n\left[\int_{\tau_L}^{\tau_U}\Xi(\tau)\left\{\sqrt{n}\boldsymbol{D}_n(\tau)\boldsymbol{E}_n(\tau)\hat{\boldsymbol{\zeta}}_i(\tau)\right\}\mathrm{d}\tau - \left\{\frac{1}{\tau_L-\tau_U}\int_{\tau_L}^{\tau_U}\sqrt{n}\boldsymbol{D}_n(\tau)\boldsymbol{E}_n(\tau)^{-1}\hat{\boldsymbol{\zeta}}_i(\tau)\mathrm{d}\tau\right\}\right]^{\otimes 2}.$$

A Wald-type test can thus be developed for H_{02} accordingly. In practice, one may choose $\Xi(\tau)$ based on the observed $\hat{\boldsymbol{\beta}}(\tau)$ or the scientifically conjectured trajectory of $\boldsymbol{\beta}_0(\tau)$ in order to better capture the departure from H_{02} and therefore boost the power.

2.2 Numerical Studies

2.2.1 Simulations

Simulation studies were conducted to examine the finite-sample performance of the proposed methods. Specifically, we generated (T, T_2) as

$$\begin{aligned} T &= I(\varepsilon = 1)\exp(\boldsymbol{\gamma}^T \tilde{\boldsymbol{Z}} + e_2 - e_1) + I(\varepsilon = 0)\exp(\boldsymbol{c}^T \tilde{\boldsymbol{Z}} + e_2), \\ T_2 &= \exp(\boldsymbol{c}^T \widetilde{\boldsymbol{Z}} + e_2). \end{aligned}$$

Here the covariate vector $\widetilde{\boldsymbol{Z}} = (Z_1, Z_2)^T$ with $Z_1 \sim Unif(0,1)$ and $Z_2 \sim Bernoulli(0.5)$. We chose $e_2 = \log\{-\log[Unif(0,1)]\}/3.2$ and $c = (0.3, 0.3)^T/3.2$ such that T_2 followed a Cox-Weibull model with shape $= 3.2$, scale $= 1$, and coefficients$= (-0.3, -0.3)^T$ (Bender et al., 2005). We set $\varepsilon \sim Bernoulli(p_1)$, $p_1 = 0.7$, $e_{11} \sim Beta(2,2)/2$, and $e_{12} \sim Exponential(1)/2$. It can be shown that the underlying quantile regression model satisfies

$$\log\{Q_1(\tau|\boldsymbol{Z})\} = \widetilde{Q}_{11}\left(\frac{\tau}{p_1}\right) + \gamma^{(1)} Z_1 + \left\{\gamma^{(2)} + \widetilde{Q}_{12}\left(\frac{\tau}{p_1}\right) - \widetilde{Q}_{11}\left(\frac{\tau}{p_1}\right)\right\} Z_2,$$

where $\widetilde{Q}_{11}(\cdot)$ and $\widetilde{Q}_{12}(\cdot)$ are quantile functions of $e_2 - e_{11}$ and $e_2 - e_{12}$ respectively. Therefore, Z_1 has a constant effect on $Q_1(\tau|\boldsymbol{Z})$, while the effect of Z_2 increases with τ. We set $\boldsymbol{\gamma} \equiv (\gamma^{(1)}, \gamma^{(2)})^T$ to be either $(0,0)^T$ or $(-0.5, 0)^T$ in the simulation studies.

Mimicking the Denmark diabetes registry study, we generated left truncation time L from a mixture of a point mass at zero and a positive random variable, and let independent censoring time C have a point mass at a common upper bound. We considered the following two set-ups:

(A) $L = r_L \times 1.48Beta(2.5, 2)$ and $C = r_C \times [L + Beta(3.5, 1)(1.7 - L)] + (1 - rC) \times 1.7$;

(B) $L = r_L \times 1.48Beta(4.7, 2)$, $C = r_C \times [L + Beta(3.8, 1)(1.7 - L)] + (1 - rC) \times 1.7$,

where $r_L \sim Bernoulli(0.8)$ and $r_C \sim Bernoulli(0.8)$. The truncation rates are 30%

and 45% in set-up (A) and set-up (B) respectively. Under each set-up, we implemented the proposed method on 2000 simulated datasets with $g(\cdot) = \exp(\cdot)$, $M_1 = 10^6$ and sample size $n = 200$ or 400.

In Table 2.1, we present the empirical biases (EmpBias), empirical standard deviations (EmpSD), average estimated standard deviations (EstSD), and 95% empirical coverage probabilities (COV95) for the proposed estimator of $\boldsymbol{\beta}_0(\tau)$ with $\tau = 0.1, 0.2, 0.3, 0.4$. It is observed that our estimates are virtually unbiased, and the estimated standard deviations agree well with empirical standard deviations. The standard deviations tend to be smaller when $\tau = 0.3$ and $\tau = 0.4$ compared to those when $\tau = 0.1$ and $\tau = 0.2$. The coverage probabilities of the 95% CIs match the nominal level reasonably well and improve with sample size. In Table 2.1, we also provide empirical biases of the estimate that does not account for left truncation (PFBias), obtained by implementing Peng and Fine (2009)'s method which treats all truncation times as 0. It is shown that failing to account for left truncation can lead to substantial biases. Unreported simulations also suggest that a larger estimation bias may be resulted from ignoring left truncation when the dependence between T_1 and T_2 increases.

We further examined the proposed estimator of the trimmed mean covariate effect as well as the Wald tests for H_{01} and H_{02}. We set $\tau_L = 0.1$ and $\tau_U = 0.45$, and chose $\Xi(\tau) = 2I(\tau \leq \frac{\tau_L+\tau_U}{2})/(\tau_U - \tau_L)$. In Table 2.2, we summarize the empirical biases (EmpBias), empirical standard deviations (EmpSD), mean estimated standard deviations (EstSD) of $\Phi_{1,j}(\hat{\boldsymbol{\beta}})$ $(j = 2, 3)$, and the empirical rejection rates of the two Wald tests (EmpRR). Empirical biases of estimated trimmed mean effects are small for both Z_1 and Z_2. Standard deviation estimates are quite accurate. The test for H_{01} performs rather well regarding both size and power. For H_{02}, the size can be a little over conservative, but the performance improves with the sample size.

We performed additional simulations to assess the robustness of the proposed

Table 2.1: Summary of simulation results: empirical biases of Peng and Fine (2009)'s estimator (PFBias); empirical biases (EmpBias), empirical standard deviations (EmpSD), mean estimated standard deviations (EstSD), and 95% empirical coverage probabilities (COV95) of the proposed estimator at $\tau = 0.1, 0.2, 0.3, 0.4$.

	PFBias $\times 10^3$			**EmpBias** $\times 10^3$			**EmpSD** $\times 10^2$			**EstSD** $\times 10^2$			**COV95** $\times 10^3$		
N	$\hat{\boldsymbol{\beta}}_N^{(0)}$	$\hat{\boldsymbol{\beta}}_N^{(1)}$	$\hat{\boldsymbol{\beta}}_N^{(2)}$	$\hat{\boldsymbol{\beta}}^{(0)}$	$\hat{\boldsymbol{\beta}}^{(1)}$	$\hat{\boldsymbol{\beta}}^{(2)}$	$\hat{\boldsymbol{\beta}}^{(0)}$	$\hat{\boldsymbol{\beta}}^{(1)}$	$\hat{\boldsymbol{\beta}}^{(2)}$	$\hat{\boldsymbol{\beta}}^{(0)}$	$\hat{\boldsymbol{\beta}}^{(1)}$	$\hat{\boldsymbol{\beta}}^{(2)}$	$\hat{\boldsymbol{\beta}}^{(0)}$	$\hat{\boldsymbol{\beta}}^{(1)}$	$\hat{\boldsymbol{\beta}}^{(2)}$
							(A), $\gamma = (-0.5, 0)^T$								
200	201	−5	−32	−10	1	16	28	45	27	31	50	30	894	916	914
	158	−3	−13	−4	1	9	19	32	19	22	35	21	919	932	930
	117	−1	−5	1	0	6	16	27	16	18	30	18	938	942	938
	104	−4	−1	6	0	5	15	25	15	18	29	17	955	957	953
400	204	−5	−24	2	0	3	20	32	19	21	35	21	897	913	923
	159	−10	−2	1	1	3	14	22	14	15	24	15	928	933	927
	114	−10	6	3	1	2	11	19	11	13	21	12	943	955	945
	091	−12	8	4	1	6	10	17	10	12	19	11	952	957	949
							(B), $\gamma = (-0.5, 0)^T$								
200	245	8	−29	10	−8	0	29	47	29	32	53	32	878	906	908
	218	−1	−7	12	−10	−1	21	34	21	23	38	23	912	929	925
	190	−6	7	12	−8	−1	17	29	17	20	33	20	935	947	939
	160	−6	12	14	−6	0	16	27	16	20	32	19	954	960	955
400	253	12	−27	−10	11	13	21	33	20	23	37	22	902	908	919
	215	−1	−6	−1	0	9	15	24	15	16	27	16	929	936	929
	184	−5	7	2	2	1	12	20	13	14	23	14	941	949	944
	160	−10	13	5	2	3	11	19	12	13	22	13	956	957	951
							(A), $\gamma = (0, 0)^T$								
200	203	−9	−27	−5	−3	12	27	44	28	31	50	30	889	918	899
	152	−2	−2	−3	4	6	19	31	20	21	35	21	916	934	924
	126	−6	4	1	1	3	16	27	16	18	30	18	939	944	932
	105	−6	8	7	−2	3	15	25	15	18	29	17	949	960	949
400	196	−1	−18	6	−7	7	19	32	20	21	35	21	891	924	907
	151	−2	6	3	1	3	14	23	14	15	24	15	924	939	932
	127	−7	10	3	3	3	11	19	12	12	20	12	941	944	932
	106	−10	12	5	1	4	11	17	11	12	19	12	948	955	946
							(B), $\gamma = (0, 0)^T$								
200	241	21	−25	0	−2	10	29	47	29	33	53	32	893	914	912
	212	5	−8	5	−10	10	20	34	21	24	39	23	920	934	928
	186	−2	1	6	−7	8	17	29	17	20	34	20	938	944	936
	164	−12	7	11	−6	6	16	27	16	20	33	19	964	962	954
400	257	−5	−26	−5	4	14	21	34	21	23	38	22	896	916	919
	222	−9	−7	0	4	5	15	24	15	16	27	16	934	936	933
	190	−9	5	5	−1	4	12	21	12	14	23	14	944	948	934
	163	−14	10	6	2	5	12	19	11	13	22	13	954	954	958

Table 2.2: Summary of simulation study: empirical biases (EmpBias), empirical standard deviations (EmpSd), mean estimated standard deviations (EstSd) of $\hat{\Phi}_{1,j}(\hat{\boldsymbol{\beta}})$ ($j = 2, 3$), and empirical rejection rates of H_{01} (EmpRR1) and H_{02} (EmpRR2) at significance level of 0.05.

				$\hat{\Phi}_{1,j}(\hat{\boldsymbol{\beta}})$		H_{01}	H_{02}
	j	n	EmpBias ×1000	EmpSD ×100	EstSD ×1000	EmpRR1	EmpRR2
				$\gamma = (-0.5, 0)^T$			
(A)	2	200	1	27	28	0.463	0.027
	3	200	8	16	17	0.360	0.199
	2	400	0	19	19	0.750	0.035
	3	400	4	11	12	0.648	0.480
(B)	2	200	−7	29	30	0.417	0.022
	3	200	1	17	18	0.325	0.163
	2	400	2	20	21	0.659	0.036
	3	400	6	12	13	0.562	0.372
				$\gamma = (0, 0)^T$			
(A)	2	200	1	26	27	0.047	0.023
	3	200	5	16	17	0.376	0.199
	2	400	4	19	19	0.047	0.038
	3	400	−8	12	12	0.652	0.470
(B)	2	200	−7	29	30	0.047	0.023
	3	200	−4	17	18	0.313	0.160
	2	400	2	20	21	0.054	0.033
	3	400	−6	12	13	0.560	0.388

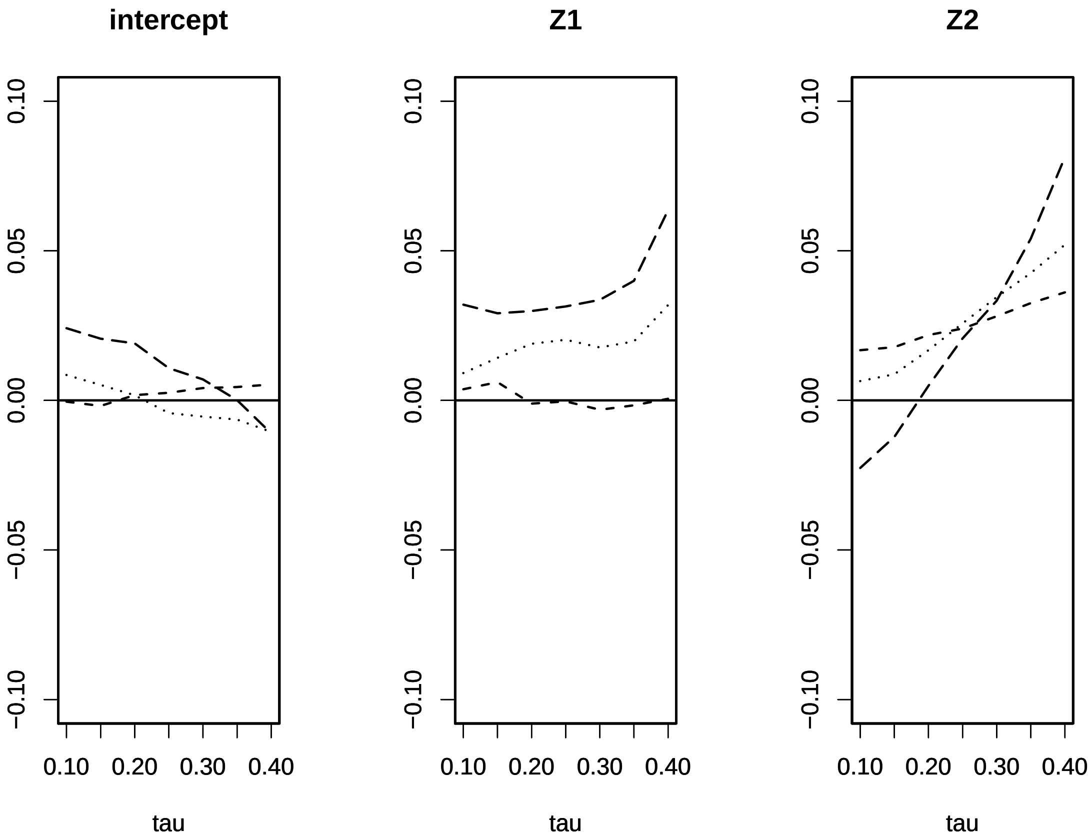
estimator of $\boldsymbol{\beta}_0(\tau)$ to misspecification of the regression model for T_2. Specifically, we replaced e_2 by $e_{2A} \sim 2Weibull(shape = 5, scale = 1) - 2$, $e_{2B} \sim 2Beta(2, 2) - 1.17$, and $e_{2C} \sim 2Unif(0, 1) - 1.17$. This leads to situations where the Cox proportional hazards model is not the true model for T_2. Among the three substitutes of e_2, e_{2A} is most similar to e_2 and e_{2C} is least similar to e_2, thereby respectively representing the cases with the smallest departure and the largest departure from the proportional hazards assumption. In Figure 2.1, we plot empirical biases observed in these scenarios where T_2 is incorrectly specified to follow a Cox model. The magnitudes of biases are rather small, mostly below 0.05. As expected, we also observe an increasing trend in bias with the degree of violation of the assumed model. Overall, our simulation results suggest quite robust performance of the proposed estimator.

2.2.2 Denmark Diabetes Registry Data Analysis

The Denmark diabetes registry study (Andersen et al., 1993) is a prospective cohort study on insulin-dependent diabetes patients aged below 31 and referred to the Steno Memorial Hospital in Greater Copenhagen between 1933 to 1981. Patients were followed until death, immigration, or the end of the study at December 31, 1984. Stratifications on diabetic onset age and birth cohort may be necessary to avoid non-homogeneity. For example, those with early diabetic onset age may be under more influence from genetic defects when compared to those who developed diabetes later. In this analysis, we focus on a subcohort of patients who were born before 1940 and had diabetic onset age greater or equal than 15. Among the 858 patients in this subcohort, 29% experienced diabetic nephropathy (DN) and 40% died in the study duration. Approximately 19% of patients had diabetic onset at the study entry and therefore had zero truncation times. This implies the identifiability of $S_{T_2|\boldsymbol{Z}=\boldsymbol{z}}(t)$.

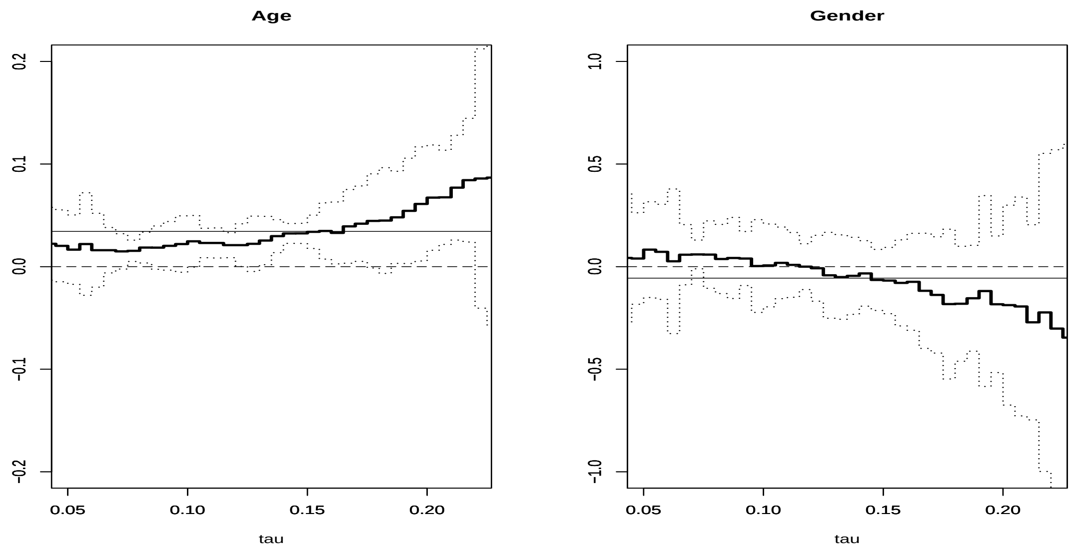
We fit model (2.1) to this dataset. Covariates considered include patient's gender and diabetes onset age. We adopt the link function $g(\cdot) = \exp(\cdot)$ and $M = 10^6$. We

Figure 2.1: Empirical biases of the proposed estimator when T_2 is assumed to follow a Cox model but is generated from an AFT model with e_{2A} (dashed lines), e_{2B} (dotted lines), and e_{2C} (long dashed lines).



shift onset age down by 15 years, so that the intercept term corresponds to female patients who had diabetic onset at age 15. Figure 2.2 plots the estimated coefficients for Age and Gender, along with the corresponding 95% confidence intervals, and the estimated trimmed mean quantile effects for $\tau \in [0.05, 0.22]$. It is observed from Figure 2 that diabetes onset age appears to have a significant effect on the cumulative incidence quantiles of DN. Patients with older diabetes onset age tend to have longer DN free survival time. It is interesting to note that the coefficients for age tend to increase with τ, suggesting a stronger influence of age at diabetes onset on time to DN for patients who developed DN rather late as compared to those who had DN shortly after the onset of diabetes. The estimated coefficients for gender and confidence intervals suggest that there may be little difference in the cumulative incidence quantiles of DN between males and females with $\tau \in [0.05, 0.22]$.

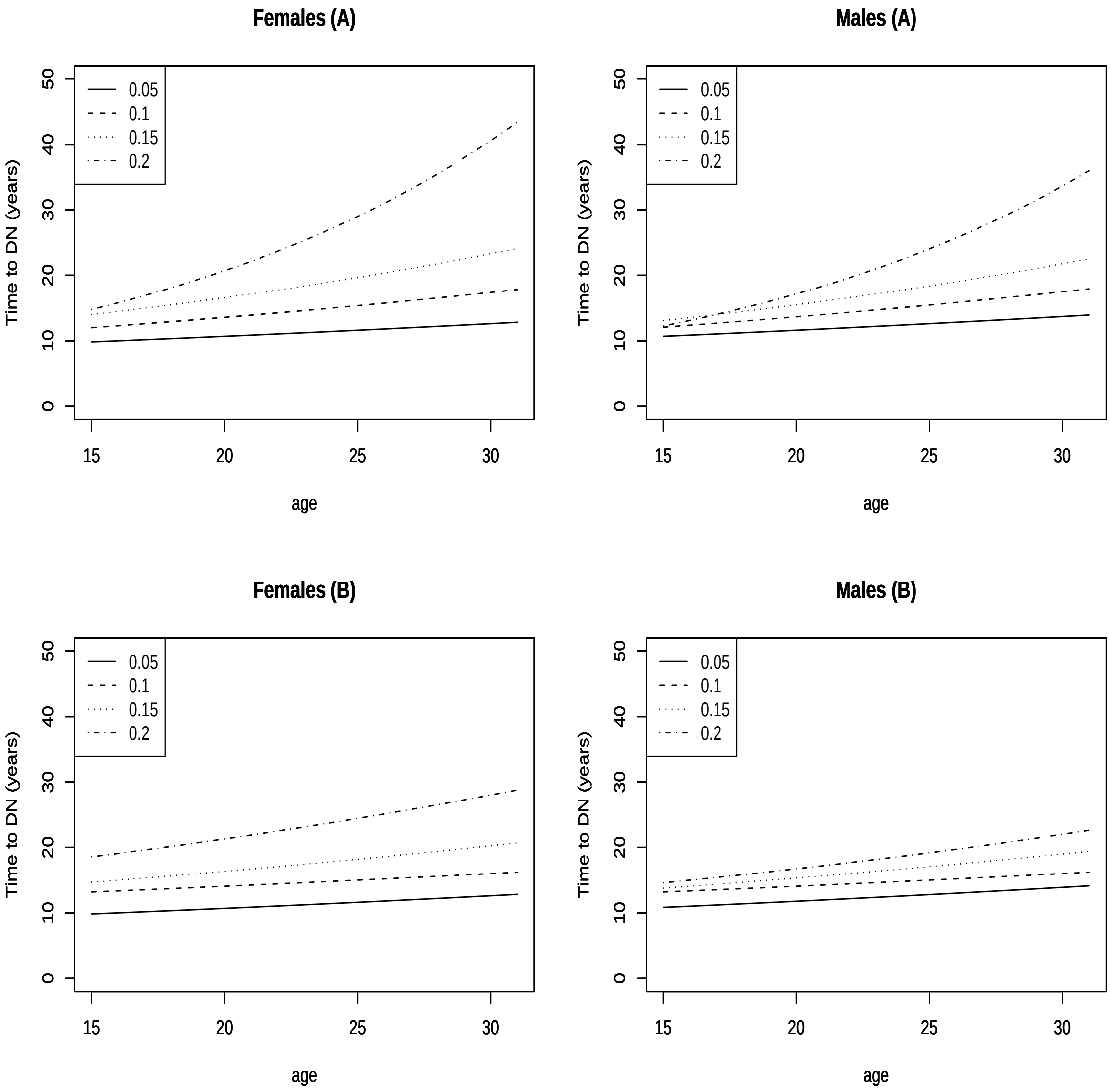
Figure 2.2: Denmark diabetes registry study: estimated coefficients (bold solid lines), the corresponding pointwise 95% confidence intervals (dotted lines), and trimmed mean effect estimates (solid lines) for age and gender.



In our second-stage exploration, we estimate the trimmed mean effect of age on DN by 0.035 and the corresponding standard deviation by 0.013. This renders a p value of 0.008 for testing H_{01} and provides some evidence for the overall significance of the diabetes onset age effect on time to DN. The estimated trimmed mean effect of gender is -0.059 with a standard error of 0.102. This is consistent with Figure 2.2 where the confidence intervals for the gender coefficients mostly cover zero. However, caution should be exercised when interpreting this trimmed mean effect, since we observe cross-over effects in gender from Figure 2.2. We also conduct hypothesis testing for H_{02} to investigate whether age's effect is constant or varying over quantiles. We choose $\Xi(\tau) = 2I\{\tau \leq (\tau_L + \tau_U)/2\}/(\tau_U - \tau_L)$ in view of the monotonicity pattern of the age coefficient estimates displayed by Figure 2.2. We obtain $\hat{\Phi}_{2,j} = -0.015$ and its standard error equals 0.007, leading to a Wald test of -2.048 and thus a p value of 0.04. This result confirms our observation from Figure 2.2 that age's effect on the cumulative incidence quantiles of DN may increase with τ. This suggests that age may play a more important role in DN progression for patients with relative lower risk of DN. Applying the same test on the gender coefficients suggests that a constant effect may be adequate ($p = 0.17$).

Based on model (2.1), we also plot predicted cumulative incidence quantiles of time to DN with $\tau = 0.05,\ 0.1,\ 0.15$ and 0.2; see Figure 2.3. For both males and females, the predicted τth quantile increases with diabetes onset age, and the age seems more influentious with larger τs. Figure 2.3 well reflects our findings from Figure 2.2. First, the increasing trend of each curve indicates that a younger diabetic onset age may be associated with a quicker progression to DN. Secondly, the rather flat quantile curves corresponding to small τ's, such as $\tau = 0.05$, reflect the diminished prognostic value of age for DN progression among high DN risk patients. Lastly, we observe that the DN cumulative incidence quantiles appear to follow similar patterns for females and males. In Figure 2.3, we also plot the predicted quantile functions obtained from using Peng and Fine (2009)'s approach. It is suggested that ignoring left truncation may not only lead to decreased age effect for both females and males, but also fail to detect the distinction in age effect between patients with high DN risk and those with low DN risk.

Figure 2.3: Denmark diabetes registry study: estimated quantiles for females and males based on the proposed approach (the two top panels) and Peng and Fine (2009)'s approach (2009) (the two bottom panels).



In summary, by modeling cumulative incidence quantiles of time to DN, our regression analysis of the diabetes registry data offers a sensible and comprehensive view about the relationship between the DN endpoint and covariates of interest, among this specific cohort of patients born before 1940 and had diabetic onset age between 15 and 31. The new regression procedure may be recommended for practical use in similar studies.

2.3 Discussion

In this Chapter, we propose a regression approach for left-truncated semi-competing risks data. By formulating covariate effects on quantiles of the conditional cumulative incidence function, our model has the flexibility to accommodate covariates with non-constant effects while providing simple and meaningful interpretations. The developed estimation and inference procedures well utilize the semi-competing risks structure and account for left truncation, and can be efficiently implemented using existing functionality in some standard statistical software.

In practice, choosing the quantile range when applying our method may follow the same recommendation for competing risks quantile regression provided in Peng and Fine (2009). That is, by regularity conditions *C4* and *C5*, the τ in model (1) should satisfy $Q_1(\tau|\boldsymbol{Z}) < \infty$. With a real dataset, one may adaptively select τ according to some exploratory analysis, for example, plotting cumulative incidence curves stratified on covariates.

It is worth noting that $\hat{\boldsymbol{\beta}}(\tau)$ is a piecewise-constant function, which only jumps at the observed T_1 and T_2 in the one-sample case. In general regression settings, it is however rather difficult to determine the exact jump points of $\hat{\boldsymbol{\beta}}(\cdot)$. In practice, one may approximate $\{\hat{\boldsymbol{\beta}}(\tau),\ \tau \in [\tau_L, \tau_U]\}$ by a step function superimposed to $\hat{\boldsymbol{\beta}}(\tau)$ evaluated on a fine τ–grid with step size of order $o(n^{-1/2})$. Provided the smoothness of $\boldsymbol{\beta}_0(\cdot)$, it can be shown that such obtained estimator of $\boldsymbol{\beta}_0(\cdot)$ retains the asymptotic properties stated in Theorems 2.1.1 and 2.1.2.

As suggested by one referee, a multi-state approach, such as the illness–death model, may be developed for left-truncated semi-competing risks data. Such a regression strategy is different from the one adopted in this paper but would allow for inference on transition from disease to death and transition from disease-free to death in addition to T_1. Investigations along this line merit future research.

In this Chapter we assume the independence between (L, C) and $\boldsymbol{Z}$ for presentation simplicity. In practice, one may relax this assumption by imposing additional semiparametric modeling of (L, C) given $\boldsymbol{Z}$ which can render a reasonable estimator for $G(y, \boldsymbol{z})$. This may only demand rather minor changes in inferences and asymptotics. Furthermore, the proposed methods can be readily adapted to the competing risks setting after slight modifications. Therefore, this work has potentially broader applications in practice.

2.4 Proofs

Define

$$\begin{aligned}
\boldsymbol{S}_n(\boldsymbol{b},\tau) &= n^{-1/2}\sum_{i=1}^{n}\boldsymbol{Z}_i^*[I\{X_i^* \le g(\boldsymbol{Z}_i^{*T}\boldsymbol{b}), \delta_i^*\eta_i^* = 1\}\hat{G}^{-1}(Y_i^*)\hat{\alpha}(\boldsymbol{Z}_i^*) - \tau],\\
\boldsymbol{S}_n^G(\boldsymbol{b},\tau) &= n^{-1/2}\sum_{i=1}^{n}\boldsymbol{Z}_i^*[I\{X_i^* \le g(\boldsymbol{Z}_i^{*T}\boldsymbol{b}), \delta_i^*\eta_i^* = 1\}G^{-1}(Y_i^*)\alpha(\boldsymbol{Z}_i^*) - \tau],\\
\boldsymbol{\mu}(\boldsymbol{b},\tau) &= n^{-1/2}E\{\boldsymbol{S}_n{}^G(\boldsymbol{b},\tau)\}.
\end{aligned}$$

Without loss of generality, we assume $a_Y = 0$. Let $\mathcal{T} = (0,\nu]$, $\mathcal{Z} = \{\boldsymbol{z} : ||\boldsymbol{z}|| \le M\}$.

2.4.1 Proof of Theorem 1

The first step is to sort out the asymptotic properties of $\hat{\alpha}(\boldsymbol{z})/\hat{G}(y)$. To this end, we need to look at each specific element of this plug-in weight. Define $R(y) = P(L < y \le Y|L < Y)$, it follows from *C2* that $\inf_{y\in(0,\nu]}R(y)$ is bounded away from 0. Define

$$\tilde{L}_i(y) = \int_y^{\nu}\frac{I(L_i^* < u \le Y_i^*)}{\{R(u)\}^2}F_L^*(\mathrm{d}u) - \frac{I(L_i^* > y)}{R(L_i^*)}$$

where $F_L^*(l) = P(L \le l|L < Y)$. By the results in Chao (1987), for the Kaplan-Meier Estimators $\hat{F}_L(\cdot)$, we have

$$\hat{F}_L(t) - F_L(t) = \frac{1}{n}\sum_{i=1}^{n}F_L(t)\tilde{L}_i(t) + o_p^{\mathcal{T}}(n^{-1/2}). \tag{2.4}$$

Here and in the sequel, $o_p^{\mathcal{S}}(n^{-1/2})$ means root n convergence to 0 in probability uniformly on set $\mathcal{S}$.

We need to pose some mild requirements on $\hat{S}_{T_2|\boldsymbol{Z}=\boldsymbol{z}}(t)$, the estimator of $S_{T_2|\boldsymbol{Z}=\boldsymbol{z}}(t) = P(T_2 > t|\boldsymbol{Z} = \boldsymbol{z})$, to establish the properties of $\hat{G}(y)$ and $\hat{\alpha}(\boldsymbol{z})$. Particularly, we require that $\hat{S}_{T_2|\boldsymbol{Z}=\boldsymbol{z}}(t)$ can be expressed in the form of

$$\hat{S}_{T_2|\boldsymbol{Z}=\boldsymbol{z}}(t) - S_{T_2|\boldsymbol{Z}=\boldsymbol{z}}(t) = \frac{1}{n}\sum_{i=1}^{n}\xi_i(t,\boldsymbol{z}) + \varepsilon_{S_{T_2}}(t,\boldsymbol{z}), \tag{2.5}$$

where $\xi_i(t,\boldsymbol{z})$ are i.i.d. influence functions, and $n^{1/2}\varepsilon_{S_{T_2}}(t,\boldsymbol{z})$ converges to 0 in probability uniformly for $t \in (0,\nu]$ and $||\boldsymbol{z}|| \leq M$. Also, we require that the functional class $\{\xi_i(t,\boldsymbol{z}) : t \in (0,\nu], ||\boldsymbol{z}|| \leq M\}$ is Donsker (Van der Vaart and Wellner, 1996) with mean 0. The specific form of $\xi_i(t,\boldsymbol{z})$, when T_2 is assumed to follow the Cox proportional hazards model, is provided in Section 2.4.4.

Now by (2.4) and (2.5), we can apply Taylor expansion to show

$$\begin{aligned}
\hat{\alpha}(\boldsymbol{z}) - \alpha(\boldsymbol{z}) &= \frac{1}{n}\sum_{i=1}^{n}\int_0^{\nu}\left[\xi_i(u,\boldsymbol{z})\mathrm{d}F_L(u) + S_{T_2|\boldsymbol{Z}=\boldsymbol{z}}(u)\mathrm{d}\{F_L(u)\tilde{L}_i(u)\}\right] + o_p^{\mathcal{Z}}(n^{-1/2}) \\
&\equiv \frac{1}{n}\sum_{i=1}^{n} a_i(\boldsymbol{z}) + o_p^{\mathcal{Z}}(n^{-1/2}).
\end{aligned}$$

It is not hard to see $Ea_i(\boldsymbol{z}) = 0$ from $E\xi_i(t,\boldsymbol{z}) = 0$ and $E\tilde{L}_i(t) = 0$. Following this, we combine Taylor expansion and some algebraic manipulations to show

$$\begin{aligned}
\frac{\hat{\alpha}(\boldsymbol{z})}{\hat{S}_{T_2|\boldsymbol{Z}=\boldsymbol{z}}(y)} - \frac{\alpha(\boldsymbol{z})}{S_{T_2|\boldsymbol{Z}=\boldsymbol{z}}(y)} &= \frac{1}{n}\sum_{i=1}^{n}\Big\{\frac{a_i(\boldsymbol{z})}{S_{T_2|\boldsymbol{Z}=\boldsymbol{z}}(y)} - \frac{\alpha(\boldsymbol{z})\xi_i(y,\boldsymbol{z})}{S^2_{T_2|\boldsymbol{Z}=\boldsymbol{z}}(y)}\Big\} + o_p^{\mathcal{T}\times\mathcal{Z}}(n^{-1/2}) \\
&\equiv \frac{1}{n}\sum_{i=1}^{n}\kappa_i(y,\boldsymbol{z}) + o_p^{\mathcal{T}\times\mathcal{Z}}(n^{-1/2}),
\end{aligned}$$

with $E\kappa_i(y,\boldsymbol{z})=0$. We can further show

$$\begin{aligned}\hat{G}(y)-G(y)=&\frac{1}{n}\sum_{i=1}^{n}\Big\{\frac{I(L_i^*<y\le Y_i^*)\alpha(\boldsymbol{Z}_i^*)}{S_{T_2|\boldsymbol{Z}=\boldsymbol{z}}(Y_i^*)}-G(y)\Big\}\\&+\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\kappa_i(Y_j^*,\boldsymbol{Z}_j^*)}{n}I(L_j^*<y\le Y_j^*)+o_p^{\mathcal{T}}(n^{-1/2})\\=&\frac{1}{n}\sum_{i=1}^{n}g_i(y)+o_p^{\mathcal{T}}(n^{-1/2})\end{aligned}\tag{2.6}$$

where

$$g_i(y)\equiv\Big\{I(L_i^*<y\le Y_i^*)\frac{\alpha(\boldsymbol{Z}_i^*)}{S_{T_2|\boldsymbol{Z}=\boldsymbol{z}}(Y_i^*)}-G(y)\Big\}+E_{\widetilde{\boldsymbol{\omega}}_j^*}\Big\{\kappa_i(Y_j^*,\boldsymbol{Z}_j^*)I(L_j^*<y\le Y_j^*)\Big\},$$

with $\widetilde{\boldsymbol{\omega}}_i^*$ denoting $(L_i^*,X_i^*,Y_i^*,\delta_i^*,\eta_i^*,\boldsymbol{Z}_i^*)$ and $E_{\widetilde{\boldsymbol{\omega}}_j^*}$ representing the expectation over $\widetilde{\boldsymbol{\omega}}_j^*,\ j=1,2,...n$. Noting that $E\big\{I(L_i^*<y\le Y_i^*)\alpha(\boldsymbol{Z}_i^*)/S_{T_2|\boldsymbol{Z}=\boldsymbol{z}}(Y_i^*)\big\}=G(y)$ and

$$E_{\widetilde{\boldsymbol{\omega}}_i^*}\big[E_{\widetilde{\boldsymbol{\omega}}_j^*}\{\kappa_i(Y_j^*,\boldsymbol{Z}_j^*)I(L_j^*<y\le Y_j^*)\}\big]=E_{\widetilde{\boldsymbol{\omega}}_j^*}\big[I(L_j^*<y\le Y_j^*)E_{\widetilde{\boldsymbol{\omega}}_i^*}\{\kappa_i(Y_j^*,\boldsymbol{Z}_j^*)\}\big]=0,$$

we have $Eg_i(y)=0$. We would show later that the functional class $\{g_i(y):y\in\mathcal{T}\}$ is Donsker thus Glivenko-cantelli, therefore $\hat{G}(y)$ is uniformly consistent for $G(y)$ on $\mathcal{T}$.

Combining the above and Taylor expansion, we have

$$\begin{aligned}\hat{\alpha}(\boldsymbol{z})/\hat{G}(y)-\alpha(\boldsymbol{z})/G(y)&=\frac{1}{n}\sum_{i=1}^{n}\left\{\frac{a_i(\boldsymbol{z})}{G(y)}-\frac{\alpha(\boldsymbol{z})g_i(y)}{G(y)^2}\right\}+o_p^{\mathcal{T}\times\mathcal{Z}}(n^{-1/2})\\&\equiv\frac{1}{n}\sum_{i=1}^{n}w_i(y,\boldsymbol{z})+o_p^{\mathcal{T}\times\mathcal{Z}}(n^{-1/2})\end{aligned}\tag{2.7}$$

with $Ew_i(y,\boldsymbol{z})=0$.

Next, we claim that $\{w_i(y,\boldsymbol{z}),\boldsymbol{z}\in\mathcal{Z},y\in\mathcal{T}\}$ form a Donsker class (Van der Vaart and Wellner, 1996). Then, by the functional law of the iterated logarithm (Goodman et al., 1981), (2.7) implies $\sup_{\boldsymbol{z}\in\mathcal{Z},y\in\mathcal{T}}|\hat{\alpha}(\boldsymbol{z})/\hat{G}(y)-\alpha(\boldsymbol{z})/G(y)| = o(n^{-1/2+r})$ for $0 < r < \frac{1}{2}$ and consequently

$$\sup_{\tau,\boldsymbol{b}}||n^{-1/2}\boldsymbol{S}_n(\boldsymbol{b},\tau) - n^{-1/2}\boldsymbol{S}_n^G(\boldsymbol{b},\tau)|| = o(n^{-1/2+r}), a.s. \tag{2.8}$$

To show $\{w_i(y,\boldsymbol{z}), \boldsymbol{z}\in\mathcal{Z}, y\in\mathcal{T}\}$ is Donsker, we first need to prove that $\{a_i(\boldsymbol{z}) : \boldsymbol{z}\in\mathcal{Z}\}$ forms a Donsker class provided $\{\xi_i(t,\boldsymbol{z}) : t\in\mathcal{T}, \boldsymbol{z}\in\mathcal{Z}\}$ is Donsker. We first examine the component, $\frac{1}{\sqrt{n}}\sum_{i=1}^n\int_0^\nu \xi_i(u,\boldsymbol{z})\mathrm{d}F_L(u)$, in $a_i(\boldsymbol{z})$ and show its weak convergence to a tight Gaussian process. It is easy to see $\frac{1}{\sqrt{n}}\sum_{i=1}^n \xi_i(t,\boldsymbol{z}) \rightsquigarrow \varphi(t,\boldsymbol{z})$ according to Donsker's property, where $\varphi(t,\boldsymbol{z})$ is a tight Gaussian process and $\rightsquigarrow$ means converge weakly. Note that

$$\sup_{\boldsymbol{z}\in\mathcal{Z}}|\int_0^\nu\{x(u,\boldsymbol{z})-y(u,\boldsymbol{z})\}\mathrm{d}F_L(u)| \le \sup_{t\in\mathcal{T},\boldsymbol{z}\in\mathcal{Z}}|x(t,\boldsymbol{z})-y(t,\boldsymbol{z})|; \quad x,y\in\ell^\infty(\mathcal{T}\times\mathcal{Z}).$$

Then the map π that maps $x(t,\boldsymbol{z})$ to $\int_0^\nu x(u,\boldsymbol{z})dF_L(u)$ is a continuous map from $\ell^\infty(\mathcal{T}\times\mathcal{Z})$ to $\ell^\infty(\mathcal{Z})$. Therefore $\frac{1}{\sqrt{n}}\sum_{i=1}^n\int_0^\nu \xi_i(u,\boldsymbol{z})dF_L(u) \rightsquigarrow \int_0^\nu \varphi(u,\boldsymbol{z})dF_L(u)$ according to the Continuous Mapping Theorem. Since π is a linear map, $\int_0^\nu \varphi(u,\boldsymbol{z})dF_L(u)$ is a mean zero Gaussian process. The continuity of π further ensures the asymptotic tightness of $\frac{1}{\sqrt{n}}\sum_{i=1}^n\int_0^\nu \xi_i(u,\boldsymbol{z})dF_L(u)$ and also the tightness of $\int_0^\nu \varphi(u,\boldsymbol{z})dF_L(u)$. Similar arguments can be applied to other components of $a_i(\boldsymbol{z})$. Hence, $\{a_i(\boldsymbol{z}) : \boldsymbol{z}\in\mathcal{Z}\}$ is a Donsker class.

By similar arguments and the boundness of $S^{-1}_{T_2|\boldsymbol{Z}=\boldsymbol{z}}(y)$ on $\mathcal{T}\times\mathcal{Z}$, we can show $\{g_i(y), y\in\mathcal{T}\}$ forms a Donsker's class. Since Donsker implies Glivenko-cantelli (Van der Vaart and Wellner, 1996) and $Eg_i(y) = 0$, $\hat{G}(y)$ is uniformly consistent for $G(y)$ on $y\in(0,\nu]$. It follows that $\{a_i(\boldsymbol{z})/G(y) - \alpha(\boldsymbol{z})g_i(y)/G(y)^2, z\in\mathcal{Z}, y\in\mathcal{T}\}$ is also Donsker, because Donsker's property is preserved under Lipschitz transformations, and both $G(y)^{-1}$ and $\alpha(\boldsymbol{z})$ are bounded on $\mathcal{T}\times\mathcal{Z}$. Therefore, we can see (2.8) holds.

Define $\mathcal{F} = \big\{\boldsymbol{Z}_i^*[I\{X_i^* \leq g(\boldsymbol{Z}_i^{*T}b), \delta_i^*\eta_i^* = 1\}G^{-1}(Y_i^*)\alpha(\boldsymbol{Z}_i^*) - \tau],\ \boldsymbol{b} \in R^{p+1}, \tau \in [\tau_L, \tau_U]\big\}$. The function class $\mathcal{F}$ is Donsker and thus Glivenko-Cantelli because the class of indicator functions is Donsker, $\boldsymbol{Z}_i^*$, $\alpha(\boldsymbol{Z}_i^*)$ are uniformly bounded and $G(Y_i^*)$ is uniformly bounded from 0 (Van der Vaart and Wellner, 1996). By the Glivenko-Cantelli Theorem, $\sup_{\tau,\boldsymbol{b}}||n^{-1/2}\boldsymbol{S}_n^G(\boldsymbol{b},\tau) - \boldsymbol{\mu}(\boldsymbol{b},\tau)|| = o(1), a.s.$ and thus $\sup_{\tau,\boldsymbol{b}}||n^{-1/2}\boldsymbol{S}_n(\boldsymbol{b},\tau) - \boldsymbol{\mu}(\boldsymbol{b},\tau)|| = o(1), a.s.$ follows from (A.5). This, coupled with the fact that $\boldsymbol{\mu}\{\boldsymbol{\beta}_0(\tau)\} = 0$ and $n^{-1/2}\ \boldsymbol{S}_n\{\hat{\boldsymbol{\beta}}(\tau), \tau\} = o(1), a.s.$, implies that

$$\sup_{\tau \in [\tau_L, \tau_U]} \|\boldsymbol{\mu}\{\hat{\boldsymbol{\beta}}(\tau), \tau\} - \boldsymbol{\mu}\{\boldsymbol{\beta}_0(\tau)\}\| = o(1),\ a.s. \tag{2.9}$$

Following the same lines of Peng and Fine (2009), we can show that Condition *C4* and the monotonicity of $\boldsymbol{\mu}(\boldsymbol{b},\tau)$ in b imply

$$\inf_{\boldsymbol{b} \notin \mathcal{B}(\rho_0), \tau \in [\tau_L, \tau_U]} ||\boldsymbol{\mu}(\boldsymbol{b},\tau) - \boldsymbol{\mu}\{\boldsymbol{\beta}_0(\tau)\}|| > c_0\rho_0.$$

Consequently, $\{\hat{\boldsymbol{\beta}}(\tau) : \tau \in [\tau_L, \tau_U]\} \subseteq \mathcal{B}(\rho_0)$ for n large enough with probability 1. Applying Taylor expansion to $\boldsymbol{\mu}\{\hat{\boldsymbol{\beta}}(\tau), \tau\}$ around $\boldsymbol{\beta}_0(\tau)$ gives

$$\begin{aligned}\sup_{\tau \in [\tau_L, \tau_U]} ||\hat{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}_0(\tau)|| &= \sup_{\tau \in [\tau_L, \tau_U]} ||\boldsymbol{A}\{\check{\boldsymbol{\beta}}(\tau)\}^{-1}[\boldsymbol{\mu}\{\hat{\boldsymbol{\beta}}(\tau), \tau\} - \boldsymbol{\mu}\{\boldsymbol{\beta}_0(\tau)]|| \\ &\leq c_0^{-1} \sup_{\tau \in [\tau_L, \tau_U]} ||\boldsymbol{\mu}\{\hat{\boldsymbol{\beta}}(\tau), \tau\} - \boldsymbol{\mu}\{\boldsymbol{\beta}_0(\tau)||,\end{aligned}$$

where $\check{\boldsymbol{\beta}}(\tau)$ lies between $\hat{\boldsymbol{\beta}}(\tau)$ and $\boldsymbol{\beta}_0(\tau)$ and is therefore within $\mathcal{B}(\rho_0)$. Note that the last inequality holds by condition *C5*. The uniform convergence of $\hat{\boldsymbol{\beta}}(\tau)$ to $\boldsymbol{\beta}_0(\tau)$ for $\tau \in [\tau_L, \tau_U]$ then follows from (2.9).

2.4.2 Proof of Theorem 2

For simplicity, we write $W_j = W(Y_j^*, \boldsymbol{Z}_j^*)$, $\hat{W}_j = \hat{W}(Y_j^*, \boldsymbol{Z}_j^*)$ and $w_{ij} = a_i(\boldsymbol{Z}_j^*)/G(Y_j^*) - \alpha(\boldsymbol{Z}_j^*)g_i(Y_j^*)/G(Y_j^*)^2$. Let $\approx$ denote asymptotic equivalence unformly in $\tau \in [\tau_L, \tau_U]$.

First, by (2.7), simple algebraic manipulations show that

$$\begin{aligned}
\boldsymbol{S}_n\{\boldsymbol{\beta}_0(\tau), \tau\} &= n^{-1/2}\sum_{j=1}^{n} \boldsymbol{Z}_j^*\big(I[X_j^* \le g\{\boldsymbol{Z}_j^{*T}\boldsymbol{\beta}_0(\tau)\}, \delta_j^*\eta_j^* = 1]\hat{W}_j^{-1} - \tau\big) \\
&= n^{-1/2}\sum_{j=1}^{n} \boldsymbol{Z}_j^* I[X_j^* \le g\{\boldsymbol{Z}_j^{*T}\boldsymbol{\beta}_0(\tau)\}, \delta_j^*\eta_j^* = 1](\hat{W}_j^{-1} - W_j^{-1}) \\
&+ n^{-1/2}\sum_{i=1}^{n} \boldsymbol{Z}_i^*\big(I[X_i^* \le g\{\boldsymbol{Z}_i^{*T}\boldsymbol{\beta}_0(\tau)\}, \delta_i^*\eta_i^* = 1]W_i^{-1} - \tau\big) \\
&\approx n^{-1/2}\sum_{j=1}^{n} \boldsymbol{Z}_j^* I[X_j^* \le g\{\boldsymbol{Z}_j^{*T}\boldsymbol{\beta}_0(\tau)\}, \delta_j^*\eta_j^* = 1]\sum_{i=1}^{n} w_{ij}/n \\
&+ n^{-1/2}\sum_{i=1}^{n} \boldsymbol{Z}_i^*\big(I[X_i^* \le g\{\boldsymbol{Z}_i^{*T}\boldsymbol{\beta}_0(\tau)\}, \delta_i^*\eta_i^* = 1]W_i^{-1} - \tau\big) \\
&\approx n^{-1/2}\sum_{i=1}^{n} \Big\{\boldsymbol{Z}_i^*\big(I[X_i^* \le g\{\boldsymbol{Z}_i^{*T}\boldsymbol{\beta}_0(\tau)\}, \delta_i^*\eta_i^* = 1]W_i^{-1} - \tau\big) \\
&+ \frac{1}{n}\sum_{j=1}^{n} \big(\boldsymbol{Z}_j^* I[X_j^* \le g\{\boldsymbol{Z}_j^{*T}\boldsymbol{\beta}_0(\tau)\}, \delta_j^*\eta_j^* = 1]w_{ij}\big)\Big\}
\end{aligned}$$

An application of the Glivenko-Cantelli Theorem to $\frac{1}{n}\sum_{j=1}^{n}\big(\boldsymbol{Z}_j^* I[X_j^* \le g\{\boldsymbol{Z}_j^{*T}\boldsymbol{\beta}_0(\tau)\}, \delta_j^*\eta_j^* = 1]w_{ij}\big)$ gives

$$\boldsymbol{S}_n\{\boldsymbol{\beta}_0(\tau), \tau\} \approx n^{-1/2}\sum_{i=1}^{n}\{\boldsymbol{\zeta}_{1i}(\tau) + \boldsymbol{\zeta}_{2i}(\tau)\},$$

where $\boldsymbol{\zeta}_{1i}(\tau) = \boldsymbol{Z}_i^*\big(I[X_i^* \le g\{\boldsymbol{Z}_i^{*T}\boldsymbol{\beta}_0(\tau)\}, \delta_i^*\eta_i^* = 1]W_i^{-1} - \tau\big)$ and $\boldsymbol{\zeta}_{2i}(\tau) = E_{\widetilde{\boldsymbol{\omega}}_j^*}\big(\boldsymbol{Z}_j^* I[X_j^* \le g\{\boldsymbol{Z}_j^{*T}\boldsymbol{\beta}_0(\tau)\}, \delta_j^*\eta_j^* = 1]w_{ij}\big)$ with $\widetilde{\boldsymbol{\omega}}_i^*$ denoting $\{L_i^*, X_i^*, Y_i^*, \delta_i^*, \eta_i^*, \boldsymbol{Z}_i^*\}$ and $E_{\widetilde{\boldsymbol{\omega}}_j^*}$ representing the expectation over $\widetilde{\boldsymbol{\omega}}_j^*,\ j = 1, 2, ...n$. Following similar arguments for $a_i(\boldsymbol{z})$ in the proof of Theorem 2.1.1, we can show that $\{\boldsymbol{\zeta}_{1i}(\boldsymbol{z}) + \boldsymbol{\zeta}_{2i}(\boldsymbol{z}),\ \boldsymbol{z} \in \mathcal{Z}\}$ is also a Donsker class. Therefore, $\boldsymbol{S}_n\{\boldsymbol{\beta}_0(\tau), \tau\}$ converges weakly to a mean zero Gaussian Process with covariance matrix $\boldsymbol{\Sigma}(\tau', \tau) = E[\boldsymbol{\zeta}(\tau')\boldsymbol{\zeta}(\tau)]$, where $\boldsymbol{\zeta}_i(\tau) = \boldsymbol{\zeta}_{1i}(\tau) + \boldsymbol{\zeta}_{2i}(\tau)$

$(i = 1, \ldots, n)$.

Next, we establish the asymptotic linearity of $\boldsymbol{S}_n^G(\boldsymbol{b}, \tau)$ in the vicinity of $\boldsymbol{b} = \boldsymbol{\beta}_0(\tau)$; that is, for any positive sequence $\{d_n\}_{n=1}^n$ such that $d_n \to 0$,

$$\sup_{b,b' \in \mathcal{B}(\rho_0); ||b-b'|| \le d_n} ||\{\boldsymbol{S}_n^G(\boldsymbol{b}, \tau) - \boldsymbol{S}_n^G(b', \tau)\} - n^{1/2}\{\boldsymbol{\mu}(\boldsymbol{b}, \tau) - \boldsymbol{\mu}(b', \tau)\}|| = o_p(1). \quad (2.10)$$

The proof for (2.10) greatly resembles the lines of Alexander (1984) and Lai and Ying (1988). The key is to show

$$\begin{aligned} &Var\left\{\boldsymbol{Z}_i^*[I\{X_i^* \le g(\boldsymbol{Z}_i^{*T}\boldsymbol{b}), \delta_i^*\eta_i^* = 1\} - I\{X_i^* \le g(\boldsymbol{Z}_i^{*T}\boldsymbol{b}'), \delta_i^*\eta_i^* = 1\}]W^{-1}(Y_i^*, \boldsymbol{Z}_i^*)\right\} \\ &\le G_0||\boldsymbol{b} - \boldsymbol{b}'||. \end{aligned}$$

This can be verified by using the uniform boundedness of the subdistribution density $f_1(t|\boldsymbol{Z})$, $\boldsymbol{Z}_i$ and $W(y, \boldsymbol{z})$.

It follows from (2.10) that

$$\begin{aligned} & \boldsymbol{S}_n\{\hat{\boldsymbol{\beta}}(\tau), \tau\} - \boldsymbol{S}_n\{\boldsymbol{\beta}_0(\tau), \tau\} \\ = \;& n^{-1/2}\sum_{i=1}^n \boldsymbol{Z}_i^* W_i^{-1} I\big(\delta_i^*\eta_i^* = 1\big)(I[X_i^* \le g\{\boldsymbol{Z}_i^*\hat{\boldsymbol{\beta}}(\tau)\}] - I[X_i^* \le g\{\boldsymbol{Z}_i^*\boldsymbol{\beta}_0(\tau)\}]\big) \\ + \;& n^{-1/2}\sum_{i=1}^n \boldsymbol{Z}_i^* I(\delta_i^*\eta_i^* = 1)(I[X_i^* \le g\{\boldsymbol{Z}_i^*\hat{\boldsymbol{\beta}}(\tau)\}] - I[X_i^* \le g\{\boldsymbol{Z}_i^*\boldsymbol{\beta}_0(\tau)\}])(\hat{W}_i^{-1} - W_i^{-1}) \\ \approx \;& n^{1/2}[\boldsymbol{\mu}\{\hat{\boldsymbol{\beta}}(\tau), \tau\} - \boldsymbol{\mu}\{\boldsymbol{\beta}_0(\tau), \tau\}] \end{aligned}$$

Along with the fact that $\boldsymbol{A}\{\boldsymbol{\beta}_0(\tau)\} = \{\partial\boldsymbol{\mu}(\boldsymbol{b}, \tau)/\partial\boldsymbol{b}\}|_{\boldsymbol{b}=\boldsymbol{\beta}_0(\tau)}$ and $\hat{\boldsymbol{\beta}}(\tau)$ uniformly converges to $\boldsymbol{\beta}_0(\tau)$, a Taylor expansion of $\boldsymbol{\mu}(\boldsymbol{b}, \tau)$ around $\boldsymbol{b} = \boldsymbol{\beta}_0(\tau)$ gives that

$$\boldsymbol{S}_n\{\hat{\boldsymbol{\beta}}(\tau), \tau\} - \boldsymbol{S}_n\{\boldsymbol{\beta}_0(\tau), \tau\} \approx \boldsymbol{A}\{\boldsymbol{\beta}_0(\tau)\} n^{1/2}\{\hat{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}_0(\tau)\}.$$

This implies

$$n^{1/2}\{\hat{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}_0(\tau)\} \approx -\boldsymbol{A}\{\boldsymbol{\beta}_0(\tau)\}^{-1}\boldsymbol{S}_n\{\boldsymbol{\beta}_0(\tau), \tau\}, \tag{2.11}$$

and then $n^{1/2}\{\hat{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}_0(\tau)\}$ converges weakly to a mean zero Gaussian process with covariance matrix

$$\boldsymbol{A}\{\boldsymbol{\beta}_0(\tau')\}^{-1}\boldsymbol{\Sigma}(\tau', \tau)\boldsymbol{A}\{\boldsymbol{\beta}_0(\tau)\}^{-T}.$$

2.4.3 Justification for the Proposed Covariance Matrix Estimate

With the strict convexity condition in *C4*, it is implied from the proof of Theorem 2.1.1 that $\left\{\boldsymbol{S}_n^{-1}\{\boldsymbol{e}_{n,j}(\tau), \tau\}, \tau \in [\tau_L, \tau_U]\right\}$ falls in $\mathcal{B}(\rho_0)$ with probability 1 when n is large enough, and uniformly converges to $\boldsymbol{\beta}_0(\tau),\ j = 1, 2, ..p + 1$. Denote $\boldsymbol{b}_{n,j}(\tau) = \boldsymbol{S}_n^{-1}\{\boldsymbol{e}_{n,j}(\tau), \tau\}$. Using arguments similar to those for (A.7), we can show that

$$\boldsymbol{S}_n\{\boldsymbol{b}_{n,j}(\tau), \tau\} - \boldsymbol{S}_n\{\hat{\boldsymbol{\beta}}(\tau), \tau\} \approx \boldsymbol{A}\{\boldsymbol{\beta}_0(\tau)\}n^{1/2}\left[\boldsymbol{b}_{n,j}(\tau) - \hat{\boldsymbol{\beta}}(\tau)\right].$$

By the definition of $\boldsymbol{D}_n(\tau)$ and $\boldsymbol{E}_n(\tau)$, this implies $\boldsymbol{E}_n(\tau) \approx \sqrt{n}\boldsymbol{A}\{\boldsymbol{\beta}_0(\tau)\}\boldsymbol{D}_n(\tau)$. Thus $\sqrt{n}\boldsymbol{D}_n(\tau)\boldsymbol{E}_n^{-1}(\tau)$ is a consistent estimator for $A^{-1}\{\boldsymbol{\beta}_0(\tau)\}$. It follows immediately that

$$n\boldsymbol{D}_n(\tau')\boldsymbol{E}_n^{-1}(\tau')\hat{\boldsymbol{\Sigma}}(\tau', \tau)\boldsymbol{E}_n(\tau)^{-1}\boldsymbol{D}_n(\tau)^T$$

is a consistent estimator for the asymptotic covariance matrix of $n^{1/2}\{\hat{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}_0(\tau)\}$.

2.4.4 The Form of $\xi_i(t, \boldsymbol{z})$ under the Cox Proportional Hazard Model

Here we present the form of $\xi_i(t, \boldsymbol{z})$ when the Cox proportional hazards model is assumed for T_2. In that case, $P(T_2 > t|\boldsymbol{Z}) = \exp\{-\Lambda_0(t)\exp(\boldsymbol{\gamma}_0^T\widetilde{\boldsymbol{Z}})\}$, where $\boldsymbol{\gamma}_0$ is the $p \times 1$ regression coefficient, and $\Lambda_0(t)$ is the baseline cumulative hazard function.

Define $N_i(t) = I(Y_i^* \le t)\eta_i^*$, $M_i(t) = N_i(t) - \Lambda_0(t)\exp(\boldsymbol{\gamma}_0^T \widetilde{\boldsymbol{Z}}_i^*)$, $R_i(t) = I(L_i^* < t \le Y_i^*)$, $\boldsymbol{S}^{(j)}(\boldsymbol{\gamma}, t) = 1/n \sum_{i=1}^n (\widetilde{\boldsymbol{Z}}_i^*)^{\otimes j} \times R_i(t) \exp\{\boldsymbol{\gamma}^T \widetilde{\boldsymbol{Z}}_i^*\}$, where $j = 0, 1, 2$. Let $\boldsymbol{E}(\boldsymbol{\gamma}, t) = \boldsymbol{S}^{(1)}(\boldsymbol{\gamma}, t)/S^{(0)}(\boldsymbol{\gamma}, t)$, $\boldsymbol{V}(\boldsymbol{\gamma}, t) = \dfrac{\boldsymbol{S}^{(2)}(\boldsymbol{\gamma}, t)}{S^{(0)}(\boldsymbol{\gamma}, t)} - \boldsymbol{E}(\boldsymbol{\gamma}, t)^{\otimes 2}$, $\boldsymbol{G} = \int_0^\nu \boldsymbol{V}(\boldsymbol{\gamma}_0, t) S^{(0)}(\boldsymbol{\gamma}_0, t) \mathrm{d}\Lambda_0(t)$, and $\boldsymbol{P}(t) = -\int_0^t \boldsymbol{E}(\boldsymbol{\gamma}_0, u)\mathrm{d}\Lambda_0(u)$. Let $\boldsymbol{\eta}_0(t) = \{\Lambda_0(t), \boldsymbol{\gamma}_0\}^T$. Adapting Andersen and Gill (1982)'s results, we can show

$$
\begin{aligned}
& \hat{\boldsymbol{\eta}}(t) - \boldsymbol{\eta}(t) \\
= & \begin{pmatrix} \frac{1}{n}\sum_{i=1}^n [\boldsymbol{P}(t)^T \int_0^\nu \boldsymbol{G}^{-1}\{\widetilde{\boldsymbol{Z}}_i^* - \boldsymbol{E}(\boldsymbol{\gamma}_0, u)\}\mathrm{d}M_i(u) + \int_0^t \dfrac{1}{S^{(0)}(\boldsymbol{\gamma}_0, u)}\mathrm{d}M_i(u)] \\ \frac{1}{n}\sum_{i=1}^n \int_0^\nu \boldsymbol{G}^{-1}\{\widetilde{\boldsymbol{Z}}_i^* - \boldsymbol{E}(\boldsymbol{\gamma}_0, u)\}\mathrm{d}M_i(u)\} \end{pmatrix} + o_p^{\mathcal{T}}(n^{-1/2}) \\
\equiv & \frac{1}{n}\sum_{i=1}^n \{p_i^{(1)}(t), \boldsymbol{p}_i^{(2)}\}^T + o_p^{\mathcal{T}}(n^{-1/2}),
\end{aligned}
$$

and $\left\{\left(p_i^{(1)}(t), \boldsymbol{p}_i^{(2)}\right)^T, t \in \mathcal{T}\right\}$ forms a Donsker's class.

Applying Taylor expansions gives that

$$
\hat{S}_{T_2|\boldsymbol{Z}=\boldsymbol{z}}(t) - S_{T_2|\boldsymbol{Z}=\boldsymbol{z}}(t) \quad = \quad -\frac{1}{n}\sum_{i=1}^n \exp({\boldsymbol{\gamma}_0}^T z) S_{T_2|\boldsymbol{Z}=\boldsymbol{z}}(t)\{p_i^{(1)}(t) + \Lambda_0(t){\boldsymbol{p}_i^{(2)}}^T \boldsymbol{z}\} + o_p^{\mathcal{T}\times\mathcal{Z}}(n^{-1/2}).
$$

Therefore, it is easy to see that the influence function of $\hat{S}_{T_2|\boldsymbol{Z}=\boldsymbol{z}}(t)$ is given by

$$
\xi_i(t, \boldsymbol{z}) \equiv -\exp({\boldsymbol{\gamma}_0}^T \boldsymbol{z}) S_{T_2|\boldsymbol{Z}=\boldsymbol{z}}(t)\{p_i^{(1)}(t) + \Lambda_0(t){\boldsymbol{p}_i^{(2)}}^T \boldsymbol{z}\}.
$$

We can show $\{\xi_i(t, \boldsymbol{z}), t \in \mathcal{T}, \boldsymbol{z} \in \mathcal{Z}\}$ is Donsker following similar arguments for $a_i(\boldsymbol{z})$ in 2.4.1.

Chapter 3

Quantile Regression adjusting for Dependent Censoring

3.1 Regression Procedures

3.1.1 Data and Model

We start with a formal introduction of data and notations. Let T_1 be time to the primary endpoint and T_2 be time to dependent censoring. Let $\boldsymbol{Z} = (1, \widetilde{\boldsymbol{Z}}^T)^T$ be the $(p+1) \times 1$ covariate vector. There may also be an administrative censoring time C that is conditionally independent of (T_1, T_2) given $\boldsymbol{Z}$. Define $X = T_1 \wedge T_2 \wedge C$, $Y = T_2 \wedge C$, $\delta = I(T_1 \leq T_2 \wedge C)$, $\eta = I(T_2 \leq C)$, where $\wedge$ is the minimum operator and $I(\cdot)$ is the indicator function. Observed data includes n identically and independently distributed replicates of $\{X, Y, \delta, \eta, \boldsymbol{Z}\}$, denoted by $\{X_i, Y_i, \delta_i, \eta_i, \boldsymbol{Z}_i\}_{i=1}^n$.

Define the τth conditional quantile of an event time T_1 as $Q_{T_1}(\tau|\boldsymbol{Z}) = \inf\{t : \Pr(T_1 \leq t|\boldsymbol{Z}) \geq \tau\}$, we assume that

$$Q_{T_1}(\tau|\boldsymbol{Z}) = \exp\{\boldsymbol{Z}^T \boldsymbol{\beta}_0(\tau)\}, \quad 0 < \tau < 1, \tag{3.1}$$

where $\boldsymbol{\beta}_0(\tau)$ is a vector of unknown regression coefficient representing covariate effects at the τth quantile of the event time.

To estimate model (3.1), it is important to recognize that dependent censoring by T_2 induces the nonparametric nonidentifiability of the marginal distribution of T_1 (Tsiatis, 1975). This necessitates additional assumptions on the dependence structure between T_1 and T_2 because model (3.1) concerns the marginal quantiles of T_1. Such a strategy has been widely adopted in previous work on dependent censoring, which either restricts the joint distribution using semiparametric or parametric models (Link 1989; Emoto and Matthews 1990) or performs a sensitivity analysis (Peterson 1976; Slud and Rubinstein 1983; Klein and Moeschberger 1988; Zheng and Klein 1995; Scharfstein and Robins 2002; among others.). Following a similar idea, we assume the association structure between T_1 and T_2 follows a copula model:

$$\Pr(T_1 > s, T_2 > t|\boldsymbol{Z}) = \Psi\big\{1 - F_1(s|\boldsymbol{Z}), 1 - F_2(t|\boldsymbol{Z}), g(\bar{\boldsymbol{Z}}^T \boldsymbol{r}_0)\big\}, \tag{3.2}$$

where $F_i(t|\boldsymbol{Z}) = \Pr(T_i \leq t|\boldsymbol{Z})$ is the distribution function of T_i $(i = 1, 2)$, and $\bar{\boldsymbol{Z}} = (1, \widetilde{\boldsymbol{Z}}_0)$, with $\widetilde{\boldsymbol{Z}}_0$ being a sub-vector of $\widetilde{\boldsymbol{Z}}$, or $\widetilde{\boldsymbol{Z}}$ itself. Here $\Psi(\cdot)$ is a known copula function, for example, Clayton's copula (Clayton, 1978) or Frank's copula (Genest, 1987). The copula parameter, here specified as $g(\bar{\boldsymbol{Z}}^T \boldsymbol{r}_0)$, is often closely connected to a measure that characterizes the dependence strength between T_1 and T_2. For example, under the Clayton's copula model where $\Psi(u, v, \theta) = (u^{1-\theta} + v^{1-\theta} - 1)^{1/(1-\theta)}$, $(\theta - 1)/(\theta + 1)$ equals the Kendall's tau coefficient (Kendall and Gibbons, 1962). In model (3.2), $\boldsymbol{r}_0$ is an unknown parameter used to characterize the dependent censoring mechanism that may vary according to $\bar{\boldsymbol{Z}}$. When $\bar{\boldsymbol{Z}} = 1$, the association between T_1 and T_2 is assumed to be homogeneous for all subjects.

While model (3.2) helps identify the marginal quantiles of T_1, the estimation of model (3.1) is further facilitated by the fact that T_2 is only subject to independent censoring by C and therefore standard censored regression techniques can be applied to estimate $F_2(t|\boldsymbol{Z})$. While in theory the regression model for $F_2(t|\boldsymbol{Z})$ may be specified arbitrarily, it is natural to assume the form of the model for T_2 is the same as that for T_1. That is,

$$Q_{T_2}(\tau|\boldsymbol{Z}) = \exp\{\boldsymbol{Z}^T \boldsymbol{\alpha}_0(\tau)\}, \ \tau \in (0, 1), \tag{3.3}$$

where $\boldsymbol{\alpha}_0(\tau)$ is a $(p + 1) \times 1$ vector of regression coefficient, which can be estimated by using Peng and Huang (2008)'s approach. Denote the resulting estimator as $\widehat{\boldsymbol{\alpha}}(\cdot)$. Without further mentioning, models (3.1)–(3.3) are assumed for the estimation and inferences presented in the sequel.

3.1.2 Estimating Equations

Our proposal for estimating $\boldsymbol{\beta}_0(\tau)$ and $\boldsymbol{r}_0$ is based on the following two equalities:

$$\Pr(X > t|Y > t, \boldsymbol{Z}) = \frac{\Pr(T_1 > t, T_2 > t|\boldsymbol{Z})}{\Pr(T_2 > t|\boldsymbol{Z})} = K_A\{F_1(t|\boldsymbol{Z}), F_2(t|\boldsymbol{Z}), g(\bar{\boldsymbol{Z}}^T\boldsymbol{r}_0)\}, \tag{3.4}$$

and

$$\Pr(X \leq s|Y > t, \boldsymbol{Z}) = \frac{\Pr(T_1 \leq s, T_2 > t|\boldsymbol{Z})}{\Pr(T_2 > t|\boldsymbol{Z})} = K_B\{F_1(s|\boldsymbol{Z}), F_2(t|\boldsymbol{Z}), g(\bar{\boldsymbol{Z}}^T\boldsymbol{r}_0)\} \tag{3.5}$$

for $s \leq t$, where $K_A(u, v, \theta) = \Psi(1-u, 1-v, \theta)/(1-v)$ and $K_B(u, v, \theta) = \{1 - v - \Psi(1-u, 1-v, \theta)\}/(1-v)$. We can cancel out $\Pr(C > t|\boldsymbol{Z})$ in both (3.4) and (3.5) by conditioning on $I(Y > t)$, basing on the conditional independence between C and (T_1, T_2). Note that (3.4) stems from the assumed joint distribution of (T_1, T_2) on the diagonal line, while (3.5) is derived from the information on the upper wedge of (T_1, T_2), which is uniquely available in the semi-competing setting.

To utilize the above facts to construct estimating equations for $\boldsymbol{\beta}_0(\tau)$ and $\boldsymbol{r}_0$, we need to further bridge $\boldsymbol{\alpha}_0(\cdot)$ and $\boldsymbol{\beta}_0(\cdot)$ with the distribution functions of T_1 and T_2. This may be done as follows:

$$F_2(t|\boldsymbol{Z}) \wedge \tau = \int_0^\tau I\big\{F_2(t|\boldsymbol{Z}) \geq u\big\}\mathrm{d}u = \int_0^\tau I\big[t \geq \exp\{\boldsymbol{Z}^T\boldsymbol{\alpha}_0(u)\}\big]\mathrm{d}u, \tag{3.6}$$

and $F_1[\exp\{\boldsymbol{Z}^T\boldsymbol{\beta}_0(\tau)\}|\boldsymbol{Z}] = \tau$ for any $\tau \in (0, 1)$.

Motivated by (3.4), (3.5) and (3.6), we consider two estimating functions taking the following forms:

$$\boldsymbol{S}_n\{\boldsymbol{\beta}, \widehat{\boldsymbol{\alpha}}, \boldsymbol{r}, \tau\} = n^{-1}\sum_{i=1}^n \boldsymbol{Z}_i P_i\{\boldsymbol{\beta}, \widehat{\boldsymbol{\alpha}}, \boldsymbol{r}, \tau\},\ \boldsymbol{W}_n\{\boldsymbol{\beta}, \widehat{\boldsymbol{\alpha}}, \boldsymbol{r}, \tau\} = n^{-1}\sum_{i=1}^n \bar{\boldsymbol{Z}}_i Q_i\{\boldsymbol{\beta}, \widehat{\boldsymbol{\alpha}}, \boldsymbol{r}, \tau\},$$

where

$$P_i\{\boldsymbol{\beta},\boldsymbol{\alpha},\boldsymbol{r},\tau\} = I\{\log X_i > \boldsymbol{Z}_i^T\boldsymbol{\beta}(\tau)\} - I\{\log Y_i > \boldsymbol{Z}_i^T\boldsymbol{\beta}(\tau)\}$$
$$\times K_A\{\tau, \int_0^{\tau_{U,2}} I\{\boldsymbol{Z}_i^T\boldsymbol{\beta}(\tau) \geq \boldsymbol{Z}_i^T\boldsymbol{\alpha}(u)\}\mathrm{d}u, g(\bar{\boldsymbol{Z}}_i^T\boldsymbol{r})\},$$
$$Q_i\{\boldsymbol{\beta},\boldsymbol{\alpha},\boldsymbol{r},\tau\} = I\big[Y_i > 2\exp\{\boldsymbol{Z}_i^T\boldsymbol{\beta}(\tau)\}\big] \times \Bigg(I\big[X_i \leq \exp\{\boldsymbol{Z}_i^T\boldsymbol{\beta}(\tau)\}\big]$$
$$- K_B\big\{\tau, \int_0^{\tau_{U,2}} I\big[2\exp\{\boldsymbol{Z}_i^T\boldsymbol{\beta}(\tau)\} \geq \exp\{\boldsymbol{Z}_i^T\boldsymbol{\alpha}(u)\}\big]\mathrm{d}u, g(\bar{\boldsymbol{Z}}_i^T\boldsymbol{r})\big\}\Bigg).$$

Note that we restrict our attention to $\boldsymbol{\alpha}_0(\tau)$ with $\tau \in (0, \tau_{U,2}]$, where $\tau_{U,2}$ is a constant less than 1 subject to certain identifiability constraints, as $\boldsymbol{\alpha}_0(\tau)$ may not be identifiable for all $\tau \in (0,1)$ due to the censoring by C (Peng and Huang, 2008). Similarly, we restrict our attention to $\boldsymbol{\beta}_0(\tau)$ with $\tau \in (0, \tau_{U,1}]$. where $\tau_{U,1} < 1$ is a constant satisfying mild identifiability conditions, as stated in regularity condition C5 in Chapter 3.4.1.

We propose to estimate $\boldsymbol{\beta}_0(\tau)$ and $\boldsymbol{r}_0$ by solving the following estimating equations:

$$n^{1/2}\boldsymbol{S}_n\{\boldsymbol{\beta},\widehat{\boldsymbol{\alpha}},\boldsymbol{r},\tau\} = 0, \; n^{1/2}\int_{\tau_a}^{\tau_b} \boldsymbol{W}_n\{\boldsymbol{\beta},\widehat{\boldsymbol{\alpha}},\boldsymbol{r},\tau\}\mathrm{d}\tau = 0 \tag{3.7}$$

where τ_a and τ_b are prespecified constant in $(0, \tau_{U,1}]$. We integrate $\boldsymbol{W}_n\{\boldsymbol{\beta},\widehat{\boldsymbol{\alpha}},\boldsymbol{r},\tau\}$ over $\tau \in [\tau_a, \tau_b)$ to make a better use of semi-competing risks information, which would lead to increased estimation efficiency and improved numerical stability. Under the assumed models (3.1)–(3.3), it is easy to verify that $E\boldsymbol{S}_n\{\boldsymbol{\beta}_0(\tau),\boldsymbol{\alpha}_0,\boldsymbol{r}_0,\tau\} = 0$ for any $\tau \in (0, \tau_{U,1}]$, and $E\big[\int_{\tau_a}^{\tau_b} \boldsymbol{W}_n\{\boldsymbol{\beta}_0(\tau),\boldsymbol{\alpha}_0,\boldsymbol{r}_0,\tau\}\mathrm{d}\tau\big] = 0$.

In practice, $\tau_{U,j}$, $j = 1, 2$, may be selected in an adaptive manner as suggested by Peng and Huang (2008) for randomly censored data. To accommodate the identifiability consideration, one may further "truncate" equations (3.4) and (3.5) by restricting the range of t, and solving

$$n^{1/2}\boldsymbol{S}_n^*\{\boldsymbol{\beta},\widehat{\boldsymbol{\alpha}},\boldsymbol{r},\tau\}=0,\ n^{1/2}\int_{\tau_a}^{\tau_b}\boldsymbol{W}_n^*\{\boldsymbol{\beta},\widehat{\boldsymbol{\alpha}},\boldsymbol{r},\tau\}\mathrm{d}\tau=0 \tag{3.8}$$

instead of (3.7), where

$$\begin{aligned}\boldsymbol{S}_n^*\{\boldsymbol{\beta},\boldsymbol{\alpha},\boldsymbol{r},\tau\}=&n^{-1}\sum_{i=1}^n\boldsymbol{Z}_iI\{\boldsymbol{Z}_i^T\boldsymbol{\beta}(\tau)\le\boldsymbol{Z}_i^T\boldsymbol{\alpha}(\tau_{U,2})\}P_i\{\boldsymbol{\beta},\boldsymbol{\alpha},\boldsymbol{r},\tau\},\\ \boldsymbol{W}_n^*\{\boldsymbol{\beta},\boldsymbol{\alpha},\boldsymbol{r},\tau\}=&n^{-1}\sum_{i=1}^n\bar{\boldsymbol{Z}}_iI[2\exp\{\boldsymbol{Z}_i^T\boldsymbol{\beta}(\tau)\}\le\exp\{\boldsymbol{Z}_i^T\boldsymbol{\alpha}(\tau_{U,2})\}]Q_i\{\boldsymbol{\beta},\boldsymbol{\alpha},\boldsymbol{r},\tau\}.\end{aligned}$$

The inclusion of the two indicator functions ensures $\boldsymbol{S}_n^*\{\boldsymbol{\beta}_0,\boldsymbol{\alpha}_0,\boldsymbol{r}_0,\tau\}$ and $\boldsymbol{W}_n^*\{\boldsymbol{\beta}_0,\boldsymbol{\alpha}_0,\boldsymbol{r}_0,\tau\}$ to be mean 0 for any $\tau\in(0,1)$, with little influence on the asymptotic results. The estimating equations in our numerical studies is given by (3.8).

Here and in the sequel we may write $\boldsymbol{\beta}$ as a shorthand for $\boldsymbol{\beta}(\tau)$, and $\boldsymbol{\alpha}$ for $\boldsymbol{\alpha}(\tau)$, with the understanding that these coefficients are τ-specific. Also, we would abbreviate $\int_{\tau_a}^{\tau_b}\boldsymbol{W}_n\{\boldsymbol{\beta},\widehat{\boldsymbol{\alpha}},\boldsymbol{r},\tau\}\mathrm{d}\tau$ as $\boldsymbol{W}_n\{\boldsymbol{\beta},\widehat{\boldsymbol{\alpha}},\boldsymbol{r}\}$, and $\int_{\tau_a}^{\tau_b}\boldsymbol{W}_n^*\{\boldsymbol{\beta},\widehat{\boldsymbol{\alpha}},\boldsymbol{r},\tau\}\mathrm{d}\tau$ as $\boldsymbol{W}_n^*\{\boldsymbol{\beta},\widehat{\boldsymbol{\alpha}},\boldsymbol{r}\}$.

3.1.3 Computational Algorithm

We developed an efficient iterative algorithm to jointly estimate $\boldsymbol{\beta}_0$ and $\boldsymbol{r}_0$ based on (3.8). Let $\mathcal{G}_{L(n)}=\{0<\tau_0<\tau_1<...<\tau_{L(n)}=\tau_{U,1}<1\}$ be a prespecified grid on $[0,1]$, and $||\mathcal{G}_{L(n)}||=\max\{\tau_j-\tau_{j-1};\ j=1,2,...,L(n)\}$ be the size of the grid. For any given $\boldsymbol{r}$, let $\widehat{\boldsymbol{\beta}}(\tau;\boldsymbol{r})$, with a shorthand $\widehat{\boldsymbol{\beta}}(\boldsymbol{r})$, denote a right-continuous step function of τ which jumps only on $\mathcal{G}_{L(n)}$ and satisfies $n^{1/2}\boldsymbol{S}_n^*\{\widehat{\boldsymbol{\beta}}(\boldsymbol{r}),\widehat{\boldsymbol{\alpha}},\boldsymbol{r},\tau\}=o(1)$ for $\tau\in\mathcal{G}_{L(n)}$. The details of the algorithm are described as follows.

Step A Estimate $\left\{\widehat{\boldsymbol{\alpha}}(\tau),\ \tau\in(0,\tau_{U,2}]\right\}$ using Peng and Huang (2008)'s method.

Step B Set $k=0$ and choose an initial value for $\widehat{\boldsymbol{\beta}}$, denoted by $\widehat{\boldsymbol{\beta}}^{[k]}$.

Step C Solve $\widehat{\boldsymbol{r}}^{[k]}$ from $n^{1/2}\boldsymbol{W}_n^*\{\widehat{\boldsymbol{\beta}}^{[k]},\widehat{\boldsymbol{\alpha}},\boldsymbol{r}\}=0$.

Step D Update $\widehat{\boldsymbol{\beta}}^{[k]}$ with $\widehat{\boldsymbol{\beta}}^{[k+1]} = \widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{r}}^{[k]})$. Then increase k by 1 and go to Step C until a certain convergence criteria is satisfied.

In practice, $\widehat{\boldsymbol{\beta}}^{[0]}$ can be set to be the naive estimate, acquired by treating (X, δ) as independently censored survival data and employing Peng and Huang (2008)'s method.

Since $n^{1/2}\boldsymbol{W}_n^*\{\widehat{\boldsymbol{\beta}}^{[k]}, \widehat{\boldsymbol{\alpha}}, \boldsymbol{r}\}$ is a smooth function of $\boldsymbol{r}$, the root-finding in Step C can be easily implemented with existing statistical functionalities, like the *optim* function in R. To solve $n^{1/2}\boldsymbol{S}_n^*\{\boldsymbol{b}, \widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{r}}^{[k]}, \tau\} = 0$ in Step D, we propose to obtain $\widehat{\boldsymbol{\beta}}(\tau; \boldsymbol{r}^{[k]})$ through the following iterative procedure:

D.0 Set $m = 0$ and let $\widehat{\boldsymbol{\beta}}^{[k+1,m]}(\tau; \widehat{\boldsymbol{r}}^{[k]}) = \widehat{\boldsymbol{\beta}}^{[k]}(\tau; \widehat{\boldsymbol{r}}^{[k]})$.

D.1 Find $\widehat{\boldsymbol{\beta}}^{[k+1,m+1]}(\tau)$ by solving

$$n^{-1/2}\sum_{i=1}^{n} \boldsymbol{Z}_i \widehat{B}_i^{[m]}\{I(\log X_i > \boldsymbol{Z}_i^T\boldsymbol{b}) - \widehat{A}_i^{[m]}(\tau)\} = 0, \tag{3.9}$$

where $\widehat{B}_i^{[m]}(\tau) = I\{\boldsymbol{Z}_i^T\widehat{\boldsymbol{\beta}}^{[k+1,m]}(\tau) \leq \boldsymbol{Z}_i^T\widehat{\boldsymbol{\alpha}}(\tau_{U,2})\}$ and $\widehat{A}_i^{[m]}(\tau) = I\{\log Y_i > \boldsymbol{Z}_i^T\widehat{\boldsymbol{\beta}}^{[k+1,m]}(\tau)\}K_A\big[\tau, \int_0^{\tau_{U,2}} I\{\boldsymbol{Z}_i^T\widehat{\boldsymbol{\beta}}^{[k+1,m]}(\tau) \geq \boldsymbol{Z}_i^T\widehat{\boldsymbol{\alpha}}(u)\}\mathrm{d}u, g(\bar{\boldsymbol{Z}}_i^T\widehat{\boldsymbol{r}}^{[k]})\big]$.

D.2 Increase m by 1 and go to D.1 until a certain convergence criteria is satisfied.

It can be shown that the estimating function in (3.9) is a monotone random field in $\boldsymbol{b}$ (Fygenson and Ritov, 1994), and equals the derivative of the following L_1-type function

$$-\frac{n^{-1/2}}{2}\Bigg[\sum_{i=1}^{n}\widehat{B}_i^{[m]}(\tau)\big|\log(X_i) - \boldsymbol{Z}_i^T\boldsymbol{b}\big| + \Big|M - \sum_{i=1}^{n}\widehat{B}_i^{[m]}(\tau)\boldsymbol{Z}_i^T\boldsymbol{b}\Big| + \Big|M + 2\sum_{i=1}^{n}\widehat{B}_i^{[m]}(\tau)\widehat{A}_i^{[m]}(\tau)\boldsymbol{Z}_i^T\boldsymbol{b}\Big|\Bigg], \tag{3.10}$$

where M is an extremely large number that can bound $|\sum_{i=1}^{n}\widehat{B}_i^{[m]}(\tau)\boldsymbol{Z}_i^T\boldsymbol{b}|$ and $|2\sum_{i=1}^{n}\widehat{B}_i^{[m]}(\tau)\widehat{A}_i^{[m]}(\tau)\boldsymbol{Z}_i^T\boldsymbol{b}|$. Such an L_1–type minimization problem can be readily implemented in the *rq* function in R or the *l1fit* function in Splus. In Section **??**, we show that the proposed algorithm produce fast and stable implementation of the proposed estimation method. More details of the algorithm, including convergence criteria adopted for iterations, are provided in Chapter 3.5.

3.1.4 Asymptotic Results

In this section we outline the asymptotic properties of the proposed estimators. In the below, Theorem 3.1.1 states the consistency of $\widehat{\boldsymbol{r}}$ and the uniform consistency of $\widehat{\boldsymbol{\beta}}(\tau)$, and Theorem 3.1.2 gives the results on the limiting distribution of $n^{1/2}(\widehat{\boldsymbol{r}} - \boldsymbol{r}_0)$ and $n^{1/2}\{\widehat{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}_0(\tau)\}$. In order to establish the two theorems, we need to require regularity conditions C1-C5, the details of which are deferred to Chapter 3.4.1.

We have the following theorems.

Theorem 3.1.1. *Under conditions C1-C5, if* $\lim_{n\to\infty} ||\mathcal{G}_{L(n)}|| = 0$*, then there exists a* $\{\widehat{\boldsymbol{\beta}}(\tau), \widehat{\boldsymbol{r}}\}$ *in a neighborhood of* $\{\boldsymbol{\beta}_0(\tau), \boldsymbol{r}_0\}$*, such that* $\sup_{\tau\in[\nu_1,\tau_{U,1}]} ||\widehat{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}_0(\tau)|| \xrightarrow{p} 0$*, for any* $0 < \nu_1 < \tau_{U,1}$ *and* $\widehat{\boldsymbol{r}} \xrightarrow{p} \boldsymbol{r}_0$*.*

Theorem 3.1.2. *Under condition C1-C5, if* $\lim_{n\to\infty} n^{1/2}||\mathcal{G}_{L(n)}|| = 0$*, then there exists a* $\{\widehat{\boldsymbol{\beta}}(\tau), \widehat{\boldsymbol{r}}\}$ *in a neighborhood of* $\{\boldsymbol{\beta}_0(\tau), \boldsymbol{r}_0\}$*, such that* $n^{1/2}(\widehat{\boldsymbol{r}} - \boldsymbol{r}_0)$ *converges in distribution to a Normal distribution with mean 0. Moreover,* $n^{1/2}\{\widehat{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}_0(\tau)\}$ *converges weakly to a mean 0 Gaussian process for* $\tau \in [\nu_1, \tau_{U,1}]$*, where* $0 < \nu_1 < \tau_{U,1}$*.*

The asymptotic arguments are heavily challenged by the need to remove the effect of dependent censoring. To show the consistency in Theorem 3.1.1, we first notice that $\int_0^{\tau_{U,2}} I[t \geq \exp\{\boldsymbol{z}^T \widehat{\boldsymbol{\alpha}}(u)\}]\mathrm{d}u$ serves as a consistent estimate of $F_2(t|\boldsymbol{z}) \wedge \tau_{U,2}$. Following that, we can employ empirical process techniques to show $\sup_{\tau\in(0,\tau_{U,1}]} \boldsymbol{s}\{\widehat{\boldsymbol{\beta}}(\boldsymbol{r}), \boldsymbol{\alpha}_0, \boldsymbol{r}, \tau\} = o_p(1)$. This implies the uniform consistency of $\widehat{\boldsymbol{\beta}}(\tau; \boldsymbol{r})$ to $\widetilde{\boldsymbol{\beta}}(\tau; \boldsymbol{r})$ for any fixed $\boldsymbol{r} \in \mathcal{R}(d_R)$, and furthermore the consistency of $\widehat{\boldsymbol{r}}$ to $\boldsymbol{r}_0$. Coupled with the result that $\widetilde{\boldsymbol{\beta}}(\tau; \boldsymbol{r})$ has bounded derivative against $\boldsymbol{r}$ at $\boldsymbol{r} = \boldsymbol{r}_0$, these give the uniform consistency of $\widehat{\boldsymbol{\beta}}(\tau) = \widehat{\boldsymbol{\beta}}(\tau; \widehat{\boldsymbol{r}})$.

Establishment of the asymptotic normality in Theorem 3.1.2 is not only challenged by the non-smoothness of the estimating equations, but also by the inexplicit relationship between $\widehat{\boldsymbol{\beta}}(\tau) = \widehat{\boldsymbol{\beta}}(\tau; \widehat{\boldsymbol{r}})$ and $\widehat{\boldsymbol{r}}$. There are two key results in our proofs. Firstly, we show the asymptotic linearity of $n^{1/2}\{\widehat{\boldsymbol{\beta}}(\tau; \widehat{\boldsymbol{r}}) - \widehat{\boldsymbol{\beta}}(\tau; \boldsymbol{r}_0)\}$ with regard to $n^{1/2}\{\widehat{\boldsymbol{r}} - \boldsymbol{r}_0\}$ by mimicking the arguments in Lai and Ying (1988) and Peng and Huang (2008). Secondly, we used extensive empirical process and integral estimating equation techniques to show that $n^{1/2}[\boldsymbol{W}_n\{\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{r}}), \widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{r}}\} - \boldsymbol{W}_n\{\widehat{\boldsymbol{\beta}}(\boldsymbol{r}_0), \widehat{\boldsymbol{\alpha}}, \boldsymbol{r}_0\}]$ can be written as sum of i.i.d. influence functions, and moreover is asymptotically linear with regard to $n^{1/2}(\widehat{\boldsymbol{r}} - \boldsymbol{r}_0)$. The combination of these two results leads to the weak convergency of the proposed estimators. We provide the detailed proofs in Chapter 3.4.2–3.4.3, along with the specific form of the influence functions for $n^{1/2}\{\widehat{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}_0(\tau)\}$ and $n^{1/2}(\widehat{\boldsymbol{r}} - \boldsymbol{r}_0)$.

3.1.5 Inferences

Due to the complexity involved in the limit distribution of $\widehat{\boldsymbol{\beta}}(\tau)$ and $\widehat{\boldsymbol{r}}$, we propose to use a bootstrap method for the inference on $\boldsymbol{\beta}_0(\tau)$ and $\boldsymbol{r}_0$. Specifically, we employ the paired-bootstrap scheme (Efron, 1979) and apply the algorithm presented in Section 3.1.3 to each of the B bootstrapped datasets to acquire $\{\widehat{\boldsymbol{\alpha}}_b^*, \widehat{\boldsymbol{\beta}}_b^*, \widehat{\boldsymbol{r}}_b^*\}_{b=1}^B$. For a fixed $\tau \in (\tau_0, \tau_{U,1}]$, we can use the sample standard deviation of $\{\widehat{\boldsymbol{\beta}}_b^*(\tau)\}_{b=1}^B$ to estimate the standard deviation of $\widehat{\boldsymbol{\beta}}(\tau)$. Similarly, the standard deviation of $\widehat{\boldsymbol{r}}$ can be approximated by the sample standard error of $\{\widehat{\boldsymbol{r}}_b^*\}_{b=1}^B$. The confidence intervals for $\boldsymbol{\beta}_0(\tau)$ and $\boldsymbol{r}_0$ can be constructed using normal approximation, or with the empirical percentiles of the bootstrap estimates.

We can also perform second stage inferences for exploring the varying pattern of covariate effects over τ. First, we define a trimmed mean effect, $\Phi_{1,q} = \{\int_l^u \boldsymbol{\beta}_0^{(q)}(\tau)\mathrm{d}\tau\}/(u-l)$, to summarize the effect of $\boldsymbol{Z}^{(q)}$. Here and hereafter, we use $u^{(q)}$ to denote the qth component of a vector $\mathbf{u}$. As a natural estimate of $\Phi_{1,q}$, $\widehat{\Phi}_{1,q} = \{\int_l^u \widehat{\beta}^{(q)}(\tau)\mathrm{d}\tau\}/(u-l)$

can be shown to be consistent and asymptotically normal under mild assumption on the functional form of $\boldsymbol{\beta}_0(\cdot)$. The standard error and the percentiles of $\widehat{\Phi}_{1,q}$ can be estimated using the bootstrap realizations $[\{\int_l^u \widehat{\boldsymbol{\beta}}_b^{*(q)}(\tau)\mathrm{d}\tau\}/(u-l)]_{b=1}^B$. This result can be easily extended to an integral test for evaluating whether a covariate $Z^{(q)}$ $(2 \le q \le p+1)$ has a significant effect on a range of quantiles with $\tau \in [l,u]$, namely, $H_{01}:\ \beta^{(q)}(\tau)=0,\ \tau \in [l,u]$. In practice, one may be interested in accessing whether the effect of a covariate $Z^{(q)}$ is constant for $\tau \in [l,u)$. The null hypothesis can be formulated as $H_{02}:\ \boldsymbol{\beta}_0^{(q)}(\tau)=c_0,\ \tau \in [l,u)$, where c_0 is an unspecified constant. With a predetermined nonconstant weight function $\Xi(\tau)$ satisfying $\int_l^u \Xi(\tau)\mathrm{d}\tau = 1$, the test statistic can be constructed as $\mathcal{T}_{2,q} = \int_l^u \Xi(\tau)\widehat{\beta}^{(q)}(\tau)\mathrm{d}\tau - \widehat{\Phi}_{1,q}$, standard error of which can be estimated using $[\int_l^u \Xi(\tau)\widehat{\beta}_b^{*(q)}(\tau)\mathrm{d}\tau - \{\int_l^u \widehat{\beta}_b^{*(q)}(\tau)\mathrm{d}\tau\}/(u-l)]_{b=1}^B$. Wald-type or percentile-based hypothesis testing can be performed accordingly. Rejection of H_{02} indicates that the effect of $Z^{(q)}$ may not be constant across the quantiles.

3.2 Numerical Studies

3.2.1 Simulations

We evaluated the finite-sample performance of the proposed estimators via extensive Monte Carlo simulations. Let the covariate vector $\mathbf{Z} = (1, Z_1, Z_2)^T$, where $Z_1 \sim Unif(0,1)$ and $Z_2 \sim Bernoulli(0.5)$. The model used for generating T_1 takes the form $\log T_1 = b_1 Z_1 + b_2 Z_2 + \varepsilon_1$, where the error term ε_1 is $Normal(0, 0.15^2)$ when $Z_2 = 0$, and $Normal(0, 0.5^2)$ when $Z_2 = 1$. With such heteroscedastic error structure, the underlying regression quantile $\boldsymbol{\beta}_0(\tau) = \{\beta_0^{(1)}(\tau), \beta_0^{(2)}(\tau), \beta_0^{(3)}(\tau)\}$ is given by $\beta_0^{(1)}(\tau) = Qnorm(\tau, 0, 0.15^2)$, $\beta_0^{(2)}(\tau) = b_1$ and $\beta_0^{(3)}(\tau) = Qnorm(\tau, 0, 0.5^2) - Qnorm(\tau, 0, 0.15^2) + b_2$. Notice that both $\beta_0^{(1)}(\tau)$ and $\beta_0^{(3)}(\tau)$ vary with τ while $\beta_0^{(2)}(\tau)$ is constant. We generated the dependent censoring time T_2 from a log-linear model with i.i.d. errors, $\log T_2 = a_1 Z_1 + a_2 Z_2 + \varepsilon_2$, where $\varepsilon_2 \sim Normal(\mu_2, 0.5^2)$. For the association structure between T_1 and T_2, we considered two types of copulas: the Clayton's copula (Oakes, 1982) and the Frank's copula (Genest, 1987). The bivariate survival function $\Pr(T_1 > s, T_2 > t|\boldsymbol{Z})$ is given by

$$\{\Pr(T_1 > s|\boldsymbol{Z})^{-\exp(r_c)} + \Pr(T_2 > t|\boldsymbol{Z})^{-\exp(r_c)} - 1\}^{-1/\exp(r_c)}$$

when the Clayton's copula is adopted, and by

$$-\frac{1}{r_f}\log\left(1 + \frac{\left[\exp\{-r_f \Pr(T_1 > s|\boldsymbol{Z})\} - 1\right] \times \left[\exp\{-r_f \Pr(T_2 > t|\boldsymbol{Z})\} - 1\right]}{\exp(-r_f) - 1}\right)$$

when the Frank's copula is adopted. For a fixed $\boldsymbol{Z}$, the Kendall's tau coefficient between T_1 and T_2 is linked to the copula parameters by $\exp(r_c)/\{\exp(r_c) + 2\}$ under the Clayton's model, and by $1 + 4\{D_1(r_f) - 1\}/r_f$ under the Frank's model, where $D_1(r) = [\int_0^r t/\{\exp(t) - 1\}\mathrm{d}t]/r$. We set $r_c = 1$ and $r_f = 7.325$, so that the corresponding Kendall's tau coefficients equal 0.576 under both copula models. The independent censoring time C was set to follow $Unif(0, U_C)$. We considered different combinations of μ_2, (a_1, a_2), (b_1, b_2) and U_C, which led to the 4 setups presented in Table 3.1.

Table 3.1: Summary of simulation setups, with $CP_1 = \Pr(T_2 < T_1)$, $CP_2 = P(Y < T_1)$ and $CP_3 = P(C < T_2)$.

Setup	Copula	μ_2	(a_1, a_2)	(b_1, b_2)	U_C	CP_1	CP_2	CP_3
S1.C	Clayton	0.1	$(0.4, 0.2)$	$(0, 0)$	18	0.85	0.80	0.90
S1.F	Frank	0.1	$(0.4, 0.2)$	$(0, 0)$	18	0.85	0.80	0.90
S2.C	Clayton	0.0	$(0.32, -0.1)$	$(-0.5, 0)$	8.5	0.80	0.72	0.85
S2.F	Frank	0.0	$(0.32, -0.1)$	$(-0.5, 0)$	8.5	0.79	0.72	0.85

Under each setup, we generated 2500 simulated datasets and implemented the proposed numerical algorithm to obtain $\widehat{\boldsymbol{\beta}}(\tau)$ and $\widehat{\boldsymbol{r}}$. We chose sample size $n = 200$, grid size $||\mathcal{G}_{L(n)}|| = 0.01$, $\tau_a = 0.15$, $\tau_b = 0.6$, and bootstrap sample size $B = 200$. The simulation results on $\widehat{\boldsymbol{\beta}}(\tau)$, $\tau = 0.2, 0.3, ..., 0.7$, are summarized in Table 3.2. We report the empirical bias (EBias), the empirical standard error (ESD), and the average resampling-based standard error (ASD) for $\widehat{\boldsymbol{\beta}}(\tau)$, as well as the empirical coverage probabilities of the 95% Wald-type ($\mathrm{ECP_W}$) and percentile-based ($\mathrm{ECP_P}$) confidence intervals of $\boldsymbol{\beta}_0(\tau)$. It is shown that the proposed $\widehat{\boldsymbol{\beta}}(\tau)$ is close to $\boldsymbol{\beta}_0(\tau)$ under all setups, and the bootstrap standard errors closely match the empirical standard errors. Both the Wald-type and the percentile confidence intervals achieve empirical coverage probabilities that are close to the nominal level. It is observed that the percentile-based confidence intervals may perform better than the Wald-type intervals when τ is small.

Table 3.3 summarizes the simulation results on $\widehat{\boldsymbol{r}}$. We present the same set of summary statistics including EBias, ESD, ASD, $\mathrm{ECP_W}$ and $\mathrm{ECP_P}$. We can see $\widehat{\boldsymbol{r}}$ is virtually unbiased, and the estimated standard errors are close to their empirical counterparts. We note that the Wald-type confidence intervals have slight over-coverages under the Clayton's model, while the percentile-based confidence intervals consistently have good coverages under both copula models. In Table 3.3, we also present the empirical bias and empirical standard error of the corresponding Kendall's tau coefficient estimates, denoted as $\mathcal{K}(\widehat{\boldsymbol{r}})$. As an established measure of association, the Kendall's tau coefficients are accurately estimated with small standard errors.

The results of second stage inferences are summarized in Table 3.4. We present the EBias, ESD and ASD of the trimmed mean effect estimates $\widehat{\Phi}_1^{(q)} = \{\int_l^u \widehat{\boldsymbol{\beta}}^{(q)}(\tau)\mathrm{d}\tau\}/(u-l)$, where $q = 2, 3$, $l = 0.1$ and $u = 0.7$. We also summarize the performances of the two hypothesis tests for H_{01} : $\{\int_l^u \boldsymbol{\beta}_0^{(q)}(\tau)\mathrm{d}\tau\}/(u-l) = 0$ and H_{02} : $\boldsymbol{\beta}_0^{(q)}(\tau) \equiv c_0$, following the procedures in Chapter 3.1.5. To construct the test statistics for H_{02}, we chose the weight function $\Xi(\tau)$ as $2I(\tau \le \frac{\tau_L+\tau_U}{2})/(\tau_U - \tau_L)$. For each test, we report the empirical rejection rates based on normal approximation ($\mathrm{ERR_W}$) and those based on percentiles ($\mathrm{ERR_P}$). We find that the trimmed mean effect estimates $\widehat{\Phi}_1^{(q)}$, $q = 2, 3$, are well-performed in terms of biases and standard error estimates. It is shown that both the Wald-type test and the percentile-based test achieve empirical sizes close to

Table 3.2: Summary of simulation results on $\widehat{\boldsymbol{\beta}}(\tau)$, which include the empirical bias ($\times 10^3$), empirical standard error ($\times 10^3$), average of resampling based standard error estimates ($\times 10^3$), as well as empirical coverages (%) of 95% wald-type and percentile confidence intervals.

τ	**EBias**			**ESD**			**ASD**			$\mathbf{ECP_W}$			$\mathbf{ECP_P}$		
	$\widehat{\boldsymbol{\beta}}^{(0)}$	$\widehat{\boldsymbol{\beta}}^{(1)}$	$\widehat{\boldsymbol{\beta}}^{(2)}$	$\widehat{\boldsymbol{\beta}}^{(0)}$	$\widehat{\boldsymbol{\beta}}^{(1)}$	$\widehat{\boldsymbol{\beta}}^{(2)}$	$\widehat{\boldsymbol{\beta}}^{(0)}$	$\widehat{\boldsymbol{\beta}}^{(1)}$	$\widehat{\boldsymbol{\beta}}^{(2)}$	$\widehat{\boldsymbol{\beta}}^{(0)}$	$\widehat{\boldsymbol{\beta}}^{(1)}$	$\widehat{\boldsymbol{\beta}}^{(2)}$	$\widehat{\boldsymbol{\beta}}^{(0)}$	$\widehat{\boldsymbol{\beta}}^{(1)}$	$\widehat{\boldsymbol{\beta}}^{(2)}$
							Setup S1.C								
0.2	8	−14	−10	71	113	84	72	116	85	93.2	94.7	93.7	94.5	95.5	95.3
0.3	5	−10	−7	65	105	78	67	108	79	93.9	94.7	93.7	95.1	95.5	94.9
0.4	3	−8	−5	60	97	77	63	102	77	94.2	95.0	93.5	94.8	95.5	94.2
0.5	2	−7	−3	57	94	76	59	98	76	95.0	95.8	93.3	94.9	96.1	94.3
0.6	1	−6	−4	54	89	75	57	95	76	94.8	95.4	93.9	95.1	96.4	94.9
0.7	0	−6	−3	52	88	78	55	94	79	95.0	95.9	94.0	95.4	96.4	95.0
							Setup S1.F								
0.2	9	−14	−7	74	117	81	74	119	84	92.5	93.6	94.7	94.3	95.2	95.4
0.3	5	−9	−6	62	99	75	65	106	77	93.9	95.2	94.2	95.2	95.7	94.8
0.4	3	−6	−4	55	91	72	59	97	74	95.0	95.6	94.3	94.9	95.4	95.0
0.5	2	−5	−4	52	86	72	54	91	73	94.9	95.4	93.4	94.7	96.1	95.1
0.6	1	−4	−3	48	82	72	52	89	73	95.2	95.6	93.6	95.3	95.9	95.0
0.7	1	−5	−3	48	83	75	52	89	76	95.6	95.8	94.0	95.5	95.9	94.9
							Setup S2.C								
0.2	14	−25	−2	74	117	90	76	121	88	93.2	94.4	93.0	94.6	95.9	94.3
0.3	11	−20	0	69	108	87	71	114	85	93.8	94.9	93.4	95.2	95.9	94.1
0.4	9	−18	3	64	103	86	67	108	84	94.6	95.0	92.6	95.4	95.7	93.9
0.5	8	−16	4	59	96	85	63	103	86	95.5	95.3	93.0	95.7	95.7	94.6
0.6	5	−14	7	55	93	87	60	100	88	95.2	95.3	93.8	95.9	95.9	94.8
0.7	4	−12	8	54	91	90	58	98	92	95.7	95.1	93.7	95.5	95.3	94.8
							Setup S2.F								
0.2	16	−25	1	74	113	92	75	118	88	92.8	94.6	92.2	94.2	96.4	93.6
0.3	11	−17	5	64	99	85	67	106	83	94.1	95.3	92.4	94.6	95.9	93.6
0.4	8	−14	7	57	91	81	61	98	80	94.8	95.2	92.7	94.8	95.8	93.8
0.5	6	−12	7	53	86	79	57	93	79	95.6	96.1	93.4	95.3	96.1	93.4
0.6	4	−10	5	50	84	79	55	92	80	95.3	95.6	93.8	95.2	96.0	94.2
0.7	5	−11	7	49	84	83	55	92	86	95.8	95.4	94.8	95.6	96.3	94.9

Table 3.3: Summary of simulation results on $\widehat{\boldsymbol{r}}$, which include the true value, empirical bias ($\times 10^3$), empirical standard error ($\times 10^3$), average of resampling based standard error estimates ($\times 10^3$), empirical coverages (%) of 95% wald-type and percentile-based confidence intervals, as well as the true value, empirical bias ($\times 10^3$) and empirical standard errors ($\times 10^3$) of the corresponding Kendall's tau coefficient estimates.

	$\widehat{\boldsymbol{r}}$						$\mathcal{K}(\widehat{\boldsymbol{r}})$		
	TRUE	**EBias**	**ESD**	**ASD**	**ECP$_\mathbf{W}$**	**ECP$_\mathbf{P}$**	**TRUE**	**EBias**	**ESD**
S1.C	1	−28	244	277	98.3	96.9	0.576	−8	59
S1.F	7.325	−131	1377	1527	95.7	95.9	0.576	−12	57
S2.C	1	−25	287	333	98.6	96.6	0.576	−7	69
S2.F	7.325	−61	1633	1840	95.7	96.5	0.576	−12	67

the nominal level 0.05. The Wald-type test may slightly over-perform the percentile-based test, particularly for H_{02}. We observe that both hypothesis testing methods are satisfactory in terms of power.

We also perform simulation studies to evaluate the robustness of the proposed estimators. Specifically, we examine the proposed estimators when the association structures are misspecified. We generated data from S2.C and S2.F, which correspond to the first and second row in Figure 3.1, respectively. For each setup, we implemented the proposed method by assuming the Clayton's copula and the Frank's copula. Figure 3.1 summarizes the empirical mean of the resulting $\widehat{\boldsymbol{\beta}}(\tau)$ under correctly or incorrectly specified copula assumptions. For comparison, we also plot the empirical mean of Peng and Huang (2008)'s estimators by naively treating T_2 as independent censoring. Not surprisingly, the naive estimator yield large bias for $\tau \in [0.1, 0.7]$ in all scenarios. We notice that the intercept term tend to be positively biased, and the bias manifest in the remainder regression coefficients as well. When the association structures are correctly specified, the empirical mean of the proposed $\widehat{\boldsymbol{\beta}}(\tau)$ closely match the true regression quantiles, which agrees with the results in Table 3.2. With misspecified association structures, the proposed estimators only exhibit small deviations from the true coefficients. For example, the bias of

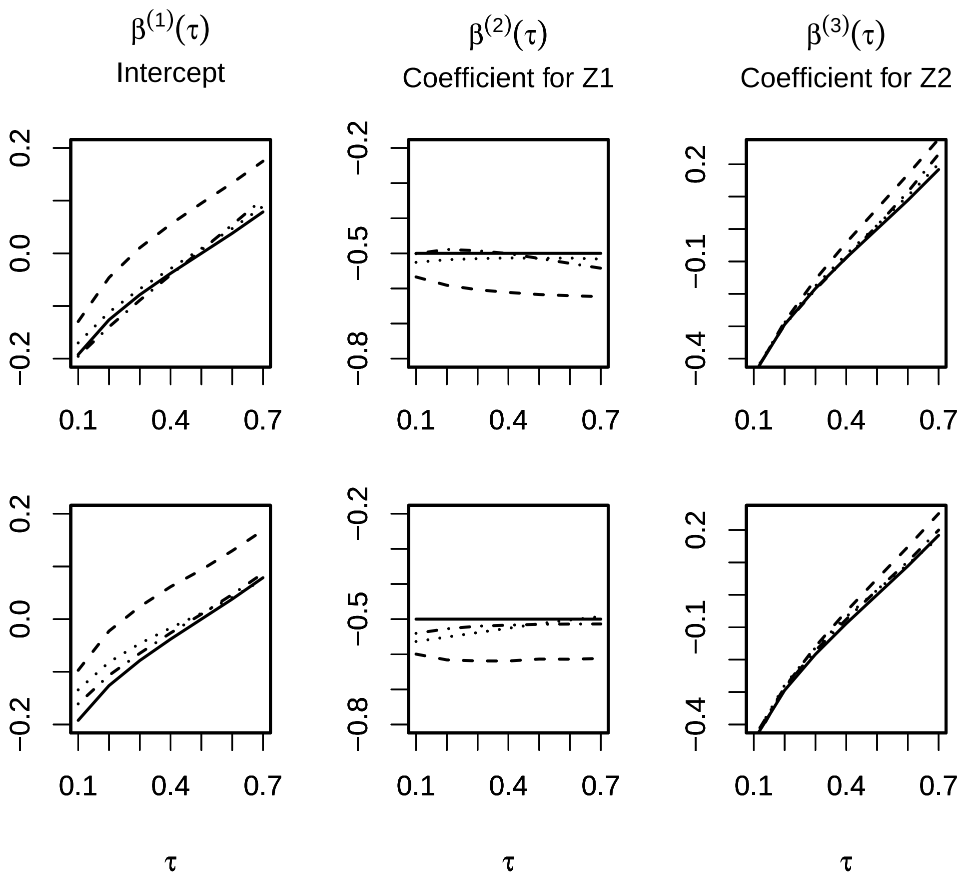
Table 3.4: Summary of simulation results on second stage inferences, which include empirical bias ($\times 10^3$), empirical standard error ($\times 10^3$) and average of resampling based standard error estimates ($\times 10^3$) of $\widehat{\Phi}_1^{(q)}$, $q = 2, 3$, as well as the empirical rejection rates of H_{01} and H_{02} with Wald-type ($\mathrm{ERR_W}$) and percentile-type ($\mathrm{ERR_P}$) methods.

		$\widehat{\Phi}_1^{(q)}$			H_{01}		H_{02}	
	q	**EBias**	**ESD**	**ASD**	$\mathbf{ERR_W}$	$\mathbf{ERR_P}$	$\mathbf{ERR_W}$	$\mathbf{ERR_P}$
S1.C	2	-10	82	81	0.057	0.059	0.046	0.033
	3	-7	64	61	0.457	0.456	1.000	1.000
S1.F	2	-9	79	78	0.059	0.067	0.043	0.032
	3	-6	61	59	0.469	0.477	1.000	1.000
S2.C	2	-19	85	85	1.000	1.000	0.047	0.032
	3	1	69	65	0.380	0.361	0.998	1.000
S2.F	2	-17	76	78	1.000	1.000	0.043	0.026
	3	5	65	61	0.386	0.373	0.999	1.000

the proposed $\widehat{\boldsymbol{\beta}}(\tau)$, when incorrectly assuming the Frank's copula, is nearly negligible in S2.C. Next, we examine the performance of association parameter estimates. The empirical mean of estimated Kendall's tau coefficient, when Frank's copula is assumed in S2.C, equals 0.568, and that when Clayton's copula is assumed in S2.F equals 0.601. Both empirical means are reasonably close to the true Kendall's coefficient 0.576. Our simulations suggest that the proposed estimators are robust to the misspecification of the association structure, in both the regression quantile estimates and the association parameter estimates.

Finally, we present summary statistics which demonstrate the computational efficiency of the proposed procedures. In Table 3.5, we report the percentage of non-convergence (NoneConv), the mean CPU time in second for each simulation without resampling (CPUtime), and the percentage of simulations which converge within 5 iterations between Step C and Step D (PerLe5). All programs were run in *R 2.9.1* on a computer with 2.66GHz core. Table 5 shows that the proposed algorithm converges fast and stably under all setups. The majority of simulations converge within 5 iterations, and the rates of non-convergence are negligible. The CPU time consumed for

Figure 3.1: Performance of the proposed method under different copula assumptions. (—— true regression quantiles; - - - empirical mean of naive estimators; · · · empirical mean of proposed estimators, assuming Clayton's copula; — · — empirical mean of proposed estimators, assuming Frank's copula. Data are generated from setup S2.C in the first row, and from setup S2.F in the second row).



each simulation increases with the censoring rate, and is affordable under all setups.

Table 3.5: Summary of convergence statistics, which include the percent of nonconvergence, the mean CPU time for one simulation, and the percent of simulations that converged within 5 iterations.

	NoneConv (%)	**CPUtime** (seconds)	**PerLe5** (%)
S1.C	0.0	1.913	99.4
S1.F	0.0	1.804	99.3
S2.C	0.0	2.351	98.1
S2.F	0.4	2.072	97.2

In summary, our simulation studies show the proposed methods perform well and are robust with realistic sample sizes. Moreover, the nice computational performance facilitates the practical utility of the proposed method.

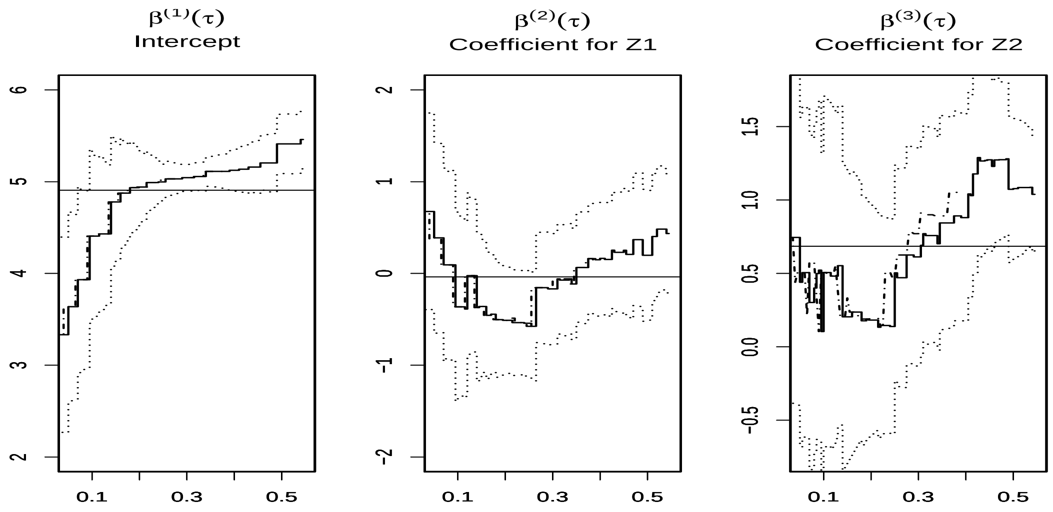
3.2.2 Aids Clinical Trial Group 364 Data Analysis

We illustrate the proposed methods by an application to the ACTG 364 study (Albrecht et al., 2001), a multicenter clinical trial in AIDS patients with plasma human immunodeficiency virus HIV RNA above 500 copies/ml. In this study, 195 subjects were enrolled and randomly assigned to one of the three treatment arms, which are protease inhibitor nelfinavir (NFV), the non-nucleoside efavirenz (EFV), and NFV+EFV. The randomization was stratified according to the treatments patients had received in previous studies. In this study, T_1 and T_2 correspond to time to first virologic failure and time to dropout, respectively, while C represents the duration of follow-up at the end of the study. In this analysis we focus on the subgroup of patients who had been treated by lamivudine before entering the study. Among the 129 patients, 44 dropped out from the assigned treatment within the follow-up period for disease-related reasons, and 81 were observed to experience first virologic failure. Among the 48 patients whose virologic endpoints were not observed, 7 were due to dropouts and the other 41 were administratively censored at the end of the study. Though the censoring by dropout is not severe, 6 out of the 7 occurred in the NFV + EFV group, which indicates the dropouts may not be random.

We consider two covariates, Z_1 and Z_2, which are treatment indicators for EFV and NFV + EFV respectively. Therefore, the first element in $\boldsymbol{\beta}_0(\tau) = \{\beta_0^{(1)}(\tau), \beta_0^{(2)}(\tau), \beta_0^{(3)}(\tau)\}$ represents the log transformed quantiles in the NFV group, while $\beta_0^{(2)}(\tau)$ and $\beta_0^{(3)}(\tau)$ represent the effects of EFV vs. NFV and NFV + EFV vs. NFV respectively, at the τth quantile of time to first virologic failure. We implement the proposed algorithm by assuming the Frank's copula, and setting $\tau_a = 0.15$, $\tau_b = 0.45$, $||\mathcal{G}_{L(n)}|| = 0.005$. The standard error estimates and confidence intervals are obtained with 1000 bootstrap resamples, out of which 7% were omitted for reason of non-convergence. We obtain $\widehat{\boldsymbol{r}} = 2.5$, with a standard error estimate of 3.05. The estimate of Kendall's tau coefficient between T_1 and T_2 equals 0.26 with a standard error estimate of 0.21. This may indicate that patients with slower virologic disease progressions tend to stay longer on the assigned treatment. However, this association is not significant at the 0.05 significance level.

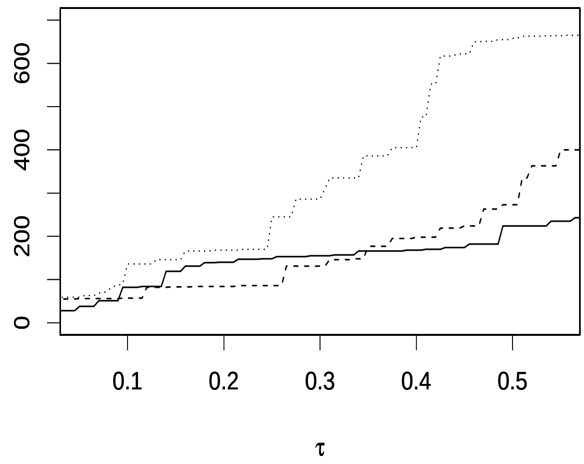
In Figure 3.2 we plot the estimated $\widehat{\boldsymbol{\beta}}(\tau)$ for $\tau \in [0.05, 0.55)$, along with the pointwise 95% Wald-type intervals. Also plotted is the naive estimate $\widehat{\boldsymbol{\beta}}_N(\tau)$ from Peng and Huang (2008)'s method, which converges up till $\tau = 0.38$. In this example, $\widehat{\boldsymbol{\beta}}(\tau)$ and $\widehat{\boldsymbol{\beta}}_N(\tau)$ are quite similar, probably because the rate of censoring by dropout is quite low. We observe a clear trend of over-estimating the regression coefficient for Z_2 by the naive method. This may not be surprising, because dropouts mainly occurred in the NFV + EFV group. The lower bound of the confidence interval for $\widehat{\beta}^{(3)}(\tau)$ stays above 0 when $\tau \geq 0.3$, suggesting that patients in the NFV + EFV arm tend to have longer virologic failure free survival time when compared to those in the NFV arm. We observe an increasing pattern of $\widehat{\beta}^{(3)}(\tau)$ with τ. This suggests that the difference in time to first virologic failure between the NFV + EFV group and the NFV group may be more pronounced for those with late onset of virologic failure compared to those who had fast progression to virologic failure. In contrast, the confidence interval for $\widehat{\beta}^{(2)}(\tau)$ always cover 0, which suggests time to first virologic failure may not be significantly different between the EFV arm and the NFV arm.

Figure 3.2: Estimated regression coefficients for the ACTG 364 study. (——, estimated regression coefficients using the proposed method; $\cdots$, 95% Wald-type pointwise confidence intervals; —, estimate of trimmed mean effects; - - -, naive estimates using Peng and Huang (2008)'s method.)



Our second stage explorations confirm our observations in Figure 3.2. The estimated trimmed mean effect of Z_2 (NFV + EFV vs. NFV), for $\tau \in [0.05, 0.55)$, is 0.685 with an estimated standard error of 0.277. This renders a p-value of 0.01. The trimmed mean effect of Z_1 (EFV vs. NFV) is estimated as -0.038. The corresponding standard error estimate is 0.241, which leads to a p-value of 0.9. These provide some evidence for the superiority of the combined treatment NFV + EFV over EFV, but not for NFV. We also conduct hypothesis testing to investigate whether covariate effects are constant over the quantiles. To compute $\widehat{\Psi}_2^{(q)}$, $q = 2, 3$, we set $l = 0.05$, $u = 0.55$, and $\Xi(\tau) = 2I(\tau \leq \frac{l+u}{2})/(u - l)$. For Z_2, we obtain $\widehat{\Psi}_2^{(3)}$ as -0.324 and its standard error estimate equals to 0.152. The resulting Wald-type p-value equals 0.033, which confirms our impression in Figure 2 that the effect of Z_2 varies with the quantiles. There are some possible explanations of this interesting pattern. For

Figure 3.3: Predicted quantiles of time to first virologic failure in days for the NFV group (—), EFV group (- - -) and NFV + EFV group (···).



example, those who had virologic failure rather quick may not have stayed on treatment long enough for the effects to manifest, or their quick failures may be mainly attributed to some other factors that dominated the treatment effects. For Z_3, we obtain the test statistics $\widehat{\Psi}_2^{(2)}$ as -0.21 and its standard error estimate as 0.14. This leads to a Wald-type p-value of 0.12. Therefore, it may be adequate to conclude that the difference in time to first virologic failure between the EFV group and the NFV arm is constant.

Finally, we plot the predicted quantiles of time to first virologic failure for each treatment groups in Figure 3.3. This plot is practically useful, as it allows us to visualize the various quantiles of time to the virologic endpoint. It is clear from Figure 3 that the predicted quantiles for the NFV+EFV group is always larger than that of the NFV group. The magnitude of difference is more pronounced in upper quantiles of T_1 than that in lower quantiles. At $\tau = 0.5$, for example, the difference is as large as 434 days, which is clinically significant. The quantile curves for the EFV group and NFV group intersect and do not show big difference.

In summary, quantile regression based on the proposed method offers a comprehensive and meaningful view of the virologic endpoint in the ACTG 364 study. Our analyses suggest that treatment NFV+EFV prolongs the virologic failure free survival time when compared to NFV alone. Moreover, the benefit of NFV+EFV over NFV is more substantial among patients subject to low or moderate risk of virologic failure.

3.3 Discussion

In this Chapter, we propose a quantile regression method that can properly accommodate dependent censoring situations that fall into the paradigm of semi-competing risks. Our method offers a useful tool for investigating nonterminating endpoints that often arise in clinical follow-up studies and their relationship with subsequent competing endpoints. The marginal quantile inference pursued in this work is sensible when studying covariate effects with the removal of dependent censoring is of substantive relevance.

We impose assumptions on the dependent censoring scheme through a general class of copula model. This is necessary for addressing the identifiability issue inherited with the dependent censoring problem. Simulations have shown that the proposed estimators are quite robust to misspecification of the copula model. This robustness feature is expected to enhance the practical utility of the proposed method.

3.4 Proofs

3.4.1 Regularity Conditions

For a vector $\boldsymbol{x}$, let $\boldsymbol{x}^{\otimes 2}$ denote $\boldsymbol{x}\boldsymbol{x}^T$, and $||\boldsymbol{x}||$ denote the Euclidean norm of $\boldsymbol{x}$. For a random variable T, let $f_T(\cdot|\boldsymbol{z})$ denote its conditional density function given

$\boldsymbol{Z} = \boldsymbol{z}$. We define $\boldsymbol{s}(\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{r}, \tau) = E\boldsymbol{S}_n(\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{r}, \tau)$, $\boldsymbol{w}(\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{r}, \tau) = E\boldsymbol{W}_n(\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{r}, \tau)$, and $\boldsymbol{w}(\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{r}) = \int_{\tau_a}^{\tau_b} \boldsymbol{w}(\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{r}, \tau)\mathrm{d}\tau$. Let $\boldsymbol{B}_b(\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{r}, \tau) = \partial \boldsymbol{s}(\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{r}, \tau)/\partial \boldsymbol{\beta}(\tau)$, $\boldsymbol{B}_r(\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{r}, \tau) = \partial \boldsymbol{s}(\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{r}, \tau)/\partial \boldsymbol{r}$, $\boldsymbol{L}_b(\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{r}, \tau) = \partial \boldsymbol{w}(\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{r}, \tau)/\partial \boldsymbol{\beta}(\tau)$, and $\boldsymbol{L}_{\boldsymbol{r}}(\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{r}) = \partial \boldsymbol{w}(\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{r})/\partial \boldsymbol{r}$. Next, define $\boldsymbol{\mu}(\boldsymbol{a}) = E[\boldsymbol{Z}I\{Y \le \exp(\boldsymbol{Z}^T\boldsymbol{a}), \eta = 1\}]$ as in Peng and Huang (2008), and $\boldsymbol{\lambda}(\boldsymbol{b}) = E[\boldsymbol{Z}I\{T_1 \le \exp(\boldsymbol{Z}^T\boldsymbol{b})\}]$. For $d > 0$, let $\mathcal{A}(d) = \{\boldsymbol{a} \in R^{p+1} : \inf_{\tau\in(0,\tau_{U,2}]} ||\boldsymbol{\mu}(\boldsymbol{a}) - \boldsymbol{\mu}\{\boldsymbol{\alpha}_0(\tau)\}|| \le d\}$ and $\mathcal{R}(d) = \{\boldsymbol{r} \in R^{q+1} : ||\boldsymbol{r} - \boldsymbol{r}_0|| \le d\}$. Define $\bar{\mathcal{B}}(d) = \{\boldsymbol{\beta} \in R^{p+1} : \max_{\tau\in(0,\tau_{U,1}]} ||\boldsymbol{\lambda}\{\boldsymbol{\beta}(\tau)\} - \boldsymbol{\lambda}\{\boldsymbol{\beta}_0(\tau)\}|| \le d\}$. Let d_A, d_B and d_R be positive constants that determine the span of the neighborhoods.

We require the following regularity conditions:

C1 The covariate space $\mathcal{Z}$ is bounded, i.e., $\sup_i ||\boldsymbol{Z}_i|| \le \infty$.

C2 (i) The regularity conditions in Peng and Huang (2008) hold for $(Y, \eta, \boldsymbol{Z})$ and $\boldsymbol{\alpha}_0(\tau)$, $\tau \in (0, \tau_{U,2}]$, (ii) for $\boldsymbol{B}_\alpha(\boldsymbol{a}) = \partial\boldsymbol{\mu}(\boldsymbol{a})/\partial\boldsymbol{a}$, there exist a constant C_F, such that each component of $||f_{T_2}\{\exp(\boldsymbol{z}^T\boldsymbol{a})|\boldsymbol{z}\}\exp(\boldsymbol{z}^T\boldsymbol{a}) \times \boldsymbol{B}_\alpha(\boldsymbol{a})^{-1}||$ is bounded by C_F uniformly in $\boldsymbol{z} \in \mathcal{Z}$ and $\boldsymbol{a} \in \mathcal{A}(d_A)$, where d_A is a positive constant. (iii) Define

$$\boldsymbol{V}_A(\boldsymbol{a}, \tau) = -\boldsymbol{Z}\big[I\{\log Y > \boldsymbol{Z}^T\boldsymbol{\beta}_0(\tau)\}d_A\big\{\tau, F_2[\exp\{\boldsymbol{Z}^T\boldsymbol{\beta}_0(\tau)\}|\boldsymbol{Z}], g(\boldsymbol{Z}^T\boldsymbol{r}_0)\big\}\times I\{\boldsymbol{Z}^T\boldsymbol{\beta}_0(\tau) \ge \boldsymbol{Z}^T\boldsymbol{a}\}\big]$$

$$\boldsymbol{V}_B(\boldsymbol{a}) = -\bar{\boldsymbol{Z}}\Big\{\int_{\tau_a}^{\tau_b} I[Y > 2\exp\{\boldsymbol{Z}^T\boldsymbol{\beta}_0(\tau)\}]d_B\big\{\tau, F_2[2\exp\{\boldsymbol{Z}^T\boldsymbol{\beta}_0(\tau)\}|\boldsymbol{Z}], g(\boldsymbol{Z}^T\boldsymbol{r}_0)\big\} I[2\exp\{\boldsymbol{Z}^T\boldsymbol{\beta}_0(\tau)\} \ge \exp\{\boldsymbol{Z}^T\boldsymbol{a}\}]\mathrm{d}\tau\Big\},$$

where $d_A(u, v, \theta) = \partial K_A(u, v, \theta)/\partial v$ and $d_B(u, v, \theta) = \partial K_B(u, v, \theta)/\partial v$. Every component of $[\partial E\{\boldsymbol{V}_A(\boldsymbol{a}, \tau)\}/\partial\boldsymbol{a}]\boldsymbol{B}_\alpha(\boldsymbol{a})^{-1}$ and $[\partial E\{\boldsymbol{V}_B(\boldsymbol{a})\}/\partial\boldsymbol{a}]\boldsymbol{B}_\alpha(\boldsymbol{a})^{-1}$ are bounded uniformly for $\boldsymbol{a} \in \mathcal{A}(d_A)$ and $\tau \in (0, \tau_{U,1}]$ when $\boldsymbol{Z}$ contains continuous components.

C3 (i)Each component of $\boldsymbol{\lambda}\{\boldsymbol{\beta}_0(\tau)\}$ is a Lipschitz function of τ when $\tau \in (0, \tau_{U,1}]$,

(ii) let $d_\theta(u,v,\theta) = \partial\Psi(u,v,\theta)/\partial\theta$, $||d_\theta\{u,v,g(\boldsymbol{z}^T\boldsymbol{r}_0)\}g'(\boldsymbol{z}^T\boldsymbol{r}_0)||$ is uniformly bounded for $\boldsymbol{z} \in \mathcal{Z}$, $u \leq \tau_{U,1}$ and $v \leq \tau_{U,2}$. (iii) the conditional density functions $f_X(t|\boldsymbol{z})$, $f_Y(t|\boldsymbol{z})$, $f_{T_1}(t|\boldsymbol{z})$ and $f_{T_2}(t|\boldsymbol{z})$ are bounded above uniformly in t and $\boldsymbol{z} \in \mathcal{Z}$.

C4 (i)$E\boldsymbol{Z}^{\otimes 2} > 0$, and $f_{T_1}[\exp\{\boldsymbol{Z}^T\boldsymbol{\beta}(\tau)\}|\boldsymbol{Z}] > 0$ for any $\boldsymbol{\beta} \in \bar{\mathcal{B}}$ and $\tau \in (0,\tau_{U,1}]$. (ii) Define $\boldsymbol{H}_b(\boldsymbol{b}) = E[\boldsymbol{Z}^{\otimes 2} f_{T_1}\{\exp(\boldsymbol{Z}^T\boldsymbol{b})|\boldsymbol{Z}\}\exp(\boldsymbol{Z}^T\boldsymbol{b})]$. For any $\nu_1 \in (0,\tau_{U,1}]$, min eig$[\boldsymbol{H}_b\{\boldsymbol{\beta}(\tau)\}]$ is bounded below by a positive number uniformly for $\boldsymbol{\beta} \in \bar{\mathcal{B}}(d_B)$ and $\tau \in [\nu_1,\tau_{U,1}]$, where eig$(\cdot)$ denotes the eigenvalues of a matrix, (iii)$\inf_{\boldsymbol{\beta}\in\bar{\mathcal{B}}(d_B),\boldsymbol{r}\in\mathcal{R}(d_R),\tau\in(0,\tau_{U,1}]}$ min eig$[-\boldsymbol{B}_b\{\boldsymbol{\beta},\boldsymbol{\alpha}_0,\boldsymbol{r},\tau\}\boldsymbol{H}_b\{\boldsymbol{\beta}(\tau)\}^{-1}] > 0$, (iv) for any fixed $\boldsymbol{r} \in \mathcal{R}(d_R)$, there exists a unique solution to $\boldsymbol{s}\{\widetilde{\boldsymbol{\beta}}(\boldsymbol{r}),\boldsymbol{\alpha}_0,\boldsymbol{r},\tau\} = 0$ in $\bar{\mathcal{B}}(d_B)$, and min $\left|\text{eig}[\partial\boldsymbol{w}\{\widetilde{\boldsymbol{\beta}}(\boldsymbol{r}),\boldsymbol{\alpha}_0,\boldsymbol{r}\}/\partial\boldsymbol{r}]\right|$ is bounded from 0 uniformly for $\boldsymbol{r} \in \mathcal{R}(d_R)$, (v) every component of $\boldsymbol{L_b}(\boldsymbol{\beta},\boldsymbol{\alpha}_0,\boldsymbol{r}_0,\tau) \times \boldsymbol{B_b}(\boldsymbol{\beta},\boldsymbol{\alpha}_0,\boldsymbol{r}_0,\tau)^{-1}$ is uniformly bounded for $\boldsymbol{\beta} \in \bar{\mathcal{B}}(d_B)$.

C5 (i)For any $\boldsymbol{z} \in \mathcal{Z}$, we have $\boldsymbol{z}^T\boldsymbol{\beta}_0(\tau_{U,1}) \leq \boldsymbol{z}^T\boldsymbol{\alpha}_0(\tau_{U,2})$, $2\exp\{\boldsymbol{z}^T\boldsymbol{\beta}_0(\tau_b)\} \leq \exp\{\boldsymbol{z}^T\boldsymbol{\alpha}_0(\tau_{U,2})\}$. (ii) Both $\inf_{\boldsymbol{z}\in\mathcal{Z}} \Pr[C > \exp\{\boldsymbol{Z}^T\boldsymbol{\beta}_0(\tau_{U,1})\}|\boldsymbol{Z}]$ and $\inf_{\boldsymbol{z}\in\mathcal{Z}} \Pr[C > 2\exp\{\boldsymbol{Z}^T\boldsymbol{\beta}_0(\tau_b)\}|\boldsymbol{Z}]$ are bounded away from 0.

Condition C1 requires bounded covariates and is often met in practice. Condition C2 imposes mild assumptions on the dependent censoring time T_2. C2(ii) ensures that $\int_0^{\tau_{U,2}} I[t \geq \exp\{\boldsymbol{z}^T\widehat{\boldsymbol{\alpha}}(u)\}]\text{d}u$ can serve as a consistent estimate for $F_2(t|\boldsymbol{z}) \wedge \tau_{U,2}$, while C2(iii) facilitates the asymptotic normality of $\boldsymbol{s}(\boldsymbol{\beta}_0,\widehat{\boldsymbol{\alpha}},\boldsymbol{r},\tau)$ and $\boldsymbol{w}(\boldsymbol{\beta}_0,\widehat{\boldsymbol{\alpha}},\boldsymbol{r})$. Condition C3 ensures that $\boldsymbol{s}\{\boldsymbol{\beta}_0(\tau),\boldsymbol{\alpha}_0,\boldsymbol{r},\tau\}$ and $\boldsymbol{w}\{\boldsymbol{\beta}_0(\tau),\boldsymbol{\alpha}_0,\boldsymbol{r},\tau\}$ are smooth in both τ and $\boldsymbol{r}$, and the conditional density functions of X, Y, T_1 and T_2 are uniformly bounded. This type of assumptions are common in quantile regression models. Condition $C4$ ensures the identifiability of the proposed estimator in a neighborhood of $\boldsymbol{\beta}_0$ and $\boldsymbol{r}$. Specifically, C4(i)–(ii) ensures that $\boldsymbol{\lambda}\{\boldsymbol{\beta}(\tau)\}$ is a one-to-one mapping from $\bar{\mathcal{B}}(d_B)$ to $\big\{\boldsymbol{\lambda}\{\boldsymbol{\beta}(\tau)\} : \boldsymbol{\beta} \in \bar{\mathcal{B}}(d_B), \tau \in (0,\tau_{U,1}]\big\}$, and has bounded derivative against $\boldsymbol{\beta}(\tau)$ for $\tau \in [\nu_1, \tau_{U,1}]$. Coupled with C4(iii), this further implies that $\inf_{\tau\in[\nu_1,\tau_{U,1}]} \min \mathrm{eig}[-\boldsymbol{B}_b\{\boldsymbol{\beta}_0, \boldsymbol{\alpha}_0, \boldsymbol{r}_0, \tau\}]$ can be bounded below by a positive number. In fact, one can use simple manipulations to get

$$\boldsymbol{B}_b\{\boldsymbol{\beta}_0, \boldsymbol{\alpha}_0, \boldsymbol{r}_0, \tau\} = -E\bigg(\boldsymbol{Z}^{\otimes 2} I[C > \exp\{\boldsymbol{Z}^T\boldsymbol{\beta}_0(\tau)\}]\Psi_1'\{1-\tau, 1 - F_{T_2}[\exp\{\boldsymbol{Z}^T\boldsymbol{\beta}_0(\tau)\}|\boldsymbol{Z}], \boldsymbol{r}_0\} \\ \times f_{T_1}[\exp\{\boldsymbol{Z}^T\boldsymbol{\beta}_0(\tau)\}|\boldsymbol{Z}]\exp\{\boldsymbol{Z}^T\boldsymbol{\beta}_0(\tau)\}\bigg),$$

where $\Psi_1'(u,v,\theta) = \partial\Psi(u,v,\theta)/\partial u \in (0,1)$. Furthermore, it can be shown that $\inf_{\tau\in(0,\tau_{U,1}]} \min \mathrm{eig}[-\boldsymbol{B}_b\{\boldsymbol{\beta}_0, \boldsymbol{\alpha}_0, \boldsymbol{r}_0, \tau\}\boldsymbol{H}_b\{\boldsymbol{\beta}_0(\tau)\}^{-1}] > 0$ under boundary conditions C5, when $\Psi(u,v,\theta)$ corresponds to the Clayton's copula. Hence it is reasonable to assume $\inf_{\tau\in(0,\tau_{U,1}]} \min \mathrm{eig}[-\boldsymbol{B}_b\{\boldsymbol{\beta}, \boldsymbol{\alpha}_0, \boldsymbol{r}, \tau\}\boldsymbol{H}_b\{\boldsymbol{\beta}(\tau)\}^{-1}] > 0$, as well as the uniqueness of $\tilde{\boldsymbol{\beta}}(\boldsymbol{r})$, for $(\boldsymbol{\beta}, \boldsymbol{r})$ in the vicinity of $(\boldsymbol{\beta}_0, \boldsymbol{r}_0)$. Finally, C5 imposes mild boundary conditions that ensures the estimating equations to be unbiased.

3.4.2 Proof of Theorem 1

Lemma 3.4.1. *For* $\boldsymbol{\mu}(\boldsymbol{a}) = E\boldsymbol{Z}I\{Y \le \exp(\boldsymbol{Z}^T\boldsymbol{a}), \eta = 1\}$, *we have*

$$\sup_{\boldsymbol{z}\in\mathcal{Z},t}\left|\int_0^{\tau_{U,2}} I[t \ge \exp\{\boldsymbol{z}^T\widehat{\boldsymbol{\alpha}}(u)\}]\,du - \int_0^{\tau_{U,2}} I[t \ge \exp\{\boldsymbol{z}^T\boldsymbol{\alpha}_0(u)\}]\,du\right| \\ \le 2C_F \sup_{\tau\in(0,\tau_{U,2}]} ||\boldsymbol{\mu}\{\widehat{\boldsymbol{\alpha}}(\tau)\} - \boldsymbol{\mu}\{\boldsymbol{\alpha}_0(\tau)\}||.$$

Proof of Lemma 3.4.1. By regularity condition C2(ii) and Taylor expansion, we have

$$\sup_{\boldsymbol{z}\in\mathcal{Z}}\left|F_2[\exp\{\boldsymbol{z}^T\widehat{\boldsymbol{\alpha}}(\tau)\}|\boldsymbol{z}] - F_2[\exp\{\boldsymbol{z}^T\boldsymbol{\alpha}_0(\tau)\}|\boldsymbol{z}]\right| \le C_F||\boldsymbol{\mu}\{\widehat{\boldsymbol{\alpha}}(\tau)\} - \boldsymbol{\mu}\{\boldsymbol{\alpha}_0(\tau)\}||,\ \tau \in (0, \tau_{U,2}],$$

where $F_2[\exp\{\boldsymbol{z}^T\boldsymbol{\alpha}_0(\tau)\}|\boldsymbol{z}] = \tau$. Moreover, let ε_F denote $\sup_{\boldsymbol{z}\in\mathcal{Z},\tau\in(0,\tau_{U,2}]}\big|F_2[\exp\{\boldsymbol{z}^T\widehat{\boldsymbol{\alpha}}(\tau)\}|\boldsymbol{z}] -$

$\tau\big|$, we can show

$$\sup_{\boldsymbol{z}\in\mathcal{Z},t}\left|\int_0^{\tau_{U,2}} I[t\geq\exp\{\boldsymbol{z}^T\widehat{\boldsymbol{\alpha}}(u)\}]\mathrm{d}u-\int_0^{\tau_{U,2}} I[t\geq\exp\{\boldsymbol{z}^T\boldsymbol{\alpha}_0(u)\}]\mathrm{d}u\right|$$
$$\leq\sup_{\boldsymbol{z}\in\mathcal{Z},t}\int_0^{\tau_{U,2}} I\big\{F_2(t|\boldsymbol{z})\in\big[F_2[\exp\{\boldsymbol{z}^T\widehat{\boldsymbol{\alpha}}(u)\}|\boldsymbol{z}]\wedge u, F_2[\exp\{\boldsymbol{z}^T\widehat{\boldsymbol{\alpha}}(u)\}|\boldsymbol{z}]\vee u\big]\big\}\mathrm{d}u$$
$$\leq\sup_{\boldsymbol{z}\in\mathcal{Z},t}\int_0^{\tau_{U,2}} I\big\{F_2(t|\boldsymbol{z})\in[u-\varepsilon_F,u+\varepsilon_F]\big\}\mathrm{d}u\leq 2\varepsilon_F.$$

Here and in the sequel, we use $\wedge$ to denote the minimum operator and $\vee$ to denote the maximum operator. This completes the proof of Lemma 3.4.1. □

For any fixed θ, the copula function $\Psi(u,v,\theta)$ satisfies the Lipschitz condition in u and v (Nelsen, 2006). Hence $K_A(u,v,\theta)$ and $K_B(u,v,\theta)$ are both Lipschitz continuous in v when $v\leq\tau_{U,2}$. Therefore, we can use Lemma 3.4.1 and the fact that $\sup_{\tau\in(0,\tau_{U,2}]}||\boldsymbol{\mu}\{\widehat{\boldsymbol{\alpha}}(\tau)\}-\boldsymbol{\mu}\{\boldsymbol{\alpha}_0(\tau)\}||\xrightarrow{p}0$ (Peng and Huang, 2008) to show

$$\sup_{\boldsymbol{z}\in\mathcal{Z},\boldsymbol{r},\boldsymbol{b},\tau}\Bigg|K_A\big\{\tau,\int_0^{\tau_{U,2}} I\{\boldsymbol{z}^T\boldsymbol{b}\geq\boldsymbol{z}^T\widehat{\boldsymbol{\alpha}}(u)\}\mathrm{d}u, g(\boldsymbol{z}^T\boldsymbol{r})\big\}$$
$$-K_A\big\{\tau,\int_0^{\tau_{U,2}} I\{\boldsymbol{z}^T\boldsymbol{b}\geq\boldsymbol{z}^T\boldsymbol{\alpha}_0(u)\}\mathrm{d}u, g(\boldsymbol{z}^T\boldsymbol{r})\big\}\Bigg|\xrightarrow{p}0.$$

It follows that

$$\sup_{\boldsymbol{\beta},\boldsymbol{r},\tau}||\boldsymbol{s}(\boldsymbol{\beta},\widehat{\boldsymbol{\alpha}},\boldsymbol{r},\tau)-\boldsymbol{s}(\boldsymbol{\beta},\boldsymbol{\alpha}_0,\boldsymbol{r},\tau)||\xrightarrow{p}0. \tag{3.11}$$

Next, we claim that $\mathcal{G}_1=\{\boldsymbol{Z}_iP_i(\boldsymbol{\beta},\boldsymbol{\alpha},\boldsymbol{r},\tau):\boldsymbol{Z}_i\in\mathcal{Z},\ \boldsymbol{\beta},\boldsymbol{\alpha}\in R^{p+1},\ \boldsymbol{r}\in R^{q+1},\ \tau\in(0,1)\}$ is Donsker thus Glivenko Cantelli. This follows by noting that the class of indicator functions is a VC-class, and by using the permanence properties of the Donsker class (Van der Vaart and Wellner, 1996). Therefore, Glivenko-Cantelli theorem gives

$$\sup_{\boldsymbol{\beta},\boldsymbol{\alpha},\tau,\boldsymbol{r}}||\boldsymbol{S}_n(\boldsymbol{\beta},\boldsymbol{\alpha},\boldsymbol{r},\tau)-\boldsymbol{s}(\boldsymbol{\beta},\boldsymbol{\alpha},\boldsymbol{r},\tau)||=o_p(1). \tag{3.12}$$

Since $\widetilde{\boldsymbol{\beta}}(\tau;\boldsymbol{r})$ is the root to $\boldsymbol{s}(\boldsymbol{\beta},\boldsymbol{\alpha}_0,\boldsymbol{r},\tau)=0$, we can use simple manipulations and get

$$||\boldsymbol{s}\{\widehat{\boldsymbol{\beta}}(\boldsymbol{r}), \boldsymbol{\alpha}_0, \boldsymbol{r}, \tau\} - \boldsymbol{s}\{\widetilde{\boldsymbol{\beta}}(\boldsymbol{r}), \boldsymbol{\alpha}_0, \boldsymbol{r}, \tau\}|| = ||\boldsymbol{s}\{\widehat{\boldsymbol{\beta}}(\boldsymbol{r}), \boldsymbol{\alpha}_0, \boldsymbol{r}, \tau\}||$$
$$\leq ||\boldsymbol{s}\{\widehat{\boldsymbol{\beta}}(\boldsymbol{r}), \widehat{\boldsymbol{\alpha}}, \boldsymbol{r}, \tau\}|| + ||\boldsymbol{s}\{\widehat{\boldsymbol{\beta}}(\boldsymbol{r}), \boldsymbol{\alpha}_0, \boldsymbol{r}, \tau\} - \boldsymbol{s}\{\widehat{\boldsymbol{\beta}}(\boldsymbol{r}), \widehat{\boldsymbol{\alpha}}, \boldsymbol{r}, \tau\}||. \quad (3.13)$$

Combining (3.12) and the fact that $\sup_{\tau\in[\nu_1,\tau_{U,1}]} ||\boldsymbol{S}_n\{\widehat{\boldsymbol{\beta}}(\tau;\boldsymbol{r}), \widehat{\boldsymbol{\alpha}}, \boldsymbol{r}, \tau\}|| = o(1)$, we get $\sup_{\tau\in[\nu_1,\tau_{U,1}]} ||\boldsymbol{s}\{\widehat{\boldsymbol{\beta}}(\tau;\boldsymbol{r}), \widehat{\boldsymbol{\alpha}}, \boldsymbol{r}, \tau\}|| \xrightarrow{p} 0$. Hence we can use (3.11) and (3.13) to get

$$\sup_{\boldsymbol{r},\tau\in[\nu_1,\tau_{U,1}]} ||\boldsymbol{s}\{\widehat{\boldsymbol{\beta}}(\boldsymbol{r}), \boldsymbol{\alpha}_0, \boldsymbol{r}, \tau\} - \boldsymbol{s}\{\widetilde{\boldsymbol{\beta}}(\boldsymbol{r}), \boldsymbol{\alpha}_0, \boldsymbol{r}, \tau\}|| \xrightarrow{p} 0. \quad (3.14)$$

By condition C4(i)–(iii), min eig$[-\boldsymbol{B}_b\{\widetilde{\boldsymbol{\beta}}(\boldsymbol{r}), \boldsymbol{\alpha}_0, \boldsymbol{r}, \tau\}]$ is bounded below by a positive constant k, uniformly for $\tau \in [\nu_1, \tau_{U,1}]$ and $\boldsymbol{r} \in \mathcal{R}(d_R)$, where ν_1 can be any constant in $(0, \tau_{U,1}]$. Therefore, one can combine the inverse function theorem and (3.14) to show that there exist a $\widehat{\boldsymbol{\beta}}(\tau, \boldsymbol{r}) \in \bar{\mathcal{B}}(d_B)$. Moreover, Taylor expansion of $\boldsymbol{s}\{\widehat{\boldsymbol{\beta}}(\boldsymbol{r}), \boldsymbol{\alpha}_0, \boldsymbol{r}, \tau\}$ around $\widetilde{\boldsymbol{\beta}}(\tau;\boldsymbol{r})$ gives

$$\sup_{\boldsymbol{r}\in\mathcal{R}(d_R),\tau\in[\nu_1,\tau_{U,1}]} ||\widehat{\boldsymbol{\beta}}(\tau;\boldsymbol{r}) - \widetilde{\boldsymbol{\beta}}(\tau;\boldsymbol{r})||$$
$$\leq \sup_{\boldsymbol{r}\in\mathcal{R}(d_R),\tau\in[\nu_1,\tau_{U,1}]} \frac{1}{k} ||\boldsymbol{s}\{\widehat{\boldsymbol{\beta}}(\boldsymbol{r}), \boldsymbol{\alpha}_0, \boldsymbol{r}, \tau\} - \boldsymbol{s}\{\widetilde{\boldsymbol{\beta}}(\boldsymbol{r}), \boldsymbol{\alpha}_0, \boldsymbol{r}, \tau\}|| \xrightarrow{p} 0. \quad (3.15)$$

This fact, combined with the Glivenko Cantelli theorem on $\boldsymbol{W}_n(\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{r})$, implies that $\sup_{\boldsymbol{r}\in\mathcal{R}(d_R)} ||\boldsymbol{W}_n(\widehat{\boldsymbol{\alpha}}, \boldsymbol{r}) - \widetilde{\boldsymbol{w}}(\widehat{\boldsymbol{\alpha}}, \boldsymbol{r})|| \xrightarrow{p} 0$, where $\boldsymbol{W}_n(\boldsymbol{\alpha}, \boldsymbol{r}) = \boldsymbol{W}_n\{\widehat{\boldsymbol{\beta}}(\boldsymbol{r}), \boldsymbol{\alpha}, \boldsymbol{r}\}$ and $\widetilde{\boldsymbol{w}}(\boldsymbol{\alpha}, \boldsymbol{r}) = \boldsymbol{w}\{\widetilde{\boldsymbol{\beta}}(\boldsymbol{r}), \boldsymbol{\alpha}, \boldsymbol{r}\}$. On the other hand, we can use Lemma 3.4.1 again to show $\sup_{\boldsymbol{r}\in\mathcal{R}(d_R)} ||\widetilde{\boldsymbol{w}}(\widehat{\boldsymbol{\alpha}}, \boldsymbol{r}) - \widetilde{\boldsymbol{w}}(\boldsymbol{\alpha}_0, \boldsymbol{r})|| \xrightarrow{p} 0$. It follows that

$$\sup_{r\in\mathcal{R}(d_R)} ||\boldsymbol{W}_n(\widehat{\boldsymbol{\alpha}}, \boldsymbol{r}) - \widetilde{\boldsymbol{w}}(\boldsymbol{\alpha}_0, \boldsymbol{r})|| \leq \sup_{r\in\mathcal{R}(d_R)} \left[||\boldsymbol{W}_n(\widehat{\boldsymbol{\alpha}}, \boldsymbol{r}) - \widetilde{\boldsymbol{w}}(\widehat{\boldsymbol{\alpha}}, \boldsymbol{r})|| + ||\widetilde{\boldsymbol{w}}(\widehat{\boldsymbol{\alpha}}, \boldsymbol{r}) - \widetilde{\boldsymbol{w}}(\boldsymbol{\alpha}_0, \boldsymbol{r})||\right] \xrightarrow{p} 0. \quad (3.16)$$

Therefore, we can see $\widetilde{\boldsymbol{w}}(\boldsymbol{\alpha}_0, \widehat{\boldsymbol{r}}) = o_p(1)$ from $\boldsymbol{W}_n(\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{r}}) = o_p(1)$. By regularity conditions C4(iv), $\boldsymbol{r}_0$ is the unique zero crossing of $\widetilde{\boldsymbol{w}}(\boldsymbol{\alpha}_0, \boldsymbol{r})$ in a neighborhood of $\boldsymbol{r}_0$, and

$\partial\widetilde{\boldsymbol{w}}(\boldsymbol{\alpha}_0, \boldsymbol{r})/\partial\boldsymbol{r}$ is of full rank for $\boldsymbol{r} \in \mathcal{R}(d_R)$. It follows that $\widehat{\boldsymbol{r}} \xrightarrow{p} \boldsymbol{r}_0$.

To complete the proof, we need to show the partial derivative of $\widetilde{\boldsymbol{\beta}}(\tau; \boldsymbol{r})$ with respect to $\boldsymbol{r}$ is bounded at $\boldsymbol{r} = \boldsymbol{r}_0$. Taking partial derivative $\partial/\partial\boldsymbol{r}$ on both sides of $\boldsymbol{s}\{\widetilde{\boldsymbol{\beta}}(\boldsymbol{r}), \boldsymbol{\alpha}_0, \boldsymbol{r}, \tau\} = 0$, we get

$$\boldsymbol{B}_b(\boldsymbol{\beta}_0, \boldsymbol{\alpha}_0, \boldsymbol{r}_0, \tau)\frac{\partial\widetilde{\boldsymbol{\beta}}(\tau; \boldsymbol{r})}{\partial\boldsymbol{r}}\Big|_{\boldsymbol{r}=\boldsymbol{r}_0} + \boldsymbol{B}_r(\boldsymbol{\beta}_0, \boldsymbol{\alpha}_0, \boldsymbol{r}_0, \tau) = 0.$$

By condition C3(ii), every component of $\boldsymbol{B}_r(\boldsymbol{\beta}_0, \boldsymbol{\alpha}_0, \boldsymbol{r}_0, \tau)$ is uniformly bounded for $\tau \in [\nu_1, \tau_{U,1}]$. Coupled with $\inf_{\tau\in[\nu_1,\tau_{U,1}]} \min \text{eig}\{-\boldsymbol{B}_b(\boldsymbol{\beta}_0, \boldsymbol{\alpha}_0, \boldsymbol{r}_0, \tau)\} > 0$, this implies

$$\frac{\partial\widetilde{\boldsymbol{\beta}}(\tau; \boldsymbol{r})}{\partial\boldsymbol{r}}\Big|_{\boldsymbol{r}=\boldsymbol{r}_0} = -\boldsymbol{B}_b(\boldsymbol{\beta}_0, \boldsymbol{\alpha}_0, \boldsymbol{r}_0, \tau)^{-1}\boldsymbol{B}_r(\boldsymbol{\beta}_0, \boldsymbol{\alpha}_0, \boldsymbol{r}_0, \tau) \tag{3.17}$$

and is uniformly bounded for $\tau \in [\nu_1, \tau_{U,1}]$. Therefore, we can show

$$\sup_{\tau\in[\nu_1,\tau_{U,1}]} ||\widetilde{\boldsymbol{\beta}}(\tau; \widehat{\boldsymbol{r}}) - \widetilde{\boldsymbol{\beta}}(\tau; \boldsymbol{r}_0)|| = o_p(1) \tag{3.18}$$

following the consistency of $\widehat{\boldsymbol{r}}$. Combining this with Equation (3.15) and the following inequality,

$$||\widehat{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}_0(\tau)|| = ||\widehat{\boldsymbol{\beta}}(\tau; \widehat{\boldsymbol{r}}) - \widetilde{\boldsymbol{\beta}}(\tau; \boldsymbol{r}_0)|| \leq ||\widehat{\boldsymbol{\beta}}(\tau; \widehat{\boldsymbol{r}}) - \widetilde{\boldsymbol{\beta}}(\tau; \widehat{\boldsymbol{r}})|| + ||\widetilde{\boldsymbol{\beta}}(\tau; \widehat{\boldsymbol{r}}) - \widetilde{\boldsymbol{\beta}}(\tau; \boldsymbol{r}_0)||,$$

we can show the uniform consistency of $\widehat{\boldsymbol{\beta}}(\tau)$ for $\tau \in [\nu_1, \tau_{U,1}]$. This completes the proof of Theorem 3.1.1.

3.4.3 Proof of Theorem 2

Lemma 3.4.2. *For any sequence* $\{\overline{\boldsymbol{\beta}}_n(\tau), \tau \in [\nu_1, \tau_{U,1}]\}_{n=1}^{\infty} \in \bar{\mathcal{B}}(d_B)$ *satisfying* $\sup_{\tau\in[\nu_1,\tau_{U,1}]} ||\boldsymbol{s}(\overline{\boldsymbol{\beta}}_n, \boldsymbol{\alpha}_0, \boldsymbol{r}_0, \tau) - \boldsymbol{s}(\boldsymbol{\beta}_0, \boldsymbol{\alpha}_0, \boldsymbol{r}_0, \tau)|| \xrightarrow{p} 0$, *we have*

$$\sup_{\tau\in[\nu_1,\tau_{U,1}]} n^{1/2}||\{\boldsymbol{S}_n(\overline{\boldsymbol{\beta}}_n, \boldsymbol{\alpha}_0, \boldsymbol{r}_0, \tau) - \boldsymbol{S}_n(\boldsymbol{\beta}_0, \boldsymbol{\alpha}_0, \boldsymbol{r}_0, \tau)\} - \{\boldsymbol{s}(\overline{\boldsymbol{\beta}}_n, \boldsymbol{\alpha}_0, \boldsymbol{r}_0, \tau) - \boldsymbol{s}(\boldsymbol{\beta}_0, \boldsymbol{\alpha}_0, \boldsymbol{r}_0, \tau)\}|| \xrightarrow{p} 0.$$

Similarly, we can show

$$\sup_{\tau\in[\nu_1,\tau_{U,1}]} n^{1/2}\Big|\Big|\boldsymbol{S}_n(\boldsymbol{\beta}_0, \overline{\boldsymbol{\alpha}}_n, \boldsymbol{r}_0, \tau) - \boldsymbol{S}_n(\boldsymbol{\beta}_0, \boldsymbol{\alpha}_0, \boldsymbol{r}_0, \tau) - \{\boldsymbol{s}(\boldsymbol{\beta}_0, \overline{\boldsymbol{\alpha}}_n, \boldsymbol{r}_0, \tau) - \boldsymbol{s}(\boldsymbol{\beta}_0, \boldsymbol{\alpha}_0, \boldsymbol{r}_0, \tau)\}\Big|\Big| \xrightarrow{p} 0$$

for any sequence $\{\overline{\boldsymbol{\alpha}}_n(\tau), \tau \in (0, \tau_{U,2}]\}_{n=1}^{\infty}$ *satisfying* $\sup_{\tau\in(0,\tau_{U,2}]} ||\boldsymbol{\mu}\{\overline{\boldsymbol{\alpha}}_n(\tau)\} - \boldsymbol{\mu}\{\boldsymbol{\alpha}_0(\tau)\}|| \xrightarrow{p} 0$.

Proof of Lemma 3.4.2. Define $\sigma_B^2(\boldsymbol{\beta}) = \sup_{\tau\in[\nu_1,\tau_{U,1}]} \mathrm{Var}\{P_i(\boldsymbol{\beta}, \boldsymbol{\alpha}_0, \boldsymbol{r}_0, \tau) - P_i(\boldsymbol{\beta}_0, \boldsymbol{\alpha}_0, \boldsymbol{r}_0, \tau)\}$, Following the arguments of Lai and Ying (1988) and Peng and Huang (2008), it is sufficient for the first part to hold if $\sigma_B^2(\overline{\boldsymbol{\beta}}_n) \xrightarrow{p} 0$. By condition C4(i)–(iii), one can use Taylor expansion on $\boldsymbol{s}(\overline{\boldsymbol{\beta}}_n, \boldsymbol{\alpha}_0, \boldsymbol{r}_0, \tau)$ around $\boldsymbol{\beta}_0(\tau)$ to show $\sup_{\tau\in[\nu_1,\tau_{U,1}]} ||\overline{\boldsymbol{\beta}}_n(\tau) - \boldsymbol{\beta}_0(\tau)|| \xrightarrow{p} 0$. Furthermore, the conditional density functions $f_X(t|\boldsymbol{z})$, $f_Y(t|\boldsymbol{z})$ and $f_{T_2}(t|\boldsymbol{z})$ are bounded above uniformly in t and $\boldsymbol{z}$, and $K_A(u, v, \theta)$ is Lipschitz continous in v. These facts allow us to mimic lemma B.1. in Peng and Huang (2008) and get $\sigma_B^2(\overline{\boldsymbol{\beta}}_n) \xrightarrow{p} 0$.

Similarly, define $\sigma_A^2(\boldsymbol{\alpha}) = \sup_{\tau\in[\nu_1,\tau_{U,1}]} \mathrm{Var}\{P_i(\boldsymbol{\beta}_0, \boldsymbol{\alpha}, \boldsymbol{r}_0, \tau) - P_i(\boldsymbol{\beta}_0, \boldsymbol{\alpha}_0, \boldsymbol{r}_0, \tau)\}$, a sufficient condition for the second part is $\sigma_A^2(\overline{\boldsymbol{\alpha}}_n) \xrightarrow{p} 0$, which follows directly from Lemma 3.4.1 and the Lipschitz continuity of $K_A(u, v, \theta)$ in v. This completes the proof of Lemma 3.4.2. □

A. Weak Convergence of $n^{1/2}[\boldsymbol{s}\{\widehat{\boldsymbol{\beta}}(\boldsymbol{r}_0), \boldsymbol{\alpha}_0, \boldsymbol{r}_0, \tau\} - \boldsymbol{s}(\boldsymbol{\beta}_0, \boldsymbol{\alpha}_0, \boldsymbol{r}_0, \tau)]$

By simple manipulations, we can show $n^{1/2}[\boldsymbol{S}_n\{\widehat{\boldsymbol{\beta}}(\boldsymbol{r}_0), \widehat{\boldsymbol{\alpha}}, \boldsymbol{r}_0, \tau\} - \boldsymbol{S}_n(\boldsymbol{\beta}_0, \boldsymbol{\alpha}_0, \boldsymbol{r}_0, \tau)]$

equals

$$n^{1/2}[\boldsymbol{S}_n(\boldsymbol{\beta}_0,\widehat{\boldsymbol{\alpha}},\boldsymbol{r}_0,\tau)-\boldsymbol{S}_n(\boldsymbol{\beta}_0,\boldsymbol{\alpha}_0,\boldsymbol{r}_0,\tau)]+n^{1/2}[\boldsymbol{S}_n\{\widehat{\boldsymbol{\beta}}(\boldsymbol{r}_0),\boldsymbol{\alpha}_0,\boldsymbol{r}_0,\tau\}-\boldsymbol{S}_n(\boldsymbol{\beta}_0,\boldsymbol{\alpha}_0,\boldsymbol{r}_0,\tau)]+\varepsilon_S \tag{3.19}$$

where $\varepsilon_S = n^{1/2}[\boldsymbol{S}_n\{\widehat{\boldsymbol{\beta}}(\boldsymbol{r}_0),\widehat{\boldsymbol{\alpha}},\boldsymbol{r}_0,\tau\} - \boldsymbol{S}_n(\boldsymbol{\beta}_0,\widehat{\boldsymbol{\alpha}},\boldsymbol{r}_0,\tau)] - n^{1/2}[\boldsymbol{S}_n\{\widehat{\boldsymbol{\beta}}(\boldsymbol{r}_0),\boldsymbol{\alpha}_0,\boldsymbol{r}_0,\tau\} - \boldsymbol{S}_n(\boldsymbol{\beta}_0,\boldsymbol{\alpha}_0,\boldsymbol{r}_0,\tau)]$. Based on the uniform consistency of $\boldsymbol{s}\{\widehat{\boldsymbol{\beta}}(\boldsymbol{r}_0),\boldsymbol{\alpha}_0,\boldsymbol{r}_0,\tau\}$ for $\tau \in [\nu_1,\tau_{U,1}]$ in Equation (3.14), and the uniform consistency of $\boldsymbol{\mu}\{\widehat{\boldsymbol{\alpha}}(\tau)\}$ for $\tau \in (0,\tau_{U,2}]$ in Peng and Huang (2008), we can use Lemma 3.4.2 to show

$$\begin{aligned} n^{1/2}[\boldsymbol{S}_n(\boldsymbol{\beta}_0,\widehat{\boldsymbol{\alpha}},\boldsymbol{r}_0,\tau) - \boldsymbol{S}_n(\boldsymbol{\beta}_0,\boldsymbol{\alpha}_0,\boldsymbol{r}_0,\tau)] &= n^{1/2}[\boldsymbol{s}(\boldsymbol{\beta}_0,\widehat{\boldsymbol{\alpha}},\boldsymbol{r}_0,\tau) - \boldsymbol{s}(\boldsymbol{\beta}_0,\boldsymbol{\alpha}_0,\boldsymbol{r}_0,\tau)] + o_p^{\tau\in[\nu_1,\tau_{U,1}]}(1), \\ n^{1/2}[\boldsymbol{S}_n\{\widehat{\boldsymbol{\beta}}(\boldsymbol{r}_0),\boldsymbol{\alpha}_0,\boldsymbol{r}_0,\tau\} - \boldsymbol{S}_n(\boldsymbol{\beta}_0,\boldsymbol{\alpha}_0,\boldsymbol{r}_0,\tau)] &= n^{1/2}[\boldsymbol{s}\{\widehat{\boldsymbol{\beta}}(\boldsymbol{r}_0),\boldsymbol{\alpha}_0,\boldsymbol{r}_0,\tau\} \\ &\quad - \boldsymbol{s}(\boldsymbol{\beta}_0,\boldsymbol{\alpha}_0,\boldsymbol{r}_0,\tau)] + o_p^{\tau\in[\nu_1,\tau_{U,1}]}(1). \end{aligned} \tag{3.20}$$

Here and in the sequel, $o_p^{\mathbf{S}}(1)$ means convergence to 0 in probability uniformly on set $\mathbf{S}$. Similarly, we use $o_p^{\mathbf{S}}(n^{-1/2})$ to denote uniform root n convergence to 0 in probability on set $\mathbf{S}$. The remainder term ε_S can be shown to be $o_p^{\tau\in[\nu_1,\tau_{U,1}]}(1)$ by similar arguments.

According to Peng and Huang (2008),

$$n^{1/2}[\boldsymbol{\mu}\{\widehat{\boldsymbol{\alpha}}(u)\} - \boldsymbol{\mu}\{\boldsymbol{\alpha}_0(u)\}] = -n^{1/2}\sum_{i=1}^{n}\phi\{\boldsymbol{Z}_i R_i^Y(\boldsymbol{\alpha}_0,u)\} + o_p^{u\in(0,\tau_{U,2}]}(1), \tag{3.21}$$

where $R_i^Y(\boldsymbol{\alpha},u) = I[Y_i \le \exp\{\boldsymbol{Z}_i^T\boldsymbol{\alpha}(u)\},\eta = 1] - \int_0^u I[Y_i \ge \exp\{\boldsymbol{Z}_i^T\boldsymbol{\alpha}(t)\}]\mathrm{d}H(t)$, $H(t) = -\log(1-t)$ for $0 \le t < 1$, and ϕ is a continuous linear map that involves product integration. The detailed form of ϕ can be found in Appendix B of Peng and Huang (2008). Using empirical process techniques, the authors showed that $-n^{1/2}\sum_{i=1}^n \phi\{\boldsymbol{Z}_i R_i^Y(\boldsymbol{\alpha}_0,u)\}$ converges weakly to a Gaussian process for $u \in (0,\tau_{U,2}]$. With these facts, now we look at the asymptotic behavior of $n^{1/2}\big[\boldsymbol{s}(\boldsymbol{\beta}_0,\widehat{\boldsymbol{\alpha}},\boldsymbol{r}_0,\tau) - \boldsymbol{s}(\boldsymbol{\beta}_0,\boldsymbol{\alpha}_0,\boldsymbol{r}_0,\tau)\big]$.

By Lemma 3.4.1 and the Lipschitz continuity of $K_A(u,v,\theta)$ in v, we can use Taylor expansion to see

$$
\begin{aligned}
&n^{1/2}\{\boldsymbol{s}(\boldsymbol{\beta}_0,\widehat{\boldsymbol{\alpha}},\boldsymbol{r}_0,\tau)-\boldsymbol{s}(\boldsymbol{\beta}_0,\boldsymbol{\alpha}_0,\boldsymbol{r}_0,\tau)\}\\
&=-n^{1/2}E\Big\{\boldsymbol{Z}I\{\log Y>\boldsymbol{Z}^T\boldsymbol{\beta}_0(\tau)\}\times d_A\big\{\tau,F_2[\exp\{\boldsymbol{Z}^T\boldsymbol{\beta}_0(\tau)\}|\boldsymbol{Z}],g(\boldsymbol{Z}^T\boldsymbol{r}_0)\big\}\times\\
&\quad[\int_0^{\tau_{U,2}}I\{\boldsymbol{Z}^T\boldsymbol{\beta}_0(\tau)\geq\boldsymbol{Z}^T\widehat{\boldsymbol{\alpha}}(u)\}\mathrm{d}u-\int_0^{\tau_{U,2}}I\{\boldsymbol{Z}^T\boldsymbol{\beta}_0(\tau)\geq\boldsymbol{Z}^T\boldsymbol{\alpha}_0(u)\}\mathrm{d}u\big]\Big\}+o_p^{\tau\in[\nu_1,\tau_{U,1}]}(1),
\end{aligned}
\tag{3.22}
$$

where $d_A(u,v,\theta)=\partial K_A(u,v,\theta)/\partial v$. We will proceed to show that $n^{1/2}\{\boldsymbol{s}(\boldsymbol{\beta}_0,\widehat{\boldsymbol{\alpha}},\boldsymbol{r}_0,\tau)-\boldsymbol{s}(\boldsymbol{\beta}_0,\boldsymbol{\alpha}_0,\boldsymbol{r}_0,\tau)\}$ converges weakly to a Gaussian process with mean 0, and present the specific form of its influence functions.

First consider the one-sample case, where $\exp\{\widehat{\boldsymbol{\alpha}}(u)\}$, $u\in(0,\tau_{U,2}]$, is asymptotically equivalent to the uth quantile of the Nelson Aalen estimator (Peng and Huang, 2008). Specifically, it can be shown that $\widehat{F}_2^{NA}[\exp\{\widehat{\boldsymbol{\alpha}}(u)\}]=u+o_p^{u\in(0,\tau_{U,2}]}(n^{-1/2})$, with $\widehat{F}_2^{NA}[\cdot]$ representing the Nelson Aalen estimator of T_2's distribution function. Therefore, we get

$$
\begin{aligned}
\int_0^{\tau_{U,2}}I\{\boldsymbol{\beta}_0(\tau)\geq\widehat{\boldsymbol{\alpha}}(u)\}\mathrm{d}u&=\int_0^{\tau_{U,2}}I\big(\widehat{F}_2^{NA}[\exp\{\boldsymbol{\beta}_0(\tau)\}]\geq u\big)\mathrm{d}u+o_p(n^{-1/2})\\
&=\widehat{F}_2^{NA}[\exp\{\boldsymbol{\beta}_0(\tau)\}]+o_p(n^{-1/2}),
\end{aligned}
$$

where $F_2[\exp\{\boldsymbol{\beta}_0(\tau)\}]\leq\tau_{U,2}$ by regularity condition C5(i). It is well-known that $n^{1/2}\{\widehat{F}_2^{NA}(t)-F_2(t)\}$ is asymptotically Gaussian and can be written as $n^{-1/2}\sum_{i=1}^n\pi_i(t)+o(1)$, where $\pi_i(t)$ are i.i.d. and form a Donsker's class with mean 0 (Kosorok, 2008). Facilitated by that, we can see the right-hand-side of Equation (3.22) equals

$$
-n^{-1/2}\sum_{i=1}^n d_A\big\{\tau,F_2[\exp\{\boldsymbol{\beta}_0(\tau)\}],g(\boldsymbol{r}_0)\big\}\pi_i[\exp\{\boldsymbol{\beta}_0(\tau)\}]\Pr\{\log Y>\boldsymbol{\beta}_0(\tau)\}+o_p^{\tau\in[\nu_1,\tau_{U,1}]}(1)
\tag{3.23}
$$

in the one-sample case. The arguments can be easily extended to the K-sample case, where $\mathcal{Z} = \{\boldsymbol{z}_1, \boldsymbol{z}_2, ..., \boldsymbol{z}_K\}$ and $\boldsymbol{Z}_i = \boldsymbol{z}_k$ if and only if observation i belongs to the k_{th} group. Similar to Equation (3.23), we can show the right-hand-side of Equation (3.22) equals

$$-n^{-1/2}\sum_{i=1}^{n}\sum_{k=1}^{K}\Big[\boldsymbol{z}_k d_A\big\{\tau, F_2[\exp\{\boldsymbol{z}_k^T\boldsymbol{\beta}_0(\tau)\}|\boldsymbol{z}_k], g(\boldsymbol{z}_k^T\boldsymbol{r}_0)\big\}\pi_{k,i}[\exp\{\boldsymbol{z}_k^T\boldsymbol{\beta}_0(\tau)\}]\times \Pr\{\log Y > \boldsymbol{z}_k^T\boldsymbol{\beta}_0(\tau)|\boldsymbol{z}_k\}\Big] + o_p^{\tau\in[\nu_1,\tau_{U,1}]}(1), \quad (3.24)$$

where $\pi_{k,i}(t)$ is the influence function for $\widehat{F}_{2,k}^{NA}(t)$, the Nelson-Aalen estimator within the k_{th} group, and equals 0 when $\boldsymbol{Z}_i \neq \boldsymbol{z}_k$. By the boundedness of $d_A(\cdot)$, $\mathcal{Z}$ and $\Pr\{\log Y > \boldsymbol{z}_k^T\boldsymbol{\beta}_0(\tau)|\boldsymbol{z}_k\}$, we can show $\sum_{k=1}^K \boldsymbol{z}_k d_A\big\{\tau, F_2[\exp\{\boldsymbol{z}_k^T\boldsymbol{\beta}_0(\tau)\}|\boldsymbol{z}_k], g(\boldsymbol{z}_k^T\boldsymbol{r}_0)\big\}\times$ $\Pr\{\log Y > \boldsymbol{z}_k^T\boldsymbol{\beta}_0(\tau)|\boldsymbol{z}_k\}\pi_{k,i}[\exp\{\boldsymbol{z}_k^T\boldsymbol{\beta}_0(\tau)\}]$ is also mean 0 and form a Donsker's class, as Donsker's property is preserved under Lipschitz transformations. It follows that (3.24) converges weakly to a mean 0 Gaussian process, by properties of Donsker's class.

Next consider the case when $\boldsymbol{Z}$ involves continous components. Define $\boldsymbol{v}_A(\boldsymbol{a}, \tau) = E\boldsymbol{V}_A(\boldsymbol{a}, \tau)$ and $\boldsymbol{J}_A(\boldsymbol{a}, \tau) = \partial \boldsymbol{v}_A(\boldsymbol{a}, \tau)/\partial \boldsymbol{a}$, where $\boldsymbol{V}_A(\boldsymbol{a}, \tau)$ is defined in condition C2(iii) in Appendix A. From condition C2(iii), each component of $\boldsymbol{J}_A\{\boldsymbol{\alpha}_0(u), \tau\}\boldsymbol{B}_a^{-1}\{\boldsymbol{\alpha}_0(u)\}$ is uniformly bounded for any $u \in (0, \tau_{U,2}]$ and $\tau \in [\nu_1, \tau_{U,1}]$. Therefore, we can combine Equation (3.21) and (3.22) to get

$$\begin{aligned} & n^{1/2}\{\boldsymbol{s}(\boldsymbol{\beta}_0, \widehat{\boldsymbol{\alpha}}, \boldsymbol{r}_0, \tau) - \boldsymbol{s}(\boldsymbol{\beta}_0, \boldsymbol{\alpha}_0, \boldsymbol{r}_0, \tau)\} \\ = & \int_0^{\tau_{U,2}} n^{1/2}[\boldsymbol{v}_A\{\widehat{\boldsymbol{\alpha}}(u), \tau\} - \boldsymbol{v}_A\{\boldsymbol{\alpha}_0(u), \tau\}]\mathrm{d}u + o_p^{\tau\in[\nu_1,\tau_{U,1}]}(1) \\ = & -n^{-1/2}\sum_{i=1}^{n}\int_0^{\tau_{U,2}} \boldsymbol{J}_A\{\boldsymbol{\alpha}_0(u), \tau\}\boldsymbol{B}_a^{-1}\{\boldsymbol{\alpha}_0(u)\}\phi\{\boldsymbol{Z}_i R_i^Y(\boldsymbol{\alpha}_0, u)\}\mathrm{d}u + o_p^{\tau\in[\nu_1,\tau_{U,1}]}(1). \end{aligned} \quad (3.25)$$

It is not hard to show the right-hand-side converges weakly to a mean 0 Gaussian process by the boundedness of $\boldsymbol{J}_A\{\boldsymbol{\alpha}_0(u),\tau\}\boldsymbol{B}_a^{-1}\{\boldsymbol{\alpha}_0(u)\}$ and the linearity of the integral operator. Therefore, we can see that $n^{1/2}[\boldsymbol{s}(\boldsymbol{\beta}_0,\widehat{\boldsymbol{\alpha}},\boldsymbol{r}_0,\tau)-\boldsymbol{s}(\boldsymbol{\beta}_0,\boldsymbol{\alpha}_0,\boldsymbol{r}_0,\tau)]$ is asymptotically equivalent to sum of i.i.d. influence functions and converges weakly to Gaussian, no matter $\boldsymbol{Z}$ is discrete or not. In the sequel we will unify (3.24) and (3.25) and write

$$n^{1/2}\{\boldsymbol{s}(\boldsymbol{\beta}_0,\widehat{\boldsymbol{\alpha}},\boldsymbol{r}_0,\tau)-\boldsymbol{s}(\boldsymbol{\beta}_0,\boldsymbol{\alpha}_0,\boldsymbol{r}_0,\tau)\}=n^{-1/2}\sum_{i=1}^{n}\boldsymbol{\rho}_i(\boldsymbol{\beta}_0,\boldsymbol{\alpha}_0,\boldsymbol{r}_0,\tau)+o_p^{\tau\in[\nu_1,\tau_{U,1}]}(1). \tag{3.26}$$

Now by $\boldsymbol{S}_n\{\widehat{\boldsymbol{\beta}}(\boldsymbol{r}_0),\widehat{\boldsymbol{\alpha}},\boldsymbol{r}_0,\tau\}=o_p^{\tau\in[\nu_1,\tau_{U,1}]}(n^{-1/2})$, we can see from Equation (3.19) and (3.20) that

$$\begin{aligned}&n^{1/2}[\boldsymbol{s}\{\widehat{\boldsymbol{\beta}}(\boldsymbol{r}_0),\boldsymbol{\alpha}_0,\boldsymbol{r}_0,\tau\}-\boldsymbol{s}(\boldsymbol{\beta}_0,\boldsymbol{\alpha}_0,\boldsymbol{r}_0,\tau)]\\&\quad=-n^{1/2}\boldsymbol{S}_n(\boldsymbol{\beta}_0,\boldsymbol{\alpha}_0,\boldsymbol{r}_0,\tau)-n^{1/2}\{\boldsymbol{s}(\boldsymbol{\beta}_0,\widehat{\boldsymbol{\alpha}},\boldsymbol{r}_0,\tau)-\boldsymbol{s}(\boldsymbol{\beta}_0,\boldsymbol{\alpha}_0,\boldsymbol{r}_0,\tau)\}+o_p^{\tau\in[\nu_1,\tau_{U,1}]}(1),\end{aligned}$$

where $-n^{1/2}\boldsymbol{S}_n(\boldsymbol{\beta}_0,\boldsymbol{\alpha}_0,\boldsymbol{r}_0,\tau)$ converges weakly to Gaussian by properties of the Donsker's class (Van der Vaart and Wellner, 1996). Moreover, we can combine Equation (3.26) and get

$$n^{1/2}[\boldsymbol{s}\{\widehat{\boldsymbol{\beta}}(\boldsymbol{r}_0),\boldsymbol{\alpha}_0,\boldsymbol{r}_0,\tau\}-\boldsymbol{s}(\boldsymbol{\beta}_0,\boldsymbol{\alpha}_0,\boldsymbol{r}_0,\tau)]=n^{-1/2}\sum_{i=1}^{n}\boldsymbol{\chi}_i(\boldsymbol{\beta}_0,\boldsymbol{\alpha}_0,\boldsymbol{r}_0,\tau)+o_p^{[\nu_1,\tau_{U,1}]}(1), \tag{3.27}$$

where $\boldsymbol{\chi}_i(\boldsymbol{\beta}_0,\boldsymbol{\alpha}_0,\boldsymbol{r}_0,\tau)=-\boldsymbol{Z}_iP_i(\boldsymbol{\beta}_0,\boldsymbol{\alpha}_0,\boldsymbol{r}_0,\tau)-\boldsymbol{\rho}_i(\boldsymbol{\beta}_0,\boldsymbol{\alpha}_0,\boldsymbol{r}_0,\tau)$ and $n^{-1/2}\sum_{i=1}^{n}\boldsymbol{\chi}_i(\boldsymbol{\beta}_0,\boldsymbol{\alpha}_0,\boldsymbol{r}_0,\tau)$ converges weakly to a Gaussian process with mean 0.

B. Weak Convergence of $n^{1/2}\boldsymbol{W}_n\{\widehat{\boldsymbol{\beta}}(\boldsymbol{r}_0),\widehat{\boldsymbol{\alpha}},\boldsymbol{r}_0\}$

Using similar arguments as Equation (3.19) and Lemma 3.4.2, we have

$$\begin{aligned}
&n^{1/2}[\boldsymbol{W}_n\{\widehat{\boldsymbol{\beta}}(\boldsymbol{r}_0),\widehat{\boldsymbol{\alpha}},\boldsymbol{r}_0\}-\boldsymbol{W}_n(\boldsymbol{\beta}_0,\boldsymbol{\alpha}_0,\boldsymbol{r}_0)]\\
=&n^{1/2}[\boldsymbol{W}_n(\boldsymbol{\beta}_0,\widehat{\boldsymbol{\alpha}},\boldsymbol{r}_0)-\boldsymbol{W}_n(\boldsymbol{\beta}_0,\boldsymbol{\alpha}_0,\boldsymbol{r}_0)]+n^{1/2}[\boldsymbol{W}_n\{\widehat{\boldsymbol{\beta}}(\boldsymbol{r}_0),\boldsymbol{\alpha}_0,\boldsymbol{r}_0\}-\boldsymbol{W}_n(\boldsymbol{\beta}_0,\boldsymbol{\alpha}_0,\boldsymbol{r}_0)]+\varepsilon_W\\
=&n^{1/2}[\boldsymbol{w}(\boldsymbol{\beta}_0,\widehat{\boldsymbol{\alpha}},\boldsymbol{r}_0)-\boldsymbol{w}(\boldsymbol{\beta}_0,\boldsymbol{\alpha}_0,\boldsymbol{r}_0)]+n^{1/2}[\boldsymbol{w}\{\widehat{\boldsymbol{\beta}}(\boldsymbol{r}_0),\boldsymbol{\alpha}_0,\boldsymbol{r}_0\}-\boldsymbol{w}(\boldsymbol{\beta}_0,\boldsymbol{\alpha}_0,\boldsymbol{r}_0)]+o_p(1)+\varepsilon_W,
\end{aligned} \tag{3.28}$$

where $\varepsilon_W=n^{1/2}\big[\boldsymbol{W}_n\{\widehat{\boldsymbol{\beta}}(\boldsymbol{r}_0),\widehat{\boldsymbol{\alpha}},\boldsymbol{r}_0\}-\boldsymbol{W}_n\{\boldsymbol{\beta}_0,\widehat{\boldsymbol{\alpha}},\boldsymbol{r}_0\}\big]-n^{1/2}\big[\boldsymbol{W}_n\{\widehat{\boldsymbol{\beta}}(\boldsymbol{r}_0),\boldsymbol{\alpha}_0,\boldsymbol{r}_0\}-\boldsymbol{W}_n(\boldsymbol{\beta}_0,\boldsymbol{\alpha}_0,\boldsymbol{r}_0)\big]$ can also be shown to be $o_p(1)$. It is easy to see $n^{1/2}\boldsymbol{W}_n(\boldsymbol{\beta}_0,\boldsymbol{\alpha}_0,\boldsymbol{r}_0)$ converges weakly to a Gaussian process by properties of Donsker's class. Moreover, we can follow the arguments in (3.24) and (3.25) to show

$$n^{1/2}\{\boldsymbol{w}(\boldsymbol{\beta}_0,\widehat{\boldsymbol{\alpha}},\boldsymbol{r}_0)-\boldsymbol{w}(\boldsymbol{\beta}_0,\boldsymbol{\alpha}_0,\boldsymbol{r}_0)\}=n^{-1/2}\sum_{i=1}^{n}\boldsymbol{\kappa}_i(\boldsymbol{\beta}_0,\boldsymbol{\alpha}_0,\boldsymbol{r}_0)+o_p(1). \tag{3.29}$$

The i.i.d. influence functions $\boldsymbol{\kappa}_i(\boldsymbol{\beta}_0,\boldsymbol{\alpha}_0,\boldsymbol{r}_0)$ take the form

$$\begin{aligned}
-\sum_{k=1}^{K}\int_{\tau_a}^{\tau_b}\Bigg[&\boldsymbol{z}_k d_B\big\{\tau,F_2[2\exp\{\boldsymbol{z}_k^T\boldsymbol{\beta}_0(\tau)\}|\boldsymbol{z}_k],g(\boldsymbol{z}_k^T\boldsymbol{r}_0)\big\}\times\\
&\Pr[Y>2\exp\{\boldsymbol{z}_k^T\boldsymbol{\beta}_0(\tau)\}|\boldsymbol{z}_k]\pi_{k,i}[2\exp\{\boldsymbol{z}_k^T\boldsymbol{\beta}_0(\tau)\}]\mathrm{d}\tau\Bigg],
\end{aligned} \tag{3.30}$$

in the K-sample case, and equals

$$-\int_0^{\tau_{U,2}}\boldsymbol{J}_B\{\boldsymbol{\alpha}_0(u)\}\boldsymbol{B}_a^{-1}\{\boldsymbol{\alpha}_0(u)\}\phi\{\boldsymbol{Z}_iR_i^Y(\boldsymbol{\alpha}_0,u)\}\mathrm{d}u \tag{3.31}$$

when $\boldsymbol{Z}$ contains continuous components, where $\boldsymbol{J}_B\{\boldsymbol{a}\}=\partial E\{\boldsymbol{V}_B(\boldsymbol{a})\}/\partial\boldsymbol{a}$ and $\boldsymbol{V}_B(\boldsymbol{a})$ was defined in regularity condition C2(iii). Following that, it is not hard to show $n^{1/2}\{\boldsymbol{w}(\boldsymbol{\beta}_0,\widehat{\boldsymbol{\alpha}},\boldsymbol{r}_0)-\boldsymbol{w}(\boldsymbol{\beta}_0,\boldsymbol{\alpha}_0,\boldsymbol{r}_0)\}$ converges weakly to a mean 0 normal distribution.

On the other hand, we can use regularity condition C4(iv) and Taylor expansion to show

$$
\begin{aligned}
& n^{1/2}[\boldsymbol{w}\{\widehat{\boldsymbol{\beta}}(\boldsymbol{r}_0),\boldsymbol{\alpha}_0,\boldsymbol{r}_0\}-\boldsymbol{w}(\boldsymbol{\beta}_0,\boldsymbol{\alpha}_0,\boldsymbol{r}_0)] \\
&= \int_{\tau_a}^{\tau_b}\Bigg[\boldsymbol{L}_{\boldsymbol{b}}\{\boldsymbol{\beta}_0(\tau),\boldsymbol{\alpha}_0,\boldsymbol{r}_0,\tau\}\boldsymbol{B}_b\{\boldsymbol{\beta}_0(\tau),\boldsymbol{\alpha}_0,\boldsymbol{r}_0,\tau\}^{-1}\times \\
&\qquad n^{1/2}[\boldsymbol{s}\{\widehat{\boldsymbol{\beta}}(\boldsymbol{r}_0),\boldsymbol{\alpha}_0,\boldsymbol{r}_0,\tau\}-\boldsymbol{s}(\boldsymbol{\beta}_0,\boldsymbol{\alpha}_0,\boldsymbol{r}_0,\tau)]\Bigg]\mathrm{d}\tau+o_p(1). \qquad (3.32)
\end{aligned}
$$

Combining Equation (3.27), (3.28), (3.29) and (3.32), we can show

$$
n^{1/2}\boldsymbol{W}_n\{\widehat{\boldsymbol{\beta}}(\boldsymbol{r}_0),\widehat{\boldsymbol{\alpha}},\boldsymbol{r}_0\}=n^{-1/2}\sum_{i=1}^{n}\boldsymbol{\iota}_i(\boldsymbol{\beta}_0,\boldsymbol{\alpha}_0,\boldsymbol{r}_0)+o_p(1) \qquad (3.33)
$$

and converges weakly to normal with mean 0, where $\boldsymbol{\iota}_i(\boldsymbol{\beta}_0,\boldsymbol{\alpha}_0,\boldsymbol{r}_0)$ equals $\bar{\boldsymbol{Z}}_iQ_i(\boldsymbol{\beta}_0,\boldsymbol{\alpha}_0,\boldsymbol{r}_0)+$ $\boldsymbol{\kappa}_i(\boldsymbol{\beta}_0,\boldsymbol{\alpha}_0,\boldsymbol{r}_0)+\int_{\tau_a}^{\tau_b}\boldsymbol{L}_{\boldsymbol{b}}\{\boldsymbol{\beta}_0(\tau),\boldsymbol{\alpha}_0,\boldsymbol{r}_0,\tau\}\boldsymbol{B}_b\{\boldsymbol{\beta}_0(\tau),\boldsymbol{\alpha}_0,\boldsymbol{r}_0,\tau\}^{-1}\boldsymbol{\chi}_i(\boldsymbol{\beta}_0,\boldsymbol{\alpha}_0,\boldsymbol{r}_0,\tau)\mathrm{d}\tau$.

C. Asymptotic Normality of $n^{1/2}(\widehat{\boldsymbol{r}}-\boldsymbol{r}_0)$ and Weak Convergence of $n^{1/2}\big\{\widehat{\boldsymbol{\beta}}(\tau)-\boldsymbol{\beta}_0(\tau)\big\}$

Using $\boldsymbol{s}\{\widetilde{\boldsymbol{\beta}}(\widehat{\boldsymbol{r}}),\boldsymbol{\alpha}_0,\widehat{\boldsymbol{r}},\tau\}=\boldsymbol{s}(\boldsymbol{\beta}_0,\boldsymbol{\alpha}_0,\boldsymbol{r}_0,\tau)=0$, we can obtain the following inequality

$$
\begin{aligned}
&||\boldsymbol{s}\{\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{r}}),\boldsymbol{\alpha}_0,\boldsymbol{r}_0,\tau\}-\boldsymbol{s}\{\widehat{\boldsymbol{\beta}}(\boldsymbol{r}_0),\boldsymbol{\alpha}_0,\boldsymbol{r}_0,\tau\}||\le\Bigg[||\boldsymbol{s}\{\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{r}}),\boldsymbol{\alpha}_0,\boldsymbol{r}_0,\tau\}-\boldsymbol{s}\{\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{r}}),\boldsymbol{\alpha}_0,\widehat{\boldsymbol{r}},\tau\}|| \\
&+||\boldsymbol{s}\{\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{r}}),\boldsymbol{\alpha}_0,\widehat{\boldsymbol{r}},\tau\}-\boldsymbol{s}\{\widetilde{\boldsymbol{\beta}}(\widehat{\boldsymbol{r}}),\boldsymbol{\alpha}_0,\widehat{\boldsymbol{r}},\tau\}||+||\boldsymbol{s}\{\widehat{\boldsymbol{\beta}}(\boldsymbol{r}_0),\boldsymbol{\alpha}_0,\boldsymbol{r}_0,\tau\}-\boldsymbol{s}(\boldsymbol{\beta}_0,\boldsymbol{\alpha}_0,\boldsymbol{r}_0,\tau)||\Bigg].
\end{aligned}
$$

According to regularity condition C3(ii), we can use $\widehat{\boldsymbol{r}}\xrightarrow{p}\boldsymbol{r}_0$ to get $\sup_{\tau\in[\nu_1,\tau_{U,1}]}||\boldsymbol{s}\{\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{r}}),\boldsymbol{\alpha}_0,\boldsymbol{r}_0,\tau\}-\boldsymbol{s}\{\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{r}}),\boldsymbol{\alpha}_0,\widehat{\boldsymbol{r}},\tau\}||\xrightarrow{p}0$. When combined with the above inequality and Equation (3.14), this implies $\sup_{\tau\in[\nu_1,\tau_{U,1}]}||\boldsymbol{s}\{\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{r}}),\boldsymbol{\alpha}_0,\boldsymbol{r}_0,\tau\}-\boldsymbol{s}\{\widehat{\boldsymbol{\beta}}(\boldsymbol{r}_0),\boldsymbol{\alpha}_0,\boldsymbol{r}_0,\tau\}||\xrightarrow{p}0$. Hence we can use regularity condition C4(ii) and Taylor expansion to show $\sup_{\boldsymbol{z}\in\mathcal{Z},\tau\in[\nu_1,\tau_{U,1}]}||\widehat{\boldsymbol{\beta}}(\tau;\widehat{\boldsymbol{r}})-\widehat{\boldsymbol{\beta}}(\tau;\boldsymbol{r}_0)||\xrightarrow{p}0$. Therefore, we can combine regularity condition C3(ii) and mimic the arguments in Lemma 3.4.2 to show

$$\sup_{\tau\in[\nu_1,\tau_{U,1}]} \mathrm{Var}[P_i\{\widehat{\boldsymbol\beta}(\widehat{\boldsymbol r}),\boldsymbol\alpha_0,\widehat{\boldsymbol r},\tau\} - P_i\{\widehat{\boldsymbol\beta}(\boldsymbol r_0),\boldsymbol\alpha_0,\boldsymbol r_0,\tau\}] \xrightarrow{p} 0$$

and

$$\sup_{\tau\in[\nu_1,\tau_{U,1}]} \Big|\Big|\boldsymbol S_n\{\widehat{\boldsymbol\beta}(\widehat{\boldsymbol r}),\boldsymbol\alpha_0,\widehat{\boldsymbol r},\tau\} - \boldsymbol S_n\{\widehat{\boldsymbol\beta}(\boldsymbol r_0),\boldsymbol\alpha_0,\boldsymbol r_0,\tau\} \\ - \boldsymbol s\{\widehat{\boldsymbol\beta}(\widehat{\boldsymbol r}),\boldsymbol\alpha_0,\widehat{\boldsymbol r},\tau\} + \boldsymbol s\{\widehat{\boldsymbol\beta}(\boldsymbol r_0),\boldsymbol\alpha_0,\boldsymbol r_0,\tau\}\Big|\Big| = o_p(n^{-1/2}). \tag{3.34}$$

On the other hand, we can show by similar arguments that

$$\sup_{\tau\in[\nu_1,\tau_{U,1}]} \Big|\Big|\boldsymbol S_n\{\widehat{\boldsymbol\beta}(\widehat{\boldsymbol r}),\widehat{\boldsymbol\alpha},\widehat{\boldsymbol r},\tau\} - \boldsymbol S_n\{\widehat{\boldsymbol\beta}(\boldsymbol r_0),\widehat{\boldsymbol\alpha},\boldsymbol r_0,\tau\} \\ - \boldsymbol S_n\{\widehat{\boldsymbol\beta}(\widehat{\boldsymbol r}),\boldsymbol\alpha_0,\widehat{\boldsymbol r},\tau\} + \boldsymbol S_n\{\widehat{\boldsymbol\beta}(\boldsymbol r_0),\boldsymbol\alpha_0,\boldsymbol r_0,\tau\}\Big|\Big| = o_p(n^{-1/2}).$$

Noting that $\sup_{\tau\in[\nu_1,\tau_{U,1}]} \boldsymbol S_n\{\widehat{\boldsymbol\beta}(\boldsymbol r),\widehat{\boldsymbol\alpha},\boldsymbol r,\tau\} = o(n^{-1/2})$, we have

$$\sup_{\tau\in[\nu_1,\tau_{U,1}]} ||\boldsymbol S_n\{\widehat{\boldsymbol\beta}(\widehat{\boldsymbol r}),\boldsymbol\alpha_0,\widehat{\boldsymbol r},\tau\} - \boldsymbol S_n\{\widehat{\boldsymbol\beta}(\boldsymbol r_0),\boldsymbol\alpha_0,\boldsymbol r_0,\tau\}|| = o_p(n^{-1/2}).$$

Combined with Equation (3.34), this further implies

$$\sup_{\tau\in[\nu_1,\tau_{U,1}]} ||\boldsymbol s\{\widehat{\boldsymbol\beta}(\widehat{\boldsymbol r}),\boldsymbol\alpha_0,\widehat{\boldsymbol r},\tau\} - \boldsymbol s\{\widehat{\boldsymbol\beta}(\boldsymbol r_0),\boldsymbol\alpha_0,\boldsymbol r_0,\tau\}|| = o_p(n^{-1/2}).$$

By the uniform convergence of $\widehat{\boldsymbol\beta}(\tau;\boldsymbol r_0)$ to $\boldsymbol\beta_0(\tau)$ in $\tau\in[\nu_1,\tau_{U,1}]$, we can use Taylor expansion and get

$$\sup_{\tau\in[\nu_1,\tau_{U,1}]} ||\widehat{\boldsymbol\beta}(\tau;\widehat{\boldsymbol r}) - \widehat{\boldsymbol\beta}(\tau;\boldsymbol r_0) + \boldsymbol B_b(\boldsymbol\beta_0,\boldsymbol\alpha_0,\boldsymbol r_0,\tau)^{-1}\boldsymbol B_r(\boldsymbol\beta_0,\boldsymbol\alpha_0,\boldsymbol r_0,\tau)(\widehat{\boldsymbol r}-\boldsymbol r_0)|| = o_p(n^{-1/2}\vee||\widehat{\boldsymbol r}-\boldsymbol r_0||). \tag{3.35}$$

By the consistency of $\widehat{\boldsymbol{r}}$ and the uniform consistency of $\widehat{\boldsymbol{\beta}}(\tau;\boldsymbol{r}_0)$, we can combind Taylor expansion and Equation (3.35) to see

$$\begin{aligned}
&\boldsymbol{w}\{\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{r}}),\boldsymbol{\alpha}_0,\widehat{\boldsymbol{r}}\}-\boldsymbol{w}\{\widehat{\boldsymbol{\beta}}(\boldsymbol{r}_0),\boldsymbol{\alpha}_0,\boldsymbol{r}_0\}\\
&=\boldsymbol{L}_r(\boldsymbol{\beta}_0,\boldsymbol{\alpha}_0,\boldsymbol{r}_0)(\widehat{\boldsymbol{r}}-\boldsymbol{r}_0)+\int_{\tau_a}^{\tau_b}\boldsymbol{L}_{\boldsymbol{b}}(\boldsymbol{\beta}_0,\boldsymbol{\alpha}_0,\boldsymbol{r}_0,\tau)\{\widehat{\boldsymbol{\beta}}(\tau;\widehat{\boldsymbol{r}})-\widehat{\boldsymbol{\beta}}(\tau;\boldsymbol{r}_0)\}\mathrm{d}\tau+o_p(n^{-1/2}\vee||\widehat{\boldsymbol{r}}-\boldsymbol{r}_0||)\\
&=\boldsymbol{D}_{\boldsymbol{r}}(\boldsymbol{\beta}_0,\boldsymbol{\alpha}_0,\boldsymbol{r}_0)\times(\widehat{\boldsymbol{r}}-\boldsymbol{r}_0)+o_p(n^{-1/2}\vee||\widehat{\boldsymbol{r}}-\boldsymbol{r}_0||), \qquad (3.36)
\end{aligned}$$

where $\boldsymbol{D}_{\boldsymbol{r}}(\boldsymbol{\beta}_0,\boldsymbol{\alpha}_0,\boldsymbol{r}_0)=\boldsymbol{L}_r(\boldsymbol{\beta}_0,\boldsymbol{\alpha}_0,\boldsymbol{r}_0)-\int_{\tau_a}^{\tau_b}\boldsymbol{L}_{\boldsymbol{b}}(\boldsymbol{\beta}_0,\boldsymbol{\alpha}_0,\boldsymbol{r}_0,\tau)\boldsymbol{B}_{\boldsymbol{b}}(\boldsymbol{\beta}_0,\boldsymbol{\alpha}_0,\boldsymbol{r}_0,\tau)^{-1}\times$ $\boldsymbol{B}_r(\boldsymbol{\beta}_0,\boldsymbol{\alpha}_0,\boldsymbol{r}_0,\tau)\mathrm{d}\tau$. In fact, it can be shown that $\boldsymbol{D}_{\boldsymbol{r}}(\boldsymbol{\beta}_0,\boldsymbol{\alpha}_0,\boldsymbol{r}_0)$ equals $\partial\boldsymbol{w}\{\widetilde{\boldsymbol{\beta}}(\boldsymbol{r}),\boldsymbol{\alpha}_0,\boldsymbol{r}\}/\partial\boldsymbol{r}\big|_{\boldsymbol{r}=\boldsymbol{r}_0}$ basing on Equation (3.17).

Again mimicking the arguments in Lemma 3.4.2, $n^{1/2}[\boldsymbol{W}_n\{\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{r}}),\widehat{\boldsymbol{\alpha}},\widehat{\boldsymbol{r}}\}-\boldsymbol{W}_n\{\widehat{\boldsymbol{\beta}}(\boldsymbol{r}_0),\widehat{\boldsymbol{\alpha}},\boldsymbol{r}_0\}]$ can be approximated by $n^{1/2}[\boldsymbol{w}\{\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{r}}),\widehat{\boldsymbol{\alpha}},\widehat{\boldsymbol{r}}\}-\boldsymbol{w}\{\widehat{\boldsymbol{\beta}}(\boldsymbol{r}_0),\widehat{\boldsymbol{\alpha}},\boldsymbol{r}_0\}\big]$, and furthermore by $n^{1/2}[\boldsymbol{w}\{\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{r}}),\boldsymbol{\alpha}_0,\widehat{\boldsymbol{r}}\}-\boldsymbol{w}\{\widehat{\boldsymbol{\beta}}(\boldsymbol{r}_0),\boldsymbol{\alpha}_0,\boldsymbol{r}_0\}]$. This fact, coupled with $\boldsymbol{W}_n\{\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{r}}),\widehat{\boldsymbol{\alpha}},\widehat{\boldsymbol{r}}\}=o(n^{-1/2})$ and Equation (3.36), gives

$$-n^{1/2}\boldsymbol{W}_n\{\widehat{\boldsymbol{\beta}}(\boldsymbol{r}_0),\widehat{\boldsymbol{\alpha}},\boldsymbol{r}_0\}=\boldsymbol{D}_{\boldsymbol{r}}\times n^{1/2}(\widehat{\boldsymbol{r}}-\boldsymbol{r}_0)+o_p(1\vee n^{1/2}||\widehat{\boldsymbol{r}}-\boldsymbol{r}_0||),$$

which further implies

$$n^{1/2}(\widehat{\boldsymbol{r}}-\boldsymbol{r}_0)=-n^{-1/2}\sum_{i=1}^{n}\boldsymbol{D}_{\boldsymbol{r}}^{-1}\boldsymbol{\iota}_i(\boldsymbol{\beta}_0,\boldsymbol{\alpha}_0,\boldsymbol{r}_0)+o_p(1). \qquad (3.37)$$

when combined with equation (3.33). It follows immediately that $n^{1/2}(\widehat{\boldsymbol{r}}-\boldsymbol{r}_0)$ converges weakly to a normal distribution with mean 0.

Finally, we combine Equation (3.27) and (3.35) to see

$$
\begin{aligned}
n^{1/2}\{\widehat{\boldsymbol{\beta}}(\tau)-\boldsymbol{\beta}_0(\tau)\} &= n^{1/2}\{\widehat{\boldsymbol{\beta}}(\tau;\widehat{\boldsymbol{r}})-\widehat{\boldsymbol{\beta}}(\tau;\boldsymbol{r}_0)\}+n^{1/2}\{\widehat{\boldsymbol{\beta}}(\tau;\boldsymbol{r}_0)-\boldsymbol{\beta}_0(\tau)\} \\
&= n^{-1/2}\sum_{i=1}^{n}\big[\boldsymbol{B}_b(\boldsymbol{\beta}_0,\boldsymbol{\alpha}_0,\boldsymbol{r}_0,\tau)^{-1}\boldsymbol{B}_r(\boldsymbol{\beta}_0,\boldsymbol{\alpha}_0,\boldsymbol{r}_0,\tau)\boldsymbol{D}_{\boldsymbol{r}}^{-1}\boldsymbol{\iota}_i(\boldsymbol{\beta}_0,\boldsymbol{\alpha}_0,\boldsymbol{r}_0) \\
&\qquad + \boldsymbol{B}_b(\boldsymbol{\beta}_0,\boldsymbol{\alpha}_0,\boldsymbol{r}_0,\tau)^{-1}\boldsymbol{\chi}_i(\boldsymbol{\beta}_0,\boldsymbol{\alpha}_0,\boldsymbol{r}_0,\tau)\big]+o_p(1),
\end{aligned}
$$

which implies that $n^{1/2}\{\widehat{\boldsymbol{\beta}}(\tau)-\boldsymbol{\beta}_0(\tau)\}$, $\tau\in[\nu_1,\tau_{U,1}]$, converges weakly to a Gaussian process with mean 0. This completes the proof of Theorem 3.1.2.

3.5 Convergence Criteria

Reliable convergence criteria are needed to determine a good termination point of the proposed iterative algorithm in Section 3.1.3. We first define convergence for step D.0-D.2, where a sequence of $\{\widehat{\boldsymbol{\beta}}^{[k,m]}(\tau): m=0,1,\dots\}$ is obtained by solving equation 3.9 iteratively to get $\widehat{\boldsymbol{\beta}}(\tau;\boldsymbol{r})$ for a fixed $\boldsymbol{r}=\boldsymbol{r}^{[k]}$. For brevity we temporarily omit k here and denote the sequence as $\{\widehat{\boldsymbol{\beta}}^{[m]}(\tau): m=1,2,\dots\}$. Define $\mathcal{D}\{\boldsymbol{\beta}_1,\boldsymbol{\beta}_2\}=\max_{i=0,1,\dots,p}\int_{\nu_1}^{\tau_{U,1}}||\boldsymbol{\beta}_2^{(i)}(\tau)-\boldsymbol{\beta}_1^{(i)}(\tau)||\mathrm{d}\tau$, which characterizes the distance between two functions $\boldsymbol{\beta}_1(\tau)$ and $\boldsymbol{\beta}_2(\tau)$ for $\tau\in[\nu_1,\tau_{U,1})$. We regard the sequence as converged at the qth iteration for a certain tolerance level tol_b if $\mathcal{D}\{\widehat{\boldsymbol{\beta}}^{[q]},\widehat{\boldsymbol{\beta}}^{[q-1]}\}\le tol_b$, and set $\widehat{\boldsymbol{\beta}}(\tau;\boldsymbol{r})$ as $\widehat{\boldsymbol{\beta}}^{[q]}(\tau)$. During the iterations, one may encounter slight oscillations at a fraction of the grid points, such that $\mathcal{D}\{\widehat{\boldsymbol{\beta}}^{[q]},\widehat{\boldsymbol{\beta}}^{[q-1]}\}>tol_b$ but $\mathcal{D}\{\widehat{\boldsymbol{\beta}}^{[q]},\widehat{\boldsymbol{\beta}}^{[q-2]}\}\le tol_b$. If that happens, we also regard the sequence as converged and set $\widehat{\boldsymbol{\beta}}(\tau;\boldsymbol{r})$ as $\{\widehat{\boldsymbol{\beta}}^{[q]}(\tau)+\widehat{\boldsymbol{\beta}}^{[q-1]}(\tau)\}/2$. If neither of $\mathcal{D}\{\widehat{\boldsymbol{\beta}}^{[q]},\widehat{\boldsymbol{\beta}}^{[q-1]}\}$ and $\mathcal{D}\{\widehat{\boldsymbol{\beta}}^{[q]},\widehat{\boldsymbol{\beta}}^{[q-2]}\}$ is less than tol_b, then convergence is not yet achieved at the qth iteration. We would proceed with the iterations until convergence, or a maximum number of iterations denoted by $max.iter_0$, at which we re-assess convergence with a relaxed tolerance level $tol_{b.relax}$. The iterations are concluded as non-convergent if convergence is not satisfied at $max.iter_0$ with $tol_{b.relax}$.

The convergence criteria for Step A-D need to concern both $\{\widehat{\boldsymbol{\beta}}^{[k]}(\tau) : k = 0, 2, \dots\}$ and $\{\widehat{\boldsymbol{r}}^{[k]}(\tau) : k = 1, 2, \dots\}$. Convergence of the former can be assessed in exactly the same way as for $\{\widehat{\boldsymbol{\beta}}^{[k,m]}(\tau) : m = 1, 2, \dots\}$, except that we may choose a different maximum number of iteration, denoted as $max.iter_1$. For convergence to be achieved at step q, we also need to require $||Ken(\boldsymbol{r}^{[q]}) - Ken(\boldsymbol{r}^{[q-1]})|| \leq tol_k$, where $Ken(\boldsymbol{r})$ maps the association parameter $\boldsymbol{r}$ to the corresponding Kendall's tau coefficient, and tol_k is a prespecified tolerence level. We specify the criteria on the Kendall's tau's scale instead of on $\boldsymbol{r}$ itself, such that we have a unified criteria for different copula functions.

In both Section 3.2.1 and 3.2.2, we choose $tol_b = 5e-4$, and $tol_{b.relax} = tol_k = 5e-3$. Also we set $max.iter_0 = 10$ and $max.iter_1 = 20$.

Chapter 4

Temporal Regression for Left-truncated Semi-competing Risks Data

4.1 Regression Procedures

4.1.1 Data and Model

The data structure considered here is identical to that in Chapter 1. Specifically, the semi-competing risks data comprises (X, Y, δ, η, Z), where $X = \min(T_1, T_2, C)$, $Y = \min(T_2, C)$, $\delta = I(T_1 \leq Y)$, $\eta = I(T_2 \leq C)$, and $Z = (1, Z_1, \ldots, Z_p)^T$ is the $(p + 1) \times 1$ covariate vector. Let L denote time to left truncation to the terminating event, data is observable only when $L < Y$. We denote the observed data as $(L_i^*, X_i^*, Y_i^*, \delta_i^*, \eta_i^*, Z_i^*)_{i=1}^n$, which are identically and independently distributed and follow the conditional distribution of $(L, X, Y, \delta, \eta, Z)$ given $L < Y$.

Here we consider regression modeling of the cumulative incidence via a class of semiparametric models that accomodate time-varying regression coefficients, as in Scheike et al. (2008). Given covariate vector Z, we assume that

$$F_1(t|Z) = g\{Z^T \beta_0(t)\}, \tag{4.1}$$

where $g(\cdot)$ is a prespecified monotone link function, and $\beta_0(t) = \{\beta_0^{(0)}(t), \beta_0^{(1)}(t), ..., \beta_0^{(p)}(t)\}^T$ is the vector of unknown regression coefficients. The first component of $\beta_0(t)$ corresponds to the g^{-1} transformed baseline cumulative incidence when $Z = (1, 0^T)^T$, and the remaining p components correspond to covariate effects on the cumulative incidence function at time t.

When $g(x) = 1 - \exp\{-\exp(x)\}$, model (4.1) can be viewed as an extension of the proportional subdistribution hazards model (Fine and Gray, 1999), which corresponds to the situation when $\beta^{(j)}(t)$, $j = 1, 2, ..., p$, are constant across t. Alternatively, model (4.1) becomes the proportional odds model when $g(x) = \exp(x)/\{1 + \exp(x)\}$. This allows us to investigate the odds of observing the nonterminating event before time t. Another useful link is $g(x) = 1 - \exp(-x)$, which leads to an additive risk model for T_1.

It is worth pointing out that the model (4.1) considered here differs from the cumulative incidence quantile regression model (2.1) in Chapter 2 in two major aspects. First, model (4.1) formulates the effects of covariates directly on the cumulative incidence function, as compared to model (2.1) that formulates the effects on cumulative incidence through its quantile functions. Second, the results from model (4.1) and (2.1) should be interpreted in different ways. In the Denmark diabetes registry data, for example, the regression coefficients in (2.1) reflect how covariates shifts the time to which a certain percentage of patients experience the DN. By constrast, the regression coefficients in (4.1) characterize the relationship between covariates and patients' DA status at each specific time points. The two models can both offer meaningful and yet complimentary interpretations on the mechanisms of DN progression.

4.1.2 Estimation

To estimate $\beta_0(t)$, we similarly utilize the following relationship between $I(X_i^* \leq t, \delta_i^* = 1, \eta_i^* = 1)$ and the cumulative incidence function. Letting $G(y,z) = P(L < y \leq C|Z = z)$ and $\alpha(z) = P(L < Y|Z = z)$, we have

$$E\left\{\frac{I(X_i^* \leq t, \delta_i^* = 1, \eta_i^* = 1)}{G(Y_i^*, Z_i^*)/\alpha(Z_i^*)}\middle| Z_i^*\right\} = F_1(t|Z_i^*). \tag{4.2}$$

In Chapter 2, we propose a consistent estimator of $W(y,z) \equiv G(y,z)/\alpha(z)$, denoted as $\hat{W}(y,z)$. Therefore, an IPCW-type estimating equation can be constructed as follows:

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} D_i\{\beta(t)\}V_i\{\beta(t)\}\big[\frac{I(X_i^* \leq t, \delta_i^* = 1, \eta_i^* = 1)}{\hat{W}(Y_i^*, Z_i^*)} - g\{Z_i^{*T}\beta(t)\}\big] = 0, \tag{4.3}$$

where $D_i\{\beta(t)\} = -\partial g\{Z_i^{*T}\beta(t)\}/\partial\beta(t)$, and $V_i\{\beta(t)\}$ is a scaler weight function. If we set $V_i\{\beta(t)\} = 1$, the root finding in (4.3) is equivalent to the minimization of the following objective function

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\big[\frac{I(X_i^* \le t, \delta_i^* = 1, \eta_i^* = 1)}{\hat{W}(Y_i^*, Z_i^*)} - g\{Z_i^{*T}\beta(t)\}\big]^2,$$

the minimizor of which can be found via standard statistical software, such as the *optim* function in R. Alternatively, we can also solve (4.3) through existing root-finding functionalities for generalized linear models, like the R function *glm*. The resulting $\hat{\beta}(t)$ is piecewise constant and jumps only at the observed failure times of T_1^*.

In Chapter 2, we showed that $\hat{W}(y,z) - W(y,z) = n^{-1}\sum_{i=1}^{n} w_i(y,z) + o_p(n^{-1/2})$, where $w_i(y,z)$ is of expectation zero and can be consistently estimated by $\hat{w}_i(y,z)$. The detailed form of $w_i(y,z)$ and $\hat{w}_i(y,z)$ can be found in Section 2.4.1. Combining this with empirical process arguments, we can show that $\hat{\beta}(t)$ is uniformly consistent for $t \in [l,u]$, where l and u are positive constants subject to certain regularity conditions. Moreover, it can be shown that $\sqrt{n}\{\hat{\beta}(t) - \beta_0(t)\}$ converges weakly to a mean-zero Gaussian process with covariance matrix $\Sigma(s,t) = E\{\xi(s)\xi(t)^T\}$. A consistent estimator of $\Sigma(s,t)$ may be given by $n^{-1}\sum_{i=1}^{n}\hat{\xi}_i(s)\hat{\xi}_i(t)^T$, where

$$\begin{aligned}
\hat{\xi}_i(t) &= -\big[n^{-1}\sum_{i=1}^{n} D_i\{\hat{\beta}(t)\}^{\otimes 2}V_i\{\hat{\beta}(t)\}\big]^{-1} \\
&\times \Bigg(D_i\{\hat{\beta}(t)\}V_i\{\hat{\beta}(t)\}[\frac{I(X_i^* \le t, \delta_i^* = 1, \eta_i^* = 1)}{\hat{W}(Y_i^*, Z_i^*)} - g\{Z_i^*\hat{\beta}(t)\}] \\
&- \frac{1}{n}\sum_{j=1}^{n}\frac{D_j\{\hat{\beta}(t)\}V_j\{\hat{\beta}(t)\}I(X_j^* \le t, \delta_j^* = 1, \eta_j^* = 1)}{\hat{W}(Y_j^*, Z_j^*)^2}\hat{w}_i(Y_j^*, Z_j^*)\Bigg).
\end{aligned}$$

4.1.3 Second Stage Inferences

Besides evaluating $\beta_0(t)$ on a range of t, it is also practically meaningful to provide summary measures of the temporal covariate effects. We define the trimmed mean effect $\eta^{(j)} = \{\int_l^u \beta_0^{(j)}(t)dt\}/(u-l)$, $j = 1, 2, ..., p$, which can be interpreted as the average effect of Z_j on the transformed cumulative incidence function when t ranges from u to l. A consistent estimator of $\eta^{(j)}$ is given by $\widehat{\eta}^{(j)} = \{\int_l^u \widehat{\beta}_0^{(j)}(t)dt\}/(u-l)$, where $n^{1/2}(\widehat{\eta}^{(j)} - \eta^{(j)})$ converges in distribution to normal distribution with mean 0. The standard error of $n^{-1/2}\{\hat{\eta}^{(j)} - \eta^{(j)}\}$ can be consistently estimated by $n^{-1}\sum_{i=1}^n \hat{\iota}_i^{(j)2}$, where $\hat{\iota}^{(j)} = \{\int_l^u \hat{\xi}_i^{(j)}(s)ds\}/(u-l)$.

We also conduct formal hypothesis testing to access whether covariate Z_j has a significant effect on the cumulative incidence. The null hypothesis is specified as $H_{01}^{(j)} : \beta_0^{(j)}(t) = 0$. A natural Wald-type test statistics $\mathcal{T}_{mean}^{(j)}$ can be constructed based on the trimmed mean effect estimate $\widehat{\eta}^{(j)}$. Alternatively, one can define a more robust test statistics $\mathcal{T}_{sup}$ as

$$\mathcal{T}_{sup}^{(j)} = n^{1/2} \sup_{t \in [l,u]} |\widehat{\beta}^{(j)}(t)\Phi(t)|,$$

where the weight function $\Phi(t)$ can be chosen as the inverse of the estimated standard error of $\widehat{\beta}^{(j)}(t)$. Based on the supreme-norm, $\mathcal{T}_{sup}^{(j)}$ is omnibus for all departures from $H_{01}^{(j)}$. The distribution of $\mathcal{T}_{sup}^{(j)}$ under the null hypothesis is rather complicated but can be approximated via resampling.

Furthermore, it is also important to access whether the effect of Z_j is constant for $t \in [l, u]$. To this end, we specify the null hypothesis as $H_{02}^{(j)} : \beta_0^{(j)}(t) = c_0$ and define $\mathcal{T}_{cons}^{(j)} = \int_l^u \Xi(t)\widehat{\beta}^{(j)}(t)dt - \widehat{\eta}^{(j)}$, where $\Xi(t)$ is a nonconstant function satisfying $\int_l^u \Xi(t) = 1$. Again wald-type hypothesis testing can be conducted, utilizing influence function based standard error estimates. In practice, the form of $\Xi(t)$ can be chosen based on scientific interests or prior knowledge of the covariate effects. For example, one may set $\Xi(t) = 2I\{t \leq (l+u)/2\}/(u-l)$ if there is interest in accessing monotone trend in $\beta_0^{(j)}(t)$.

4.2 Numerical Studies

4.2.1 Simulations

We conduct some preliminary simulations to evaluate the performance of the proposed estimator. Here we study the situation when $g(x) = \log\{-\log(1-x)\}$. The covariate vector $Z = (1, Z_1, Z_2)^T$, where $Z_1 \sim Uniform(0,1)$ and $Z_2 \sim Bernoulli(0.5)$. The cumulative incidence of T_1 satisfies

$$\log[-\log\{1 - F_1(t|Z)\}] = \log\big(- \log[\{1 - p + p\exp(-t)\}\big) + Z_1 b_1(t) + Z_2 b_2(t)$$

for $t \leq 3$, where $p = 0.5$, $b_1(t) = 0.75$, and $b_2(t) = 0.4t/(0.4t + 1)$ is a monotone function of t. We generated T_2 from a Cox proportional hazards model with regression coefficient $\gamma_0 = (0.4, -0.2)^T$ and set the baseline to be $Weibull(shape = 0.9, scale = 5)$. The left truncation time $L = r_L \times Uniform(0, 2.1)$, where $r_L \sim Bernoulli(0.8)$. This truncation scheme leads to 20% left truncation on T_2. We let the censoring time $C = L + D$, where $D \sim Uniform(6, 7.5)$ and is independent of L. Overall, the frequency of T_1 and T_2 being observed is 62% and 75% respectively.

We performed the proposed analyses on 1000 simulated datasets with sample size $n = 200$ and 400. The simulation results are summaried in Table 1 below. We reported the empirical bias (EmpBias), empirical standard error (EmpSD), and average of estimated standard errors (EstSD) of $\hat{\beta}(t)$ at $t = 0.8, 1.6, 2.4$, as well as the empirical coverage probabilities of the 95% Wald-type confidence intervals (COV95). It is observed that $\hat{\beta}(t)$ is virtually unbiased. The relatively larger bias at small t may have been because that $\beta^{(0)}(t)$ goes to $-\infty$ when $t \to 0$. We observe that the magnitude of bias shrinks with the sample size. The standard error estimates are quite close to their empirical counterparts. The empirical coverages of the confidence intervals based on normal approximation agree with the nominal level 95%.

Table 4.1: Simulation studies: empirical biases, empirical standard errors, average standard error estimates of $\hat{\beta}(t)$, and the empirical coverage probabilities of Wald-type confidence intervals.

t	EmpBias			EmpSD			EstSD			COV95		
	$\hat{\beta}^{(0)}$	$\hat{\beta}^{(1)}$	$\hat{\beta}^{(2)}$	$\hat{\beta}^{(0)}$	$\hat{\beta}^{(1)}$	$\hat{\beta}^{(2)}$	$\hat{\beta}^{(0)}$	$\hat{\beta}^{(1)}$	$\hat{\beta}^{(2)}$	$\hat{\beta}^{(0)}$	$\hat{\beta}^{(1)}$	$\hat{\beta}^{(2)}$
						n=200						
0.8	-55	27	9	424	610	346	417	610	349	944	944	949
1.6	-34	18	8	320	474	272	333	506	286	950	956	956
2.4	-28	14	8	311	479	272	324	510	283	955	961	955
						n=400						
0.8	-19	-1	6	280	411	234	291	424	244	966	966	957
1.6	-3	-17	5	219	322	187	234	353	200	964	966	963
2.4	0	-18	3	216	334	185	228	357	198	964	966	966

We also performed second stage inferences on $\beta_0^{(j)}(t)$, $j = 1, 2$. Table 4.2 reports the summary statistics of the trimmed mean effect estimates $\hat{\eta}^{(j)}(t)$. We see that the trimmed mean effects are accurately estimated, and the average estimated standard errors of $\hat{\eta}^{(j)}(t)$ matches the empirical values. In addition, we conducted the significance test for Z_1 and Z_2 via test statistics $\mathcal{T}_{mean}^{(j)}$, $j = 1, 2$, and the constancy test through $\mathcal{T}_{cons}^{(j)}$. We summarized the performances of the two tests via their empirical rejection rates (ERR), under columns $H_{01}^{(j)}$ and $H_{02}^{(j)}$ respectively (Table 4.2). Judging by these results, the significance test has reasonably good power in detecting the departure from the null hypothesis, and the power increases with the sample size. Furthermore, the constancy test is satisfactory in terms of size and power.

Therefore, we can see that the proposed estimators work well under this setup. In near future, we will investigate their performances with other truncation schemes and link functions.

Table 4.2: Simulation studies: empirical bias, empirical standard errors, average standard error estimates of the trimmed mean effect estimates $\hat{\eta}^{(j)}$, and the empirical rejection rates of the hypothesis tests.

n	j	$\eta^{(j)}$			$H_{01}^{(j)}$	$H_{02}^{(j)}$
		EmpBias	EmpSD	EstSD	ERR	ERR
200	1	0.021	0.474	0.502	0.323	0.049
	2	0.008	0.270	0.284	0.265	0.277
400	1	0.000	0.341	0.351	0.556	0.055
	2	0.006	0.188	0.200	0.522	0.461

4.2.2 Analysis of Denmark Diabetes Registry Study

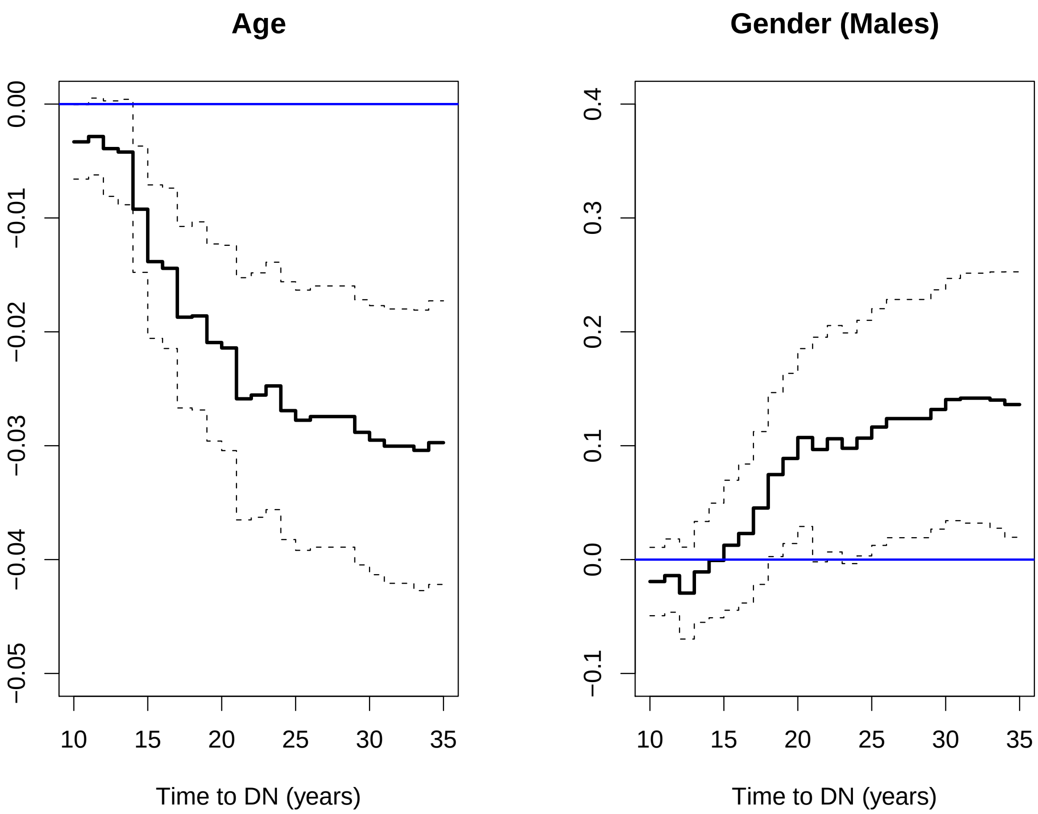
In Section 2.2.2, we performed quantile regression analysis for the cumulative incidence of diabetic nephropathy (DN) , among the cohort of subjects that were born before 1940 and had diabetic onset age greater or equal to 15 in the Denmark diabetes registry study. There were two covariates, Z_1 and Z_2, which are diabetic onset age and gender respectively. The analysis suggests that older diabetic onset age is associated with slower DN progression, and the association is more pronounced among subjects with relatively lower risks of DN. The analysis does not detect significant gender effect for $\tau \in [0.05, 0.22]$.

In this subsection, we perform an alternative analysis by fitting model (4.1) to the same cohort of subjects. Our target here is to evaluate the temporal relationship between covariates and subjects' DN status. We adopt the additive risk model for T_1 by setting $g(x) = 1 - \exp(-x)$, where negative regression coefficients correspond to protective covariate effects on DN.

Figure 4.1 shows the estimated regression coefficient $\hat{\beta}(t)$ for $t \in [10, 35]$ (years), as well as the corresponding 95% Wald-type confidence intervals. First, we observe that the confidence interval for age mostly excludes 0, suggesting some protective effect of older diabetic onset age on DN progression. Furthermore, the magnitude of the regression coefficient for age seems to increase with time, which suggests that patients with slower DN progression are more susceptible to the influence of diabetic onset age when compared to those who had fast progression to DN. These implications agree well with those offered by the quantile regression model. Concerning the gender effect on DN, we observe that the difference between males and females is not significant for relatively smaller t. However, it appears that males may experience faster DN progression than females at larger time points. Interestingly, our quantile regression analysis in Section 2.2.2 does not suggest a significant gender effect for $\tau \in [0.05, 0.22]$, where the upper bound of the τ range is selected based on some identifiability considerations. A possible reason is that the τ range under consideration in Section 2.2.2 mainly corresponds to the first half of the time range $t \in [10, 35]$.

In our second stage inferences, we obtain a trimmed mean effect estimate for age as -0.021 with an estimated standard error of 0.004. The corresponding p-value for testing $H_{01}^{(1)}$ is less than 0.001, suggesting highly significant effect of diabetic onset age on time to DN. In addition, the test statistics for $H_{02}^{(1)}$ equals 0.008 with an estimated standard error of 0.002. This gives a p-value < 0.001, which indicates that the age effect on DN progression may strengthen over time. We obtain a trimmed mean effect estimate for gender as 0.082. The corresponding standard error estimate equals 0.038, yielding a p-value of 0.02. This result suggests that there exists some difference between male and female subjects in terms of DN progression. We obtain $\mathcal{T}_{cons}^{(2)}$ as -0.048 with an standard error of 0.02. The corresponding p-value for testing $H_{02}^{(2)}$ equals 0.007, which confirms our observation from Figure 4.1 that the gender effect varies with time.

Figure 4.1: Denmark diabetes registry study: estimated regression coefficients based on the temporal regression model. (bold solid line: estimated regression coefficients; dashed line: 95% Wald-type pointwise confidence intervals).



Chapter 5

Summary and Future Work

5.1 Summary

In this dissertation we study two scenarios of semi-competing risks data that are commonly encountered in biomedical studies. We propose regression methods under the frameworks of quantile regression and temporal regression, two varying-coefficient regression models that are fastly emerging in survival analysis. The proposed methods are well fitted to the scenarios of interest for scientific implications.

We first study the mortality-morbidity scenario, with an additional complication by administrative left truncation. We develop inference procedures for the conditional quantiles of the cumulative incidence function, which facilitate straightforward scientific interpretation. The proposed inference procedures well-utilize the semi-competing risks structure and do not require artificial truncation. The proposed estimator can be acquired based on standard statistical software, and can be shown to be uniformly consistent and weakly converge to Gaussian. Simulation studies show the proposed method performs well with moderate sample size. We apply the proposed method to Denmark diabetes registry study to illustrate its practical utility.

Next, we focus on the endpoint-dropout scenario. We develop inference procedures for the conditional quantiles of the marginal distribution function, which well-characterize the underlying biological process, as well as the dependence structure between the primary endpoint and the dependent censoring. The proposed inference procedure makes good use of all available information in the semi-competing risks structure. We develop an efficient iterative algorithm, which can be readily and stably implemented. Using sophisticated arguments involving empirical process and stochastic integral equations, we established the asymptotic properties of the proposed estimator. Monte Carlo simulations demonstrate the satisfactory performance of the proposed estimator with moderate sample size. An application to the ACTG 364 study shows the applicational power of the proposed methods.

Furthermore, we study temporal regression under the mortality-morbidity scenario. We proposed estimation and inference procedures for the temporal covariate effects on the cumulative incidence by properly accounting for left-truncation. The proposed methods can be implemented with standard statistical software, and leads to straightforward interpretations of covariate effects on the cumulative incidence. Some preliminary simulation studies suggest satisfactory performances of the proposed estimator.

5.2 Future Work

In this subsection we discuss work to be done in the near future and possible extensions of this dissertation work.

First, we will conduct more simulation studies for our third project to get a more comprehensive evaluation of the numerical performances of the proposed estimator. Specifically, we will consider simulation setups with different truncation schemes and link functions. In addition, we will provide detailed justifications for the asymptotic properties of the proposed estimators.

It is worthwhile to consider the case of multiple event times, some of which are nonterminating while others are terminating. A general solution to such problems under the quantile regression and temporal regression frameworks would be especially useful in practice.

Furthermore, it is interesting to study varying-coefficient regression on the crude quantities of left-truncated competing risks data. Recently, some research efforts have been devoted to regression modeling with such data structure by assuming a proportional hazards model (Geskus, 2010; Shen, 2011; Zhang et al., 2011). It would be desirable to relax the constant covariate effect assumption for more robust and comprehensive analyses.

Bibliography

Albrecht, M., Bosch, R., Hammer, S., Liou, S., Kessler, H., Para, M., Eron, J., Valdez, H., Dehlinger, M., and Katzenstein, D. AIDS Clinical Trials Group 364 Study Team.(2001). Nelfinavir, efavirenz, or both after the failure of nucleoside treatment of HIV infection. *New England Journal of Medicine* **345,** 398–407.

Albrecht, M., Bosch, R., Hammer, S., Liou, S., Kessler, H., Para, M., Eron, J., Valdez, H., Dehlinger, M., Katzenstein, D., et al. (2001). Nelfinavir, Efavirenz, or Both after the Failure of Nucleoside Treatment of HIV Infection. *New England Journal of Medicine* **345,** 398–407.

Alexander, K. (1984). Probability Inequalities for Empirical Processes and a Law of the Iterated Logarithm. *The Annals of Probability* **12,** 1041–1067.

Andersen, P., Borgan, Ø., Gill, R., and Keiding, N. (1993). *Statistical Models based on Counting Processes*. Springer.

Andersen, P. and Gill, R. (1982). Cox's Regression Model for Counting Processes: a Large Sample Study. *The Annals of Statistics* **10,** 1100–1120.

Asgharian, M., M'Lan, C., and Wolfson, D. (2002). Length-Biased Sampling With Right Censoring. *Journal of the American Statistical Association* **97,** 201–209.

Bender, R., Augustin, T., and Blettner, M. (2005). Generating Survival Times to Simulate Cox Proportional Hazards Models. *Statistics in Medicine* **24,** 1713–1723.

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

Chao, M. (1987). Influence Curves for Randomly Truncated Data. *Biometrika* **74,** 426–429.

Clayton, D. (1978). A Model for Association in Bivariate Life Tables and its Application in Epidemiological Studies of Familial Tendency in Chronic Disease Incidence. *Biometrika* **65,** 141–151.

Efron, B. (1967). The Two Sample Problem with Censored Data. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability* **4,** 831–853.

Efron, B. (1979). Bootstrap Methods: Another Look at the Jackknife. *The Annals of Statistics* **7,** 1–26.

Emoto, S. and Matthews, P. (1990). A Weibull Model for Dependent Censoring. *The Annals of Statistics* **18,** 1556–1577.

Fine, J. and Gray, R. (1999). A Proportional Hazards Model for the Subdistribution of a Competing Risk. *Journal of the American Statistical Association* **94,** 496–497.

Fine, J., Jiang, H., and Chappell, R. (2001). On Semi-competing Risks Data. *Biometrika* **88,** 907–919.

Fine, J., Yan, J., and Kosorok, M. (2004). Temporal process regression. *Biometrika* **91,** 683.

Fygenson, M. and Ritov, Y. (1994). Monotone Estimating Equations for Censored Data. *The Annals of Statistics* **22,** 732–746.

Genest, C. (1987). Frank's Family of Bivariate Distributions. *Biometrika* **74,** 549–555.

Geskus, R. (2010). Cause-specific cumulative incidence estimation and the fine and gray model under both left truncation and right censoring. *Biometrics* .

Goodman, V., Kuelbs, J., and Zinn, J. (1981). Some Results on the LIL in Banach Space with Applications to Weighted Empirical Processes. *The Annals of Probability* pages 713–752.

Heckman, J. and Honoré, B. (1989). The Identifiability of the Competing Risks Model. *Biometrika* pages 325–330.

Hsieh, J., Wang, W., and Adam Ding, A. (2008). Regression Analysis based on Semi-competing Risks Data. *Journal of the Royal Statistical Society: Series B(Statistical Methodology)* **70,** 3–20.

Huang, Y. (2002). Calibration Regression of Censored Lifetime Medical Cost. *Journal of the American Statistical Association* **97,** 318–328.

Huang, Y. (2010). Quantile Calculus and Censored Regression. *The Annals of Statistics* **38,** 1607–1637.

Huang, Y. and Wang, M. (1995). Estimating the Occurrence Rate for Prevalent Survival Data in Competing Risks Models. *Journal of the American Statistical Association* **90,** 1406–1415.

Jiang, H., Fine, J., Kosorok, M., and Chappell, R. (2005). Pseudo Self-consistent Estimation of a Copula Model with Informative Censoring. *Scandinavian Journal of Statistics* **32,** 1–20.

Jin, Z., Ying, Z., and Wei, L. (2001). A Simple Resampling Method by Perturbing the Minimand. *Biometrika* **88,** 381–390.

Kendall, M. and Gibbons, J. (1962). *Rank Correlation Methods.* Charles Griffin, London.

Klein, J. and Andersen, P. (2005). Regression Modeling of Competing Risks Data based on Pseudovalues of the Cumulative Incidence Function. *Biometrics* **61,** 223–229.

Klein, J. and Moeschberger, M. (1988). Bounds on Net Survival Probabilities for Dependent Competing Risks. *Biometrics* **44,** 529–538.

Koenker, R. (2005). *Quantile Regression*. Cambridge Univ Pr.

Koenker, R. and Bassett, G. (1978). Regression Quantiles. *Econometrica* **46,** 33–50.

Kosorok, M. (2008). *Introduction to Empirical Processes and Semiparametric Inference*. Springer Verlag.

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

Lin, D., Robins, J., and Wei, L. (1996). Comparing two Failure Time Distributions in the presence of Dependent Censoring. *Biometrika* **83,** 381–393.

Link, W. (1989). A Model for Informative Censoring. *Journal of the American Statistical Association* **84,** 749–752.

Lynden-Bell, D. (1971). A method of allowing for known observational selection in small samples applied to 3CR quasars. *Monthly Notices of the Royal Astronomical Society* **155,** 95–118.

Martinussen, T. and Scheike, T. (2002). A flexible additive multiplicative hazard model. *Biometrika* **89,** 283.

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

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

Nelsen, R. (2006). *An Introduction to Copulas*. Springer Us.

Oakes, D. (1982). A Model for Association in Bivariate Survival Data. *Journal of the Royal Statistical Society. Series B (Methodological)* **44,** 414–422.

Parzen, M., Wei, L., and Ying, Z. (1994). A Resampling Method based on Pivotal Estimating Functions. *Biometrika* **81,** 341–350.

Peng, L. and Fine, J. (2006a). Nonparametric Estimation with Left-truncated Semicompeting Risks Data. *Biometrika* **93,** 367–383.

Peng, L. and Fine, J. (2006b). Rank Estimation of Accelerated Lifetime Models with Dependent Censoring. *Journal of the American Statistical Association* **101,** 1085–1093.

Peng, L. and Fine, J. (2007). Regression Modeling of Semicompeting Risks Data. *Biometrics* **63,** 96–108.

Peng, L. and Fine, J. (2009). Competing Risks Quantile Regression. *Journal of the American Statistical Association* **104,** 1440–1453.

Peng, L. and Huang, Y. (2007). Survival analysis with temporal covariate effects. *Biometrika* **94,** 719.

Peng, L. and Huang, Y. (2008). Survival Analysis with Quantile Regression Models. *Journal of the American Statistical Association* **103,** 637–649.

Peng, L., Jiang, H., Chappell, R., and Fine, J. (2008). An Overview of the Semi–Competing Risks Problem. In Biswas, A., Datta, S., Fine, J., and Segal, M., editors, *Statistical Advances in the Biomedical Sciences: Clinical trials, Epidemiology, Survival Analysis, and Bioinformatics*, pages 177–192. Wiley-Blackwell.

Peterson, A. (1976). Bounds for a Joint Distribution Function with Fixed Subdistribution Functions: Application to Competing Risks. *Proceedings of the National Academy of Sciences of the United States of America* **73,** 11–13.

Portnoy, S. (2003). Censored Regression Quantiles. *Journal of the American Statistical Association* **98,** 1001–1013.

Portnoy, S. and Lin, G. (2010). Asymptotics for Censored Regression Quantiles. *Journal of Nonparametric Statistics* **22,** 115–130.

Powell, J. (1984). Least Absolute Deviations Estimation for the Censored Regression Model. *Journal of Econometrics* **25,** 303–325.

Powell, J. (1986). Censored Regression Quantiles. *Journal of Econometrics* **32,** 143–155.

Prentice, R., Kalbfleisch, J., Peterson Jr, A., Flournoy, N., Farewell, V., and Breslow, N. (1978). The Analysis of Failure Times in the presence of Competing Risks. *Biometrics* **34,** 541–554.

Robins, J. and Rotnitzky, A. (1992). Recovery of Information and Adjustment for Dependent Censoring using Surrogate Markers. *Aids Epidemiology, Methodological issues* pages 297–331.

Scharfstein, D. and Robins, J. (2002). Estimation of the Failure Time Distribution in the presence of Informative Censoring. *Biometrika* **89,** 617–634.

Scheike, T., Zhang, M., and Gerds, T. (2008). Predicting cumulative incidence probability by direct binomial regression. *Biometrika* **95,** 205.

Shen, P. (2011). Proportional subdistribution hazards regression for left-truncated competing risks data. *Journal of Nonparametric Statistics* **1,** 11.

Slud, E. and Rubinstein, L. (1983). Dependent Competing Risks and Summary Survival Curves. *Biometrika* **70,** 643–649.

Tian, L., Zucker, D., and Wei, L. (2005). On the cox model with time-varying regression coefficients. *Journal of the American statistical Association* **100,** 172–183.

Tsiatis, A. (1975). A Nonidentifiability Aspect of the Problem of Competing Risks. *Proceedings of the National Academy of Sciences* **72,** 20–22.

Valsecchi, M., Silvestri, D., and Sasieni, P. (1996). Evaluation of long-term survival: Use of diagnostics and robust estimators with cox's proportional hazards model. *Statistics in Medicine* **15,** 2763–2780.

Van der Vaart, A. and Wellner, J. (1996). *Weak Convergence and Empirical Processes: with Applications to Statistics*. Springer Verlag.

Wang, M. (1991). Nonparametric estimation from cross-sectional survival data. *Journal of the American Statistical Association* **86,** 130–143.

Wang, W. (2003). Estimating the Association Parameter for Copula Models under Dependent Censoring. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* **65,** 257–273.

Winnett, A. and Sasieni, P. (2003). Iterated residuals and time-varying covariate effects in cox regression. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **65,** 473–488.

Woodroofe, M. (1985). Estimating a Distribution Function with Truncated Data. *The Annals of Statistics* **13,** 163–177.

Yang, S. (1999). Censored Median Regression using Weighted Empirical Survival and Hazard Functions. *Journal of the American Statistical Association* **94,** 137–146.

Ying, Z., Jung, S., and Wei, L. (1995). Survival Analysis with Median Regression Models. *Journal of the American Statistical Association* **90,** 178–184.

Zhang, X., Zhang, M., and Fine, J. (2011). A proportional hazards regression model for the subdistribution with right-censored and left-truncated competing risks data. *Statistics in Medicine* .

Zheng, M. and Klein, J. (1995). Estimates of Marginal Survival for Dependent Competing Risks based on an Assumed Copula. *Biometrika* **82,** 127–138.

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