Distribution Agreement

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

Signature:

Shuang Ji	Date

Quantile Regression for Complex Censored Data

By

Shuang Ji

Doctor of Philosophy

Biostatistics

Limin Peng, Ph.D.
Advisor

Yijian (Eugene) Huang, Ph.D.
Committee Member

Brent A. Johnson, Ph.D.
Committee Member

David H. Howard, Ph.D.
Committee Member

Accepted:

Lisa A. Tedesco, Ph.D.
Dean of the Graduate School

Date

Quantile Regression for Complex Censored Data

By

Shuang Ji

B.S., University of Science and Technology of China, 2005

M.S., University of Texas at Arlington, 2007

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

2012

Abstract

Survival data subject to complex censoring schemes are frequently encountered in biomedical research. For such data, naive application of classical approaches built for the random censoring case may lead to substantial estimation bias. In this dissertation, we focus on two different scenarios that involve complex censoring mechanisms, dependent censoring and double censoring. We develop appropriate methods under the quantile regression (Koenker and Bassett, 1978) framework, which are expected to accommodate a more dynamic relationship between covariates and survival time compared to traditional regression models in survival analysis.

The first part of this dissertation is motivated by the Warfarin-Aspirin Symptomatic Intracranial Disease (WASID) study, in which dependent censoring is posed by informative withdrawal. One scientific interest is about the analysis of time to a study endpoint defined as ischemic stroke, brain hemorrhage, or death from vascular causes, whichever happens first, corresponding to the setting where subjects do not withdraw. We propose a quantile regression procedure for such dependently censored data, along with an efficient and stable algorithm. We establish the uniform consistency and weak convergence of the resulting estimators. Extensive simulation studies demonstrate good finite-sample performance of the proposed inferential procedures. We illustrate the practical utility of our method via an application to the WASID study.

The second part of this dissertation is motivated by the US Cystic Fibrosis Foundation Patient Registry (CFFPR) study, in which double censoring presents while the left censoring variable is always observed. It is of interest to investigate the association between age at the first Pseudomonas aeruginosa (PA) infection, an important landmark event of CF pathology, and a set of risk factors. We propose a new analysis strategy for such doubly censored data and develop computationally simple estimation and inference procedures. Moreover, we propose conditional inference to address the special identifiability issues attached to the doubly censoring setting. Asymptotic properties are established for the resulting estimators, and the finite-sample performance is assessed by simulation studies. Analysis of the CFFPR study is also conducted based on our method.

In the third part, we study a double censoring data structure with unobservable left censoring times. We develop a self-consistent estimating equation along with an iterative algorithm. Our simulation studies demonstrate good finite-sample properties of the proposed method. We also apply the proposed method to the CFFPR study.

In summary, this dissertation work provides useful quantile regression tools for analyzing complex survival data, which have broad applications in medical and public health research.

Acknowledgments

I would like to sincerely and gratefully thank my advisor, Dr. Limin Peng, for her guidance, understanding, and patience. Not only has she provided tremendous help to my research by her insight and expertise, she sets a role model for me as a researcher and mentor.

I would also like to thank Dr. Eugene Huang, Dr. Brent Johnson and Dr. David Howard for their time and effort in serving as my committee members. They provided very helpful suggestions that has lead to substantial improvement in this dissertation.

I owe special thanks to Dr. Qi Long for providing guidance on my long-term collaboration projects with the Department of Psychiatry & Behavioral Sciences at Emory University, which constitute an important part of my PhD training. I am also grateful to Dr. Amita Manatunga, Dr. Brent Johnson, and Mr. Kirk Easley for various research and consulting opportunities.

I would like to extend my appreciation to all faculty and staff members in the Department of Biostatistics and Bioinformatics at Emory, and to all my friends, with whom I shared many happy moments over the past five years.

Finally, I would like to deeply thank my family for their unwavering support and unconditional love. Their never-fading faith in me musters my courage to overcome difficult times.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background

In survival analysis, one important task is to deal with various types of censoring and truncation. Many efforts have been devoted to handle random right censoring (i.e., event time and censoring time are independent given covariates if any). For example, classical approaches, such as the Kaplan-Meier estimator (Kaplan and Meier, 1958) and the Cox proportional hazards model (Cox, 1972), were originally built for this scenario. However, in practice, researchers frequently encounter more complex censoring scenarios, where naive applications of existing methods for dealing with random right censoring can lead to substantial biases.

Motivated by several large biomedical studies, we develop appropriate methods for analyzing survival data subject to complex censoring mechanisms. Specifically, in this dissertation, we study three research problems on two scenarios commonly encountered: dependent censoring and double censoring. In the first project, we investigate cases where the independent censoring assumption is not appropriate, for example, when there is informative dropout in a clinical trial. In the second and third projects, we concern doubly censored data, which arise when left censoring is present in addition to right censoring. As elaborated later, considerable statistical challenges are involved in developing valid statistical methods that can appropriately accommodate these situations.

Throughout this dissertation research, we focus on the quantile regression modeling (Koenker and Bassett, 1978), which has emerged as a valuable alternative to the popular Cox model and accelerated failure time (AFT) model. With quantile regression, one would be able to assess how covariates impact various quantiles of event time without imposing constant effects as contrast to the traditional Cox regression and AFT model. Such analyses may also help detect population inhomogeneous risk patterns, for example, covariate effects that vary between patients who have high susceptibility to the event of interest (e.g., onset of a certain disease, disease progression,

or death) versus those who are less prone to this event. In the presence of complex censoring schemes, such as the dependent censoring and double censoring scenarios considered here, little has been studied under the quantile regression framework. We aim to fill this gap in this dissertation.

In the rest of this chapter, we first describe two motivating examples, the Warfarin Aspirin Symptomatic Intracranial Disease (WASID) study and the Cystic Fibrosis Foundation Patient Registry (CFFPR) study, and present literature reviews on analysis of dependently censored data and analysis of doubly censored data separately. This is followed by a general review of quantile regression for survival data. An outline of this dissertation is given at the end of this chapter.

1.2 Motivating Examples

1.2.1 The Warfarin-Aspirin Symptomatic Intracranial Disease (WASID) Study

Dependent censoring is of concern in many clinical studies. A good example comes from the WASID study, the first clinical trial that compared warfarin and aspirin in treating atherosclerotic intracranial arterial stenosis, an important cause of stroke (Chimowitz et al., 2005). In this trial, 569 patients who had stroke or transient ischemic attack resulting from stenosis of a major intracranial artery were randomized to receive either warfarin or aspirin. The primary endpoint was ischemic stroke, brain hemorrhage, or death from vascular causes other than stroke, whichever happens first. The mean follow-up duration was 1.8 years, and there was administrative censoring, which is independent of the primary endpoint.

Of note, study medications were terminated early for 125 patients due to various disease-related reasons, such as adverse events and changes in health conditions. Subsequently, these patients were given treatments other than the original assignments,

and the follow-up continued. In the primary analysis (Chimowitz et al., 2005), an intent-to-treat (ITT) strategy was followed. That is, for patients whose assigned treatments were terminated early, no distinction was made between the follow-up information before and after the withdrawal. It is of interest to conduct a secondary on-treatment analysis that considers the withdrawals as censoring to the disease endpoint and thus confers the effect pertaining to the originally assigned treatment. Such an analysis would provide a useful complement to the primary analysis. For the new analysis considered here, one complication is that the withdrawals may be correlated with the study endpoint and thus pose a dependent censoring scenario. In addition, among those 125 withdrawals, 44 occurred in the aspirin arm and 81 in the warfarin arm. Such an unbalanced allocation can amplify the estimation bias for treatment effect caused by falsely treating withdrawals as independent censoring (Huang and Zhang, 2008). These considerations necessitate properly adjusting for dependent censoring. Inferences on event time corresponding to the removal of patient withdrawals are of direct scientific relevance and are studied in our first project.

1.2.2 The Cystic Fibrosis Foundation Patient Registry (CFFPR) Study

Double censoring often arises in registry data and other observational studies. An example is the US CFFPR study, which documents the diagnosis and annual follow-ups of all known CF patients. As one of the most common and life-shortening genetic disorders affecting the lungs and digestive systems, Cystic Fibrosis (CF) has a rather complex progression path, one landmark event of which is the onset of Pseudomonas aeruginosa (PA) infection. PA affects 80% of CF patients under age 18 by accelerating decline in lung function (Kosorok et al., 2001), and has been believed to be an important predictor of mortality in CF (Retsch-Bogart et al. 2008). Therefore, it is of scientific interest to investigate the association between onset ages of the first de-

tected PA infection and its risk factors, such as gender, diagnosis mode and diagnosis year.

To this end, one complication is that ages at the first PA infection were not observed for two types of patients: those who had been infected and thus were positive at the first documented CFFPR visit, and those who had no detected or reported PA infection by the end of follow-up. This poses a double censoring scenario. With the time origin set as birth, time to first PA infection was subject to left censoring by time to registry entry, which was always recorded, and also right censoring by time to last follow-up visit, which was not always known in advance due to the occurrence of random dropout. Among the 12,818 CF patients whose data were collected by the CFFPR study during the period between 1986 and 2005, 3,343 (26.1%) patients had PA infection at study entry (i.e., age at the first PA infection is left censored) and 2,213 (17.3%) patients had no PA infection documented by December 2005 (i.e., age at the first PA infection is right censored). Such a double censoring setting, with the left censoring time always observed, is the focus of the second part in my dissertation. In the third project, we study a similar data structure, but further relax the assumption by not requiring the left censoring time to be always known.

1.3 Literature Review

1.3.1 Existing Work on Dealing with Dependent Censoring

By viewing the occurrence of dependent censoring as a distinct endpoint, we may formulate the survival data subject to dependent censoring as competing risks data. As a result, the dependent censoring problem may be tackled by employing techniques for handling competing risks, which are generally classified into two categories (Kalbfleisch and Prentice, 2002): approaches based on *crude quantities*, and approaches that focus on *net quantities*.

Crude quantities, such as the cause-specific hazard and the cumulative incidence function, reflect the failure process in the presence of the competing risks and are desirable when the problem involves multiple types of failure, such as morbidity and mortality. Nonparametric estimation of such quantities has been well studied (Gray, 1988; Pepe, Pepe; Lin, 1997). As for regression methods based on *crude quantities*, one may naively fit a Cox model or an AFT model by treating dependent censoring as independent censoring, and the results actually correspond to the cause-specific hazard. A popular alternative is to fit a proportional hazards type model with respect to the cumulative incidence function (Fine and Gray, 1999).

Analyses based on the *net quantities*, such as the marginal distribution function, are sensible when it is of interest to make inference on the study endpoint while hypothesizing the removal of the competing risks. For example, when dependent censoring is caused by events that preclude the observation of but not the development of the endpoint of interest, such as the informative withdrawals in the WASID study, methods based on the *net quantities* may be preferred because it would produce inference which corresponds to the setting without the interruption of the observation process. By this consideration, in the first project of this dissertation, we focus on developing methods based on the *net quantities*, which may yield scientifically relevant and meaningful application for the WASID study.

There has been rich literature on competing risks approaches based on *net quantities*. As a common feature of this type of methods, additional assumptions on the relationship among times to distinct failure types are required because the marginal and joint distributions are not nonparametrically identifiable (Tsiatis, 1975). For example, in the one-sample case, much previous work with dependently censored data restricts the joint distribution using either semiparametric or parametric models (Link, 1989; Emoto and Matthews, 1990, among others). Due to lack of sufficient information to verify the assumed dependence structure, performing a sensitivity

analysis (Peterson, 1976; Slud and Rubinstein, 1983; Klein and Moeschberger, 1988; Zheng and Klein, 1995; Scharfstein et al., 2001; Scharfstein and Robins, 2002, among others) has been advocated to yield bounds for the estimands of interest under various plausible assumptions on the joint distribution of the event time and the censoring time.

The first project of this dissertation concerns the general regression setting. Among existing work based on the *net quantities*, Huang and Zhang (2008) extended Zheng and Klein (1995)'s approach to a bivariate Cox proportional hazards model, where the joint distributions of competing risks are linked to their marginal distributions through a known copula. More recently, Chen (2010) developed a non-parametric maximum likelihood approach for a general class of semiparametric transformation models, similarly assuming a copula model to address the identifiability issue. Both of these regression methods base inference on models that only allow for constant effects, which may not be adequate in many real data datasets (Kaslow et al., 1987; Dickson et al., 1989; Thorogood et al., 1990; Verweij and Van Houwelingen, 1995; Carey et al., 1995; Jensen et al., 1997, among others).

1.3.2 Existing Work on Dealing with Double Censoring

The simultaneous presence of left censoring and right censoring can bring many complexities to the analysis of doubly censored data. For example, the distribution estimator in the one-sample case is generally presented in the self-consistent manner (Turnbull, 1974; Tsai and Crowley, 1985; Chang and Yang, 1987; Samuleson, 1989; Chang, 1990; Gu and Zhang, 1993; Zhan and Wellner, 1995; Mykland and Ren, 1996, among others) and does not have a closed form. For the two-sample problem, Gehan (1965) and Mantel (1967) studied an extension of the Wilcoxon test. More recently, Ren (2008) proposed a weighted empirical likelihood-based semiparametric maximum likelihood estimator as a unified approach for the two-sample problem with various

censoring schemes including double censoring. Overall, these approaches are more complicated than their counterparts for randomly right censored data, for example, Kaplan-Meier estimator (Kaplan and Meier, 1958) or log-rank test (Mantel, 1966).

The second and third projects of this dissertation are concerned with the general regression setting. Among existing work, Zhang and Li (1996) proposed a Buckey-James-Ritvo-type M-estimator. Their estimating equation is neither monotone nor continuous and thus may necessitate special efforts to address some computational issues. More recently, Ren and Gu (1997) and Ren (2003) proposed a parallel regression M-estimator. This approach requires the independence between censoring variables and covariates, and thus imposes a stronger random censoring assumption than the usual one. For scenarios where both left and right censoring times are always observed, Cai and Cheng (2004) studied semiparametric transformation models (Cheng et al., 1995), and Yan et al. (2009) adapted temporal process regression (Fine et al., 2004) to doubly censored data. These two approaches are not suitable for the CFFPR example either because the right censoring time may not be known in advance due to random loss to follow-up. Without requiring censoring times to be known, Lin et al. (2012) developed a self-consistent estimator as an extension of Portnoy (2003) to the double censoring case, but did not provide asymptotic investigation.

1.4 Quantile Regression for Survival Data

In this dissertation, we propose new regression methods adjusting for dependent censoring and double censoring respectively. These methods are based on the quantile regression modeling, which was first introduced by Koenker and Bassett (1978). In contrast to the traditional linear regression which models the relationship between the mean of the response variable and the covariates, quantile regression seeks to model a spectrum of quantile functions of the response variable conditional on the

covariates instead of a single mean. It has received increasing attention in survival analysis. As elaborated in Portnoy (2003) and Peng and Huang (2008), the use of quantile regression in survival analysis offers straightforward interpretation on event times as well as extra model flexibility to accommodate varying covariate effects.

Let T denote the event time of interest, and $\tilde{\boldsymbol{Z}}$ be the $p \times 1$ vector of recorded covariates. Define $\boldsymbol{Z} = (1, \tilde{\boldsymbol{Z}})^T$. The conditional τ–th quantile of T given $\boldsymbol{Z}$ is defined as $Q_T(\tau|\boldsymbol{Z}) = \inf\{t : F_T(t|\boldsymbol{Z}) \geq \tau\}$, where $F_T(t|\boldsymbol{Z}) = \Pr(T \leq t|\boldsymbol{Z})$. A linear quantile regression model may take the form

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

where $g(\cdot)$ is a known monotone link function, and $\boldsymbol{\beta}_0(\tau)$ is a vector of unknown coefficients representing covariate effects on $Q_{g^{-1}(T)}(\tau|\boldsymbol{Z})$. As noted in Peng and Huang (2008), model (1.1) reduces to the accelerated failure time (AFT) model when $g(\cdot) = \exp(\cdot)$, and $\boldsymbol{\beta}(\tau|\boldsymbol{Z}) = (Q_\epsilon(\tau), \boldsymbol{b}^T)^T$, where $\boldsymbol{b}$ is an unknown vector of parameters and ϵ is an independently and identically distributed (i.i.d.) error term.

Substantial work has been done on applying quantile regression to survival data with independent right censoring. Among the earliest breakthroughs, Powell (1984, 1986) extended the least absolute deviation (LAD) from traditional quantile regression to censored quantile regression, assuming the censoring variables are fixed or always observable. Later efforts have been made to accommodate non-fixed censoring which is not always known by requiring additional restrictions such as unconditional independent censoring (Ying et al., 1995; Honore et al., 2002, among others), or nearly i.i.d. errors (Yang, 1999). More recently, without imposing these constraints, Portnoy (2003) proposed a recursively reweighted estimator as a generalization of the Kaplan-Meier estimator, with subsequent work by Neocleous et al. (2006) and Portnoy and Lin (2010) devoted to further polishing the algorithm and the asymptotic

theory. By utilizing the martingale structure of randomly right censored data Peng and Huang (2008) proposed an alternative approach, the resulting estimator of which reduces to the Nelson-Aelan estimator in the one-sample case. Their approach is well justified in theory and also has a convenient implementation. Wang and Wang (2009) employed a local reweighting scheme to relax the assumption of global linearity at all quantiles. Huang (2010) developed fundamental quantile calculus as the base and proposed a grid-free estimation procedure that is asymptotically equivalent to the procedure of Peng and Huang (2008). More recently, Peng (2011) proposed a self-consistent estimation based on stochastic integral equations and established the asymptotic equivalence between the proposed estimator and that of Peng and Huang (2008). This work also reveals the close connection between the proposed estimator and Portnoy (2003)'s estimator in their asymptotic behaviors.

To the best of our knowledge, there has been little work on developing quantile regression methods tailored to complex censoring schemes, such as the double censoring and dependent censoring settings considered in this proposal. For the double censoring case, Lin et al. (2012) proposed an iterative algorithm for estimating regression quantiles in a self-consistent manner, but did not develop asymptotic properties for the resulting estimators. For competing risks data, Peng and Fine (2009) proposed a quantile regression method based on the cumulative incidence function, which is not applicable to draw inference on *net quantities* as desired in the WASID study.

1.5 Outline

In Chapter 2 we develop a new quantile regression method for survival data subject to dependent censoring. We propose unbiased estimation equations for obtaining regression quantiles of the event time of and censoring time simultaneously, with the dependence structure formulated via a copula. We develop an efficient iterative algo-

rithm to solve the proposed estimating equations. Asymptotic properties, including the uniform consistency and weak convergence, are established for the resulting estimators. We report results from simulation studies which show satisfactory empirical performance of the proposed method, and illustrate the practical utility of our method by an analysis of the WASID study.

In Chapter 3 we propose a new quantile regression method suitable for doubly censored data with know left censoring time. We develop a computationally simple estimation and inference procedure by appropriately using the embedded martingale structure, and establish asymptotic properties for the resulting estimators. Moreover, we propose conditional inference to address the special identifiability issues attached to the doubly censoring setting. We further show that the proposed method can be readily adapted to handle left truncation. Results from simulation studies are shown to demonstrate good finite-sample performance. Finally we apply our method to the CFFPR data.

In Chapter 4 we investigate quantile regression for another double censoring scenario, in which neither left nor right censoring time is always observed. We propose an estimation procedure in a self-consistent manner, in which we utilize stochastic integrals to facilitate computation and theoretical developments. We report some preliminary simulation studies, which suggest proper finite sample performance of the proposed method, and analysis results for the CFFPR study.

In Chapter 5 we provide a summary of our completed work and propose plans for future research.

Chapter 2

Quantile Regression for Dependently Censored Data

2.1 Quantile Regression Procedures

2.1.1 Data and Model

Let T denote the failure time, D denote time to dependent censoring, and C be an additional independent censoring time. Let $\tilde{\boldsymbol{Z}}$ be a $p \times 1$ covariate vector. Define $\tilde{T} = T \wedge D$, $X = \tilde{T} \wedge C$, and $\boldsymbol{Z} = (1, \tilde{\boldsymbol{Z}}^T)^T$. Let $\tilde{\delta} = I(\tilde{T} \leq C)$. The censoring indicator is defined as $\delta = \tilde{\delta}$ if $T \leq D$, and $\delta = 2\tilde{\delta}$ if $D < T$. The observed data consist of n replicates of $(X, \delta, \boldsymbol{Z})$, denoted by $\{(X_i, \delta_i, \boldsymbol{Z}_i),\ i = 1, \cdots, n\}$.

Define the conditional τ–th quantile of a random variable Y given $\boldsymbol{Z}$ by $Q_Y(\tau|\boldsymbol{Z}) = \inf\{t : F_Y(t|\boldsymbol{Z}) \geq \tau\}$, where $F_Y(t|\boldsymbol{Z}) = \Pr(Y \leq t|\boldsymbol{Z})$. We consider the quantile regression model for T that takes the form

$$Q_T(\tau|\boldsymbol{Z}) = g\{\boldsymbol{Z}^T\boldsymbol{\beta}_0(\tau)\}, \quad \tau \in (0, 1), \tag{2.1}$$

where $g(\cdot)$ is a known monotone link function, and the unknown vector $\boldsymbol{\beta_0}(\tau)$ represents the covariate effects on $Q_T(\tau|\boldsymbol{Z})$. For simplicity, we adopt a similar model for D:

$$Q_D(\tau|\boldsymbol{Z}) = g\{\boldsymbol{Z}^T\boldsymbol{\alpha_0}(\tau)\}, \quad \tau \in (0, 1). \tag{2.2}$$

It is important to note that, due to the dependence between T and D given the covariates, models concerning the marginal probabilities, such as (2.1) and (2.2), cannot be identified without additional assumptions on the dependence structure between T and D (Tsiatis, 1975). To address this identifiability issue, we specify the dependence structure by a copula model which relates the joint survival function of (T, D) to the marginal distributions as follows:

$$\Pr(T > t_1, D > t_2|\boldsymbol{Z}) = H\{\Pr(T > t_1|\boldsymbol{Z}), \Pr(D > t_2|\boldsymbol{Z})\}, \tag{2.3}$$

where $H(\cdot, \cdot)$ is a known copula function.

2.1.2 Copula Functions

As can be seen in (2.3), a two-dimensional copula function maps the square region $[0, 1] \times [0, 1]$ to $[0, 1]$. It was established in Sklar's representation theorem (Sklar, 1959) that, for any random variables W_1 and W_2, there exists a copula function H such that

$$\Pr(W_1 > w_1, W_2 > w_2) = H\{\Pr(W_1 > w_1), \Pr(W_2 > w_2)\}.$$

If W_1 and W_2 are continuous variables, then H is uniquely determined by

$$H(u_1, u_2) = \Pr\{W_1 > Q_{W_1}(1 - u_1), W_2 > Q_{W_2}(1 - u_2)\}.$$

A trivial example of the copula functions is the independence copula, which takes the form $H(u, v) = uv$. Among commonly adopted copulas, the Clayton copula (Clayton, 1978) is given by $H(u, v) = \{u^{-r} + v^{-r} - 1\}^{-\frac{1}{r}}, r > 0$, and the Frank copula (Genest, 1987) takes the form $H(u, v) = \log_r\{1 + \frac{(r^u-1)(r^v-1)}{r-1}\}, r > 0$ and $r \neq 1$. Here r is a known copula parameter, which often contains information on the association level. For example, under the Clayton copula, the Kendall's tau equals $r/(r + 2)$. Under the Frank copula, the Kendall's tau equals $\{1 + 4(D(\nu) - 1)\}/\nu$, where $\nu = -\log r$ and $D(\nu) = [\int_0^\nu t/\{\exp(t) - 1\}\, \mathrm{d}t]/\nu$. In practice, r may be chosen according to prior knowledge on the strength of the association between T and D. Alternatively, one may obtain bounds of $\boldsymbol{\beta}_0(\tau)$ and hence $Q_T(\tau|\boldsymbol{Z})$ by perturbing r in a plausible range.

2.1.3 Estimation Equations

To estimate $\boldsymbol{\beta}_0(\tau)$ in model (2.1), we utilize the martingales associated with cause-specific hazard functions. Denote the counting process for T by $N_1(t) = I(X \leq t, \delta = 1)$. Define $M_1(t) = N_1(t) - \int_0^t Y(u)\lambda_1^*(u|\boldsymbol{Z})\,\mathrm{d}u$ where $Y(u) = I(X \geq u)$ and $\lambda_1^*(t|\boldsymbol{Z})$ is the cause-specific hazard for T. As shown by Kalbfleisch and Prentice (2002), $M_1(t)$ is a martingale with respect to the filtration $\mathscr{F}_t = \{N_1(u), Y(u+), \boldsymbol{Z}\}$. This implies

$$E\{N_1(t) - \int_0^t Y(s)\lambda_1^*(s|\boldsymbol{Z})\,\mathrm{d}s\} = 0, \forall t \geq 0. \tag{2.4}$$

Using the fact that $\lambda_1^*(t|\boldsymbol{Z}) = -\partial \log[H\{\bar{F}_T(t_1|\boldsymbol{Z}), \bar{F}_D(t_2|\boldsymbol{Z}); r\}]/\partial t_1|_{t_1=t_2=t}$ (Kalbfleisch and Prentice, 2002) and by variable transformation inside the integral, we can show that

$$\int_0^t Y(s)\lambda_1^*(s|\boldsymbol{Z})\,\mathrm{d}s = \int_0^{F_T\{t|\boldsymbol{Z}\}} Y\{Q_T(u|\boldsymbol{Z})\}\phi_1(1-u, \bar{F}_D\{Q_T(u|\boldsymbol{Z})|\boldsymbol{Z}\})\,\mathrm{d}u, \tag{2.5}$$

where $\phi_1(v_1, v_2) = \partial \log\{H(v_1, v_2)\}/\partial v_1$ and $\bar{F}_W(t)$ denotes the survival function for a random variable W. Furthermore, we note that under models (2.1) and (2.2),

$$F_D(t|\boldsymbol{Z}) = \int_0^1 I\{v \leq F_D(t|\boldsymbol{Z})\}\,\mathrm{d}v = \int_0^1 I[g\{\boldsymbol{Z}_i^T\boldsymbol{\alpha}_0(v)\} \leq t]\,\mathrm{d}v$$

and therefore

$$\bar{F}_D\{Q_T(u|\boldsymbol{Z})|\boldsymbol{Z}\} = 1 - \int_0^1 I\{\boldsymbol{Z}_i^T\boldsymbol{\alpha}_0(v) \leq \boldsymbol{Z}_i^T\boldsymbol{\beta}_0(u)\}\,\mathrm{d}v. \tag{2.6}$$

From (2.1),(2.4), (2.5) and (2.6) we then have

$$E\Big[\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{Z}_i\Big\{N_{1i}[g\{\boldsymbol{Z}_i^T\boldsymbol{\beta}_0(\tau)\}]-\int_0^{\tau}Y[g\{\boldsymbol{Z}_i^T\boldsymbol{\beta}_0(u)\}] \times \phi_1\Big(1-u,1-\int_0^1 I\{\boldsymbol{Z}_i^T\boldsymbol{\alpha}_0(v)\le \boldsymbol{Z}_i^T\boldsymbol{\beta}_0(u)\}\,\mathrm{d}v\Big)\,\mathrm{d}u\Big\}\Big]=0, \tag{2.7}$$

where $N_{1i}(t)$ is the sample analog of $N_1(t)$.

By treating T as the dependent censoring to D, a parallel equality to (2.7) can be derived for $\boldsymbol{\alpha}_0(\cdot)$. Define $N_2(t)=I(X\le t,\delta=2)$ and $\{N_{2i}(t)\}_{i=1}^n$ be the sample analogs of $N_2(t)$. Define $\phi_2(v_1,v_2)=\partial\log\{H(v_1,v_2)\}/\partial v_2$. We can show that

$$E\Big[\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{Z}_i\Big\{N_{2i}[g\{\boldsymbol{Z}^T\boldsymbol{\alpha}_0(\tau)\}]-\int_0^{\tau}Y[g\{\boldsymbol{Z}^T\boldsymbol{\alpha}_0(u)\}] \times \phi_2\Big(1-\int_0^1 I\{\boldsymbol{Z}_i^T\boldsymbol{\beta}_0(v)\le \boldsymbol{Z}_i^T\boldsymbol{\alpha}_0(u)\}\,\mathrm{d}v,1-u\Big)\,\mathrm{d}u\Big\}\Big]=0. \tag{2.8}$$

Motivated by (2.7) and (2.8), we propose to estimate $\boldsymbol{\beta}_0(\tau)$ and $\boldsymbol{\alpha}_0(\tau)$ from the following estimating equations:

$$n^{\frac{1}{2}}\boldsymbol{S}_n^{(k)}(\boldsymbol{\beta},\boldsymbol{\alpha},\tau)=0,\quad k=1,2, \tag{2.9}$$

where

$$\boldsymbol{S}_n^{(1)}(\boldsymbol{\beta},\boldsymbol{\alpha},\tau)=n^{-1}\sum_{i=1}^{n}\boldsymbol{Z}_i\Big\{N_{1i}[g\{\boldsymbol{Z}_i^T\boldsymbol{\beta}(\tau)\}]-\int_0^{\tau}Y_i[g\{\boldsymbol{Z}_i^T\boldsymbol{\beta}(u)\}] \times \phi_1\Big(1-u,1-\int_0^1 I\{\boldsymbol{Z}_i^T\boldsymbol{\alpha}(v)\le \boldsymbol{Z}_i^T\boldsymbol{\beta}(u)\}\,\mathrm{d}v\Big)\,\mathrm{d}u\Big\},$$

$$\boldsymbol{S}_n^{(2)}(\boldsymbol{\beta},\boldsymbol{\alpha},\tau)=n^{-1}\sum_{i=1}^{n}\boldsymbol{Z}_i\Big\{N_{2i}[g\{\boldsymbol{Z}_i^T\boldsymbol{\alpha}(\tau)\}]-\int_0^{\tau}Y_i[g\{\boldsymbol{Z}_i^T\boldsymbol{\alpha}(u)\}] \times \phi_2\Big(1-\int_0^1 I\{\boldsymbol{Z}_i^T\boldsymbol{\beta}(v)\le \boldsymbol{Z}_i^T\boldsymbol{\alpha}(u)\}\,\mathrm{d}v,1-u\Big)\,\mathrm{d}u\Big\}.$$

Note that the estimating equation (2.9) requires $\boldsymbol{\beta}_0(\tau)$ and $\boldsymbol{\alpha}_0(\tau)$ be identifiable for all $\tau \in (0,1)$, which may not be possible due to the censoring to T or D. To circumvent this difficulty, we modify (2.9) by truncating the time scale by an upper bound, $t_{max} = g\{\boldsymbol{Z}^T\boldsymbol{\beta}_0(\tau_{U,1})\} \wedge g\{\boldsymbol{Z}^T\boldsymbol{\alpha}_0(\tau_{U,2})\}$, where $\tau_{U,1}$, $\tau_{U,2} \in (0,1)$. This leads to a new estimating equation,

$$n^{\frac{1}{2}}\boldsymbol{S}_n^{*(k)}(\boldsymbol{\beta},\boldsymbol{\alpha},\tau) = 0, \quad k = 1,2, \tag{2.10}$$

where

$$\begin{aligned}
\boldsymbol{S}_n^{*(1)}(\boldsymbol{\beta},\boldsymbol{\alpha},\tau) = n^{-1}\sum_{i=1}^{n}\boldsymbol{Z}_i\Big\{ & N_{1i}[g\{\boldsymbol{Z}_i^T\boldsymbol{\beta}(\tau)\}]I\{g^{-1}(X_i) \le \boldsymbol{Z}_i^T\boldsymbol{\alpha}(\tau_{U,2})\} \\
& - \int_0^{\tau} Y_i[g\{\boldsymbol{Z}_i^T\boldsymbol{\beta}(u)\}]I\{\boldsymbol{Z}_i^T\boldsymbol{\beta}(u) \le \boldsymbol{Z}_i^T\boldsymbol{\alpha}(\tau_{U,2})\} \\
& \times \phi_1\Big(1-u, 1-\int_0^{\tau_{U,2}} I\{\boldsymbol{Z}_i^T\boldsymbol{\alpha}(v) \le \boldsymbol{Z}_i^T\boldsymbol{\beta}(u)\}\,\mathrm{d}v\Big)\,\mathrm{d}u\Big\}, \\
\boldsymbol{S}_n^{*(2)}(\boldsymbol{\beta},\boldsymbol{\alpha},\tau) = n^{-1}\sum_{i=1}^{n}\boldsymbol{Z}_i\Big\{ & N_{2i}[g\{\boldsymbol{Z}_i^T\boldsymbol{\alpha}(\tau)\}]I\{g^{-1}(X_i) \le \boldsymbol{Z}_i^T\boldsymbol{\beta}(\tau_{U,1})\} \\
& - \int_0^{\tau} Y_i[g\{\boldsymbol{Z}_i^T\boldsymbol{\alpha}(u)\}]I\{\boldsymbol{Z}_i^T\boldsymbol{\alpha}(u) \le \boldsymbol{Z}_i^T\boldsymbol{\beta}(\tau_{U,1})\} \\
& \times \phi_2\Big(1-\int_0^{\tau_{U,1}} I\{\boldsymbol{Z}_i^T\boldsymbol{\beta}(v) \le \boldsymbol{Z}_i^T\boldsymbol{\alpha}(u)\}\,\mathrm{d}v, 1-u\Big)\,\mathrm{d}u\Big\}.
\end{aligned}$$

Equation (2.10) now only involves the estimation of $\{\boldsymbol{\beta}_0(\tau),\ \tau \in (0,\tau_{U,1})\}$ and $\{\boldsymbol{\alpha}_0(\tau),\ \tau \in (0,\tau_{U,2})\}$, and thus does not demand the identifiability of $\boldsymbol{\beta}_0(\tau)$ and $\boldsymbol{\alpha}_0(\tau)$ in the upper tail of τ, pointing to a more realistic scenario. The rigorous theoretical conditions for $\tau_{U,1}$ and $\tau_{U,2}$ are deferred to the statement of asymptotic results. In practice, $\tau_{U,1}$ and $\tau_{U,2}$ may need to be selected adaptively. Some empirical rules for selecting $\tau_{U,1}$ and $\tau_{U,2}$ are presented in the next subsection.

2.1.4 Computing Algorithms

We develop an iterative algorithm for finding the solution to equation (2.10), namely Algorithm A. The procedure is descried below:

Step A0. Set $m = 0$. Choose the initial value $\hat{\boldsymbol{\alpha}}^{[m]}(\tau), \tau \in (0, \tau_{U,2}]$.

Step A1. Solve $\boldsymbol{S}_n^{*(1)}(\boldsymbol{\beta}, \hat{\boldsymbol{\alpha}}^{[m]}, \tau) = 0$ for $\hat{\boldsymbol{\beta}}^{[m+1]}(\tau), \tau \in (0, \tau_{U,1}^{[m+1]}]$. Update $\tau_{U,1}$ with $\tau_{U,1}^{[m+1]}$.

Step A2. Solve $\boldsymbol{S}_n^{*(2)}(\hat{\boldsymbol{\beta}}^{[m+1]}, \boldsymbol{\alpha}, \tau) = 0$ for $\hat{\boldsymbol{\alpha}}^{[m+1]}(\tau), \tau \in (0, \tau_{U,2}^{[m+1]}]$. Update $\tau_{U,2}$ with $\tau_{U,2}^{[m+1]}$.

Step A3. Let $m = m + 1$. Repeat Steps A1 and A2 until certain convergence criteria are met.

At Step A0, one practical way to set the initial estimates is to fit model (2.1) for T and fit model (2.2) for D using existing quantile regression techniques which assume T and D are independent, for example, using Peng and Huang (2008)'s method.

At Step A1, we adopt a grid-based procedure that assumes $\hat{\boldsymbol{\beta}}^{[m+1]}(\tau)$ to be a right-continuous step function jumping only on a prespecified grid, $\mathscr{G}_{L_n} = \{0 = \tau_0 < \tau_1 < \cdots < \tau_{L_n} = \tau_{U,1}^{[m+1]} < 1\}$. The solution can be obtained by sequentially solving the following monotone estimating equation in $\boldsymbol{\beta}(\tau_j)(j = 1, \cdots, L_n)$:

$$
\begin{aligned}
n^{-\frac{1}{2}} \sum_{i=1}^{n} \boldsymbol{Z}_i \Big\{ & I[X_i \le g\{\boldsymbol{Z}_i^T \boldsymbol{\beta}(\tau_j)\}, \delta_i = 1] I\{g^{-1}(X_i) \le \boldsymbol{Z}_i^T \hat{\boldsymbol{\alpha}}^{[m]}(\tau_{U,2})\} \\
& - \sum_{l=0}^{j-1} (\tau_{l+1} - \tau_l) I[X_i \ge g\{\boldsymbol{Z}_i^T \boldsymbol{\beta}(\tau_l)\}] I\{\boldsymbol{Z}_i^T \boldsymbol{\beta}(\tau_l) \le \boldsymbol{Z}_i^T \hat{\boldsymbol{\alpha}}^{[m]}(\tau_{U,2})\} \\
& \times \phi_1 \Big(1 - \tau_l, 1 - \int_0^{\tau_{U,2}} I\{\boldsymbol{Z}_i^T \hat{\boldsymbol{\alpha}}^{[m]}(v) \le \boldsymbol{Z}_i^T \boldsymbol{\beta}(\tau_l)\} \, \mathrm{d}v \Big) \Big\} = 0 \quad (2.11)
\end{aligned}
$$

with $g\{\boldsymbol{Z}_i^T \boldsymbol{\beta}(0)\}$ set to be 0.

Due to the monotonicity of (2.11), the root finding problem in (2.11) is equivalent

to locating the minimizer of the following L_1-type convex function:

$$
\begin{aligned}
l_j(\boldsymbol{h}) = & \sum_{i=1}^{n} |I(\delta_i = 1) I\{g^{-1}(X_i) \le \boldsymbol{Z}_i^T \hat{\boldsymbol{\alpha}}^{[m]}(\tau_{U,2})\} g^{-1}(X_i) \\
& \qquad - I(\delta_i = 1) I\{g^{-1}(X_i) \le \boldsymbol{Z}_i^T \hat{\boldsymbol{\alpha}}^{[m]}(\tau_{U,2})\} \boldsymbol{h}^T \boldsymbol{Z}_i| \\
& + |R^* - \boldsymbol{h}^T \sum_{l=1}^{n} \{-I(\delta_l = 1)\} I\{g^{-1}(X_l) \le \boldsymbol{Z}_l^T \hat{\boldsymbol{\alpha}}^{[m]}(\tau_{U,2})\} \boldsymbol{Z}_l| \\
& + \Big| R^* - \boldsymbol{h}^T \sum_{r=1}^{n} \Big(2\boldsymbol{Z}_r \sum_{s=0}^{j-1} I[g^{-1}(X_r) \ge \boldsymbol{Z}_r^T \boldsymbol{\beta}(\tau_s)] I\{\boldsymbol{Z}_r^T \boldsymbol{\beta}(\tau_s) \le \boldsymbol{Z}_r^T \hat{\boldsymbol{\alpha}}^{[m]}(\tau_{U,2})\} \\
& \qquad \times \phi_1\Big(1 - \tau_s, 1 - \int_0^{\tau_{U,2}} I\{\boldsymbol{Z}_r^T \hat{\boldsymbol{\alpha}}^{[m]}(v) \le \boldsymbol{Z}_r^T \boldsymbol{\beta}(\tau_s)\} \, \mathrm{d}v \Big) \times (\tau_{s+1} - \tau_s) \Big) \Big|,
\end{aligned}
\tag{2.12}
$$

where R^* is a very large number.

Note that $\tau_{U,1}$ is adaptively selected and may vary at each iteration. At the end of the m–th iteration, we choose $\tau_{U,1}$ to be $\tau_{U,1}^{[m+1]}$, the largest quantile at which $\hat{\boldsymbol{\beta}}^{[m+1]}(\cdot)$ can be solved. For example, we may examine the distance between $\hat{\boldsymbol{\beta}}^{[m+1]}(\tau_j)$ and $\hat{\boldsymbol{\beta}}^{[m+1]}(\tau_{j+1})$ for each j, namely d_j, and stop the sequential procedure if d_j exceeds a moderate pre-specified threshold and let $J = j$. The choice of the threshold can be quite flexible, but should avoid values that are too small (e.g., less than 1) or too large (e.g., greater than 100). In our numerical studies we set the threshold to 10. We set $\tau_{U,1}^{[m+1]} = \tau_J$, for which the underlying rationale is that, given a fine grid $\mathscr{G}_{L_n}$, $\hat{\boldsymbol{\beta}}^{[m+1]}(\tau_j)$ and $\hat{\boldsymbol{\beta}}^{[m+1]}(\tau_{j+1})$ are expected be very close in the identifiable τ–region for $\hat{\boldsymbol{\beta}}(\cdot)$ when $j > 0$.

Similarly as in Step A1, the root-finding procedure at Step A2 can be transformed to minimizing a L_1-type convex function parallel to (2.12) and we omit the exact expressions here. The L_1–minimization problem can be readily solved by using existing packages implemented in standard statistical software, such as $l1fit()$ function in S-PLUS and $rq()$ function in R. A similar adaptive strategy as for selecting $\tau_{U,1}$ can be adopted for $\tau_{U,2}$, which is updated at the m–th iteration with $\tau_{U,2}^{[m+1]}$, the largest

quantile at which $\hat{\boldsymbol{\alpha}}^{[m+1]}(\cdot)$ can be identified.

Based on our experience, we have found the numerical performance of the above algorithm sometimes unstable if there is heavy censoring on D. For example, in the context of the WASID study, about 80% of the observations on D were censored by either T or C. This is quite expected in a well-designed study when D represents informative dropouts. In such a case, a more restrictive version of model (2.2) for D may improve the efficiency and thus help increase the numerical stability. Following this rationale, we propose to adopt an AFT model for D, which only allows the intercept $\alpha_0^{(0)}(\tau)$ to vary with τ but imposes constancy on each covariate effect $\alpha_0^{(k)}(\tau)$ for $k = 1, \cdots, p$.

With an AFT model assumed for D, the algorithm is modified as follows:

Step B0. Set $m = 0$. Obtain the initial values $\hat{\boldsymbol{\alpha}}^{[m]}(\tau), \tau \in (0, \tau_{U,2}]$ by fitting an AFT model using Jin et al. (2003)'s method.

Step B1. Solve $\boldsymbol{S}_n^{*(1)}(\boldsymbol{\beta}, \hat{\boldsymbol{\alpha}}^{[m]}, \tau) = 0$ for $\hat{\boldsymbol{\beta}}^{[m+1]}(\tau), \tau \in (0, \tau_{U,1}^{[m+1]}]$. Update $\tau_{U,1}$ with $\tau_{U,1}^{[m+1]}$.

Step B2. Obtain $\hat{\boldsymbol{\alpha}}^{[m+1]}(\tau), \tau \in (0, \tau_{U,2}^{[m+1]}]$ via the following procedure:

(a) Solve $\boldsymbol{S}_n^{*(2)}(\hat{\boldsymbol{\beta}}^{[m+1]}, \boldsymbol{\alpha}, \tau) = 0$ for $\tilde{\boldsymbol{\alpha}}^{[m+1]}(\tau), \tau \in (0, \tilde{\tau}_{U,2}]$.

(b) Obtain the constant $\hat{\alpha}^{[m+1](k)}$ by taking the average of $\tilde{\alpha}^{[m+1](k)}(\tau)$ over $\tau \in [\tau_a, \tau_b]$ for $k = 1, \cdots, p$, where $\tau_a \in (0, \tilde{\tau}_{U,2})$ and $\tau_b \in (\tau_a, \tilde{\tau}_{U,2})$ are prespecified constants that represent a well identified region for $\tilde{\boldsymbol{\alpha}}^{[m+1]}(\tau)$.

(c) Compute the residual on the g^{-1} scale, i.e., $g^{-1}(\epsilon_i) = g^{-1}(X_i) - Q_i$, where $Q_i = \tilde{\boldsymbol{Z}}_i^T(\hat{\alpha}^{[m+1](1)}, \cdots, \hat{\alpha}^{[m+1](p)})$.

(d) Obtain $\hat{\alpha}^{(0)[m+1]}(\tau)$ for $\tau \in (0, \tau_{U,2}^{[m+1]}]$ by solving

$$\boldsymbol{S}_n^{**(2)}(\hat{\boldsymbol{\beta}}^{[m+1]}, \alpha^{(0)}, \tau) = 0,$$

where

$$\boldsymbol{S}_n^{**(2)}(\hat{\boldsymbol{\beta}}^{[m+1]}, \alpha^{(0)}, \tau) = n^{-1}\sum_{i=1}^{n}\boldsymbol{Z}_i\Big\{I[g^{-1}(\epsilon_i) \le \alpha^{(0)}(\tau), \delta_i = 2]I\{g^{-1}(X_i) \le \boldsymbol{Z}_i^T\hat{\boldsymbol{\beta}}^{[m+1]}(\tau_{U,1})\}$$
$$- \int_0^{\tau} I[g^{-1}(\epsilon_i) \ge \alpha^{(0)}(\tau)]I\{\alpha^{(0)}(u) \le \boldsymbol{Z}_i^T\hat{\boldsymbol{\beta}}^{[m+1]}(\tau_{U,1}) - Q_i\}$$
$$\times \phi_2\Big(1 - \int_0^{\tau_{U,1}} I\{\boldsymbol{Z}_i^T\hat{\boldsymbol{\beta}}^{[m+1]}(v) - Q_i \le \alpha^{(0)}(u)\}\,\mathrm{d}v, 1-u\Big)\,\mathrm{d}u\Big\}.$$

Update $\tau_{U,2}$ with $\tau_{U,2}^{[m+1]}$.

Step B3. Let $m = m + 1$. Repeat Steps B1 and B2 until certain convergence criteria are met.

Note that in this above procedure, $\tau_{U,2}$ is selected in a slightly different manner in contrast with Algorithm A. Specifically, at the m–th iteration, we set $\tau_{U,2}$ at $\tau_{U,2}^{[m+1]}$, the largest τ at which the intercept $\hat{\alpha}^{(0)}(\tau)$ can be obtained. We still select $\tau_{U,1}$ based on the identifiability of the $p+1$ vector $\hat{\boldsymbol{\beta}}(\tau)$. As in Steps A1 and A2, equations involved in Steps B1 and B2 can also be treated as L_1 minimization problems and thus conveniently solved. Details of the convergence criteria for Steps A3 and B3 are provided in Section 2.6.

2.1.5 Asymptotic Results

Under regularity conditions C1-C5 (provided in Section 2.5.1), we establish the the uniform consistency and weak convergence for $\hat{\boldsymbol{\beta}}(\tau)$ and $\hat{\boldsymbol{\alpha}}(\tau)$ stated in the following theorems.

Theorem 2.1.1. *Assuming conditions C1-C5 hold and* $\lim_{n\to\infty}\|\mathscr{G}_{L_n}\| = 0$*, then* $\sup_{\tau\in[\nu_1,\tau_{U,1}]}\|\hat{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}_0(\tau)\| \xrightarrow{p} 0$ *and* $\sup_{\tau\in[\nu_2,\tau_{U,2}]}\|\hat{\boldsymbol{\alpha}}(\tau) - \boldsymbol{\alpha}_0(\tau)\| \xrightarrow{p} 0$*, where* $0 < \nu_1 < \tau_{U,1}$ *and* $0 < \nu_2 < \tau_{U,2}$.

Theorem 2.1.2. *Assuming conditions C1-C5 hold and* $\lim_{n\to\infty} n^{1/2}\|\mathscr{G}_{L_n}\| = 0$*, then* $n^{1/2}\{\hat{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}_0(\tau)\}$ *converges weakly to a Gaussian process for* $\tau \in [\nu_1, \tau_{U,1}]$ *with*

$0 < \nu_1 < \tau_{U,1}$, *and* $n^{1/2}\{\hat{\boldsymbol{\alpha}}(\tau) - \boldsymbol{\alpha}_0(\tau)\}$ *converges weakly to a Gaussian process for* $\tau \in [\nu_2, \tau_{U,2}]$ *with* $0 < \nu_2 < \tau_{U,2}$.

The proofs of these two theorems can be viewed as extensions of those in Peng and Huang (2008) to the bivariate case, which however are not straightforward. To prove Theorem 2.1.1, we first note that the proposed estimating functions $\boldsymbol{S}_n^{(1)}(\boldsymbol{\beta}, \boldsymbol{\alpha}, \tau)$ and $\boldsymbol{S}_n^{(2)}(\boldsymbol{\beta}, \boldsymbol{\alpha}, \tau)$ converge to their expectations $\boldsymbol{s}^{(1)}(\boldsymbol{\beta}, \boldsymbol{\alpha}, \tau)$ and $\boldsymbol{s}^{(2)}(\boldsymbol{\beta}, \boldsymbol{\alpha}, \tau)$ uniformly in τ. Second, with fixed $\boldsymbol{\alpha}$ in equations $\boldsymbol{S}_n^{(1)}(\boldsymbol{\beta}, \boldsymbol{\alpha}, \tau) = 0$ and $\boldsymbol{s}^{(1)}(\boldsymbol{\beta}, \boldsymbol{\alpha}, \tau) = 0$, the solutions for $\boldsymbol{\beta}$ can be viewed as functionals of $\boldsymbol{\alpha}$, namely $\hat{\boldsymbol{\beta}}(\boldsymbol{\alpha}, \tau)$ and $\tilde{\boldsymbol{\beta}}(\boldsymbol{\alpha}, \tau)$, respectively. We can then use $\tilde{\boldsymbol{\beta}}(\hat{\boldsymbol{\alpha}}, \tau)$ to bridge $\hat{\boldsymbol{\beta}}(\hat{\boldsymbol{\alpha}}, \tau)$ and $\boldsymbol{\beta}_0(\tau) = \tilde{\boldsymbol{\beta}}(\boldsymbol{\alpha}_0, \tau)$. Similarly we can use $\tilde{\boldsymbol{\alpha}}(\hat{\boldsymbol{\beta}}, \tau)$ to bridge $\hat{\boldsymbol{\alpha}}(\hat{\boldsymbol{\beta}}, \tau)$ and $\boldsymbol{\alpha}_0(\tau) = \tilde{\boldsymbol{\alpha}}(\boldsymbol{\beta}_0, \tau)$, where $\hat{\boldsymbol{\alpha}}(\boldsymbol{\beta}, \tau)$ and $\tilde{\boldsymbol{\alpha}}(\boldsymbol{\beta}, \tau)$ are the solutions for $\boldsymbol{\alpha}$ to $\boldsymbol{S}_n^{(2)}(\boldsymbol{\beta}, \boldsymbol{\alpha}, \tau) = 0$ and $\boldsymbol{s}^{(2)}(\boldsymbol{\beta}, \boldsymbol{\alpha}, \tau) = 0$ with fixed $\boldsymbol{\beta}$, respectively. To circumvent the difficulty that $\|\hat{\boldsymbol{\beta}}(0)\| = \infty$ and $\|\hat{\boldsymbol{\alpha}}(0)\| = \infty$, which is implied by models (2.1) & (2.2) and our estimating procedure, we consider

$$\boldsymbol{\theta}(\tau) = \boldsymbol{\mu} \begin{pmatrix} \boldsymbol{\beta}(\tau) \\ \boldsymbol{\alpha}(\tau) \end{pmatrix} = \begin{pmatrix} E\Big(\boldsymbol{Z} N_1[g\{\boldsymbol{Z}^T\boldsymbol{\beta}(\tau)\}]\Big) \\ E\Big(\boldsymbol{Z} N_2[g\{\boldsymbol{Z}^T\boldsymbol{\alpha}(\tau)\}]\Big) \end{pmatrix},$$

and prove that $\hat{\boldsymbol{\theta}}(\tau)$ converges in probability to $\boldsymbol{\theta}_0(\tau)$ uniformly for $\tau \in (0, \tau_U]$. This result further leads to the uniform convergency of $\hat{\boldsymbol{\beta}}(\tau)$ and $\hat{\boldsymbol{\alpha}}(\tau)$ for $\tau \in [\nu, \tau_U]$, where $0 < \nu < \tau_U$.

To prove Theorem 2.1.2, we first establish the connection between $n^{1/2}(\{\hat{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}_0(\tau)\}^T, \{\hat{\boldsymbol{\alpha}}(\tau) - \boldsymbol{\alpha}_0(\tau)\}^T)^T$ and $n^{1/2}\{-\boldsymbol{S}_n^{(1)}(\boldsymbol{\beta}_0, \boldsymbol{\alpha}_0, \tau)^T, -\boldsymbol{S}_n^{(2)}(\boldsymbol{\beta}_0, \boldsymbol{\alpha}_0, \tau)^T\}^T$ via a stochastic integral equation. This result allows us to express $n^{1/2}(\{\hat{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}_0(\tau)\}^T, \{\hat{\boldsymbol{\alpha}}(\tau) - \boldsymbol{\alpha}_0(\tau)\}^T)^T$ as a linear map of $n^{1/2}\{-\boldsymbol{S}_n^{(1)}(\boldsymbol{\beta}_0, \boldsymbol{\alpha}_0, \tau)^T, -\boldsymbol{S}_n^{(2)}(\boldsymbol{\beta}_0, \boldsymbol{\alpha}_0, \tau)^T\}^T$. The latter can be shown to have weak convergence, which implies the result in Theorem 2.1.2. The detailed proofs of Theorems 2.1.1 and 2.1.2 are provided in Sections 2.5.2 and 2.5.3.

2.1.6 Inferences

Given the complex limiting distributions of $\hat{\boldsymbol{\beta}}(\tau)$ and $\hat{\boldsymbol{\alpha}}(\tau)$ as shown in the proof of Theorem 2.1.2, we employ the bootstrap approach (Efron, 1979) to make inference on $\boldsymbol{\beta}_0(\tau)$ and $\boldsymbol{\alpha}_0(\tau)$. For each of the B datasets obtained from bootstrapping, we conduct the estimation procedure presented earlier in this section and obtain $\{\boldsymbol{\beta}_b^*(\tau), \tau \in (0, \tau_{U,1,b}^*]\}_{b=1}^B$ and $\{\boldsymbol{\alpha}_b^*(\tau), \tau \in (0, \tau_{U,2,b}^*]\}_{b=1}^B$. For each fixed τ, we may estimate the variances of $\hat{\boldsymbol{\beta}}(\tau)$ and $\hat{\boldsymbol{\alpha}}(\tau)$ by the sample variances of $\{\boldsymbol{\beta}_b^*(\tau)\}_{b=1}^B$ and $\{\boldsymbol{\alpha}_b^*(\tau)\}_{b=1}^B$ respectively, and then construct confidence intervals of $\boldsymbol{\beta}_0(\tau)$ and $\boldsymbol{\alpha}_0(\tau)$ using normal approximation.

Hypotheses testing can be conducted to further investigate the patterns of the covariate effects. Define $\beta_0^{(q)}$ to be the coefficient corresponding to $\tilde{\boldsymbol{Z}}^{(q)}$ for $q = 1, \cdots, p$. In practice, one may be especially interested in testing (I) the overall significance of $\beta_0^{(q)}(\tau)$ across a pre-specified range of τ, say $[l, u]$, where $0 < l < u < \tau_{U,1}$, and (II) the constancy of $\beta_0^{(q)}(\tau)$ over $\tau \in [l, u]$. We may formulate these tests as (I) $H_0 : \beta_0^{(q)}(\tau) = 0, \tau \in [l, u]$, and (II) $\tilde{H}_0 : \beta_0^{(q)}(\tau) = \rho_0, \tau \in [l, u]$, where ρ_0 is an unknown constant. To perform these tests, we first define a useful summary statistic, $\eta_{0,q} \equiv \int_l^u \beta_0^{(q)}(v)dv/(u-l)$ for $q = 1, \cdots, p$, which may be interpreted as the average covariate effect of $\tilde{\boldsymbol{Z}}^{(q)}$ across $\tau \in [l, u]$. It can be shown that $\hat{\eta}_q = \int_l^u \hat{\beta}^{(q)}(v)dv/(u-l)$ is a consistent estimator for $\eta_{0,q}$ and is asymptotically normal. Give the observed data, the limiting distribution of $\hat{\eta}_q$ can be approximated by the sample $\{\eta_{b,q}^*\}_{b=1}^B$ where $\eta_{b,q}^* = \int_l^u \beta_b^{*(q)}(v)dv/(u-l)$. Testing H_0 is equivalent to testing $\eta_{0,q} = 0$, which is straightforward given the asymptotically normal distribution of $\hat{\eta}_q$. To test $\tilde{H}_0$, one may adopt the test statistic $\tilde{\Gamma} = n^{1/2} \int_l^u \{\hat{\beta}^{(q)}(v) - \hat{\eta}_q\}\Theta(v)\, \mathrm{d}v/(u-l)$, where $\Theta(\cdot)$ is a pre-specified weight function, which may be properly chosen to emphasize the departure from $\tilde{H}_0$. The limiting distribution of $\tilde{\Gamma}$ can be approximated by the sample $\{\tilde{\Gamma}_b^*\}_{b=1}^B$, where $\tilde{\Gamma}_b^* = n^{1/2} \int_l^u \{\beta_b^{*(q)}(v) - \eta_{b,q}^*\}\Theta(v)\, \mathrm{d}v/(u-l)$, and it naturally leads to Wald-type hypothesis testing.

2.2 Simulation Studies

We studied the finite-sample performance of the proposed estimators via Monte-Carlo simulations. For the association structure between T and D, we considered the Clayton copula with association parameter r_c and the Frank's copula with association parameter r_f. We set $r_c = \exp(1)$ and $r_f = \exp(-7.325)$ and correspondingly the values of Kendall's tau are around 0.58 for both settings, representing moderate dependency.

We generated T from a log linear model with heteroscedastic errors:

$$\log T = b_1 Z_1 + b_2 Z_2 + \epsilon_1,$$

where $Z_1 \sim Unif(0,1)$, $Z_2 \sim Bernoulli(0.5)$, and the error term ϵ_1 follows $N(0, 0.2^2)$ if $Z_2 = 0$ and $N(0, 0.4^2)$ if $Z_2 = 1$. In addition, D was generated from the AFT model

$$\log D = a_1 Z_1 + a_2 Z_2 + \epsilon_2,$$

where $\epsilon_2 \sim N(\mu_2, 0.3^2)$. The independent censoring time C was assumed to follow $Unif(0, c_u)$. Under this set-up, models (2.1) and (2.2) hold with $g(\cdot) = \exp(\cdot)$. It can be shown that the underlying regression quantile $\boldsymbol{\beta}_0(\tau) = \{\beta_0^{(0)}(\tau), \beta_0^{(1)}(\tau), \beta_0^{(2)}(\tau)\}^T$, where $\beta_0^{(0)}(\tau) = Q_{N(0,0.2^2)}(\tau)$, $\beta_0^{(1)} = b_1$, and $\beta_0^{(2)} = b_2 + Q_{N(0,0.4^2)}(\tau) - Q_{N(0,0.2^2)}(\tau)$. It can also be seen that $\boldsymbol{\alpha}_0(\tau) = \{Q_{\epsilon_2}(\tau), a_1, a_2\}^T$. Under each copula, we considered two specific configurations: (I) $\mu_2 = 0, b_1 = 0.27, b_2 = 0, a_1 = 0, a_2 = 0.3, c_u = 12$, which results in 45% dependent censoring rate and 10% independent censoring rate, and (II) $\mu_2 = 0.1, b_1 = 0.27, b_2 = 0, a_1 = 0, a_2 = 0.3, c_u = 12$, which results in 30% dependent censoring rate and 45% independent censoring rate. For case (I) we assumed a general quantile regression model for D. For case (II) we adopted the modified algorithm assuming AFT model for D with $\tau_a = 0.1$ and $\tau_b = 0.4$.

Under each configuration we simulated 1000 date sets of sample size $n = 200$. An equally spaced grid on τ with size 0.01 was adopted when estimating $\boldsymbol{\beta}_0(\tau)$ and $\boldsymbol{\alpha}_0(\tau)$. We chose $B = 100$ as the number of bootstrap replicates for the variance estimation.

Table 2.1 presents the estimation results when the Clayton copula was correctly adopted. We report absolute values of biases (Bias), empirical standard deviations (EmpSD), average estimated resampling-based standard deviations (AvgSD) of $\hat{\boldsymbol{\beta}}(\tau)$ and $\hat{\boldsymbol{\alpha}}(\tau)$, and coverage rates of 95% Wald confidence intervals of $\boldsymbol{\beta}_0(\tau)$ and $\boldsymbol{\alpha}_0(\tau)$ with $\tau = 0.1, 0.3, 0.5$ and 0.7. These results show that under these set-ups the biases are small, the bootstrap standard errors agree well with the empirical ones, and the coverage rates are in general close to the nominal level. For case (I), the convergence rate was 94.5% and on average 3.7 iterations were required to achieve convergence. For case (II) the convergence rate was 99.2%, achieved by an average of 4.4 iterations. In a similar fashion, Table 2.2 presents the estimation results when the Frank copula was correctly adopted. These results are also satisfactory. For case (I), the convergence rate was 92.7% with 5.5 iterations on average. For case (II) the convergence rate was 99.7%, achieved by an average of 5.3 iterations.

We also compared our approach with a naive application of Peng and Huang (2008) by simply treating D as independent censoring. Figure 2.1 displays the mean estimated coefficients from the proposed approach and those from the naive approach along with the true coefficients under a correctly specified Clayton copula, assuming an AFT model for D in both approaches. We can see that the proposed estimator $\hat{\boldsymbol{\beta}}(\tau)$ is virtually unbiased and $\hat{\boldsymbol{\alpha}}(\tau)$ has only small bias, while the naive approach can produce substantial bias. This again suggests the importance to properly account for dependent censoring.

To assess the robustness of our methods, we also carried out estimation procedures with mis-specified copulas and compared the results to those under the correct copu-

Table 2.1: Simulation Results on Parameter Estimation under the Clayton copula. Bias: biases; AvgSD: average estimated resampling-based standard deviations; EmpSD: empirical standard deviations; Cov95: coverage rates of 95% Wald confidence intervals.

τ		Bias	EmpSD	AvgSD	Cov95		Bias	EmpSD	AvgSD	Cov95
45% dep. censoring, 10% indep. censoring, model (2) for D										
0.1	$\hat{\beta}^{(0)}$	0.02	0.10	0.11	0.95	$\hat{\alpha}^{(0)}$	0.01	0.10	0.11	0.94
	$\hat{\beta}^{(1)}$	0.02	0.19	0.22	0.95	$\hat{\alpha}^{(1)}$	-0.02	0.17	0.19	0.94
	$\hat{\beta}^{(2)}$	-0.02	0.10	0.11	0.96	$\hat{\alpha}^{(2)}$	0.02	0.10	0.11	0.95
0.3	$\hat{\beta}^{(0)}$	0.01	0.09	0.10	0.95	$\hat{\alpha}^{(0)}$	0.02	0.08	0.10	0.94
	$\hat{\beta}^{(1)}$	0.01	0.16	0.19	0.97	$\hat{\alpha}^{(1)}$	-0.02	0.14	0.16	0.94
	$\hat{\beta}^{(2)}$	-0.01	0.08	0.09	0.96	$\hat{\alpha}^{(2)}$	0.01	0.08	0.10	0.96
0.5	$\hat{\beta}^{(0)}$	0.01	0.08	0.09	0.95	$\hat{\alpha}^{(0)}$	0.02	0.09	0.13	0.97
	$\hat{\beta}^{(1)}$	0.02	0.16	0.21	0.97	$\hat{\alpha}^{(1)}$	-0.02	0.14	0.19	0.97
	$\hat{\beta}^{(2)}$	-0.01	0.08	0.09	0.96	$\hat{\alpha}^{(2)}$	0.01	0.08	0.10	0.97
30% dep. censoring, 10% indep. censoring, AFT model for D										
0.1	$\hat{\beta}^{(0)}$	0.01	0.09	0.10	0.94	$\hat{\alpha}^{(0)}$	0.02	0.08	0.10	0.96
	$\hat{\beta}^{(1)}$	0.01	0.16	0.18	0.95	$\hat{\alpha}^{(1)}$	-0.03	0.13	0.15	0.95
	$\hat{\beta}^{(2)}$	-0.01	0.09	0.10	0.96	$\hat{\alpha}^{(2)}$	0.03	0.09	0.10	0.96
0.3	$\hat{\beta}^{(0)}$	0.01	0.07	0.08	0.95	$\hat{\alpha}^{(0)}$	0.02	0.08	0.10	0.96
	$\hat{\beta}^{(1)}$	0.00	0.13	0.15	0.96	$\hat{\alpha}^{(1)}$	-0.03	0.13	0.15	0.95
	$\hat{\beta}^{(2)}$	-0.01	0.07	0.08	0.96	$\hat{\alpha}^{(2)}$	0.03	0.09	0.10	0.96
0.5	$\hat{\beta}^{(0)}$	0.00	0.07	0.08	0.96	$\hat{\alpha}^{(0)}$	0.02	0.08	0.10	0.97
	$\hat{\beta}^{(1)}$	0.00	0.13	0.15	0.97	$\hat{\alpha}^{(1)}$	-0.03	0.13	0.15	0.95
	$\hat{\beta}^{(2)}$	-0.00	0.07	0.08	0.95	$\hat{\alpha}^{(2)}$	0.03	0.09	0.10	0.96
0.7	$\hat{\beta}^{(0)}$	0.00	0.07	0.08	0.96	$\hat{\alpha}^{(0)}$	0.03	0.10	0.11	0.96
	$\hat{\beta}^{(1)}$	0.01	0.14	0.18	0.97	$\hat{\alpha}^{(1)}$	-0.03	0.13	0.15	0.95
	$\hat{\beta}^{(2)}$	-0.00	0.07	0.08	0.96	$\hat{\alpha}^{(2)}$	0.03	0.09	0.10	0.96

las. Specifically, we focused on configuration (II), the case with 30% dependent censoring. With the true Kendall's tau set to be 0.576, we first generated T and D under the Clayton copula, and then estimated the regression coefficients assuming two types of dependence structure. One is the Frank copula with Kendall's tau= 0.576, which represents the situation of mis-specified copula function with correct degree of association, and the other is the Clayton copula with Kendall's tau= $0.79, 0.33$ and 0.16, in which the copula function was true but the association parameters were not. Similarly, we also generated T and D under the Frank copula, and examined the estimation

Table 2.2: Simulation Results on Parameter Estimation under the Frank copula. Bias: biases; AvgSD: average estimated resampling-based standard deviations; EmpSD: empirical standard deviations; Cov95: coverage rates of 95% Wald confidence intervals.

τ		Bias	EmpSD	AvgSD	Cov95		Bias	EmpSD	AvgSD	Cov95
45% dep. censoring, 10% indep. censoring, model (2) for D										
0.1	$\hat{\beta}^{(0)}$	0.03	0.11	0.12	0.94	$\hat{\alpha}^{(0)}$	0.01	0.10	0.12	0.95
	$\hat{\beta}^{(1)}$	0.02	0.20	0.23	0.96	$\hat{\alpha}^{(1)}$	-0.02	0.16	0.19	0.96
	$\hat{\beta}^{(2)}$	-0.04	0.11	0.12	0.95	$\hat{\alpha}^{(2)}$	0.03	0.11	0.12	0.93
0.3	$\hat{\beta}^{(0)}$	0.01	0.08	0.09	0.95	$\hat{\alpha}^{(0)}$	0.01	0.08	0.10	0.93
	$\hat{\beta}^{(1)}$	0.01	0.16	0.19	0.96	$\hat{\alpha}^{(1)}$	-0.02	0.14	0.15	0.95
	$\hat{\beta}^{(2)}$	-0.01	0.08	0.09	0.95	$\hat{\alpha}^{(2)}$	0.02	0.08	0.09	0.96
0.5	$\hat{\beta}^{(0)}$	0.01	0.07	0.08	0.95	$\hat{\alpha}^{(0)}$	0.02	0.08	0.11	0.96
	$\hat{\beta}^{(1)}$	0.01	0.14	0.19	0.98	$\hat{\alpha}^{(1)}$	-0.03	0.12	0.16	0.97
	$\hat{\beta}^{(2)}$	-0.01	0.07	0.08	0.98	$\hat{\alpha}^{(2)}$	0.01	0.07	0.08	0.97
30% dep. censoring, 10% indep. censoring, AFT model for D										
0.1	$\hat{\beta}^{(0)}$	0.01	0.09	0.10	0.96	$\hat{\alpha}^{(0)}$	0.02	0.09	0.09	0.93
	$\hat{\beta}^{(1)}$	0.02	0.17	0.19	0.97	$\hat{\alpha}^{(1)}$	-0.02	0.13	0.14	0.94
	$\hat{\beta}^{(2)}$	-0.02	0.09	0.10	0.96	$\hat{\alpha}^{(2)}$	0.03	0.09	0.09	0.94
0.3	$\hat{\beta}^{(0)}$	0.01	0.07	0.08	0.94	$\hat{\alpha}^{(0)}$	0.01	0.09	0.09	0.94
	$\hat{\beta}^{(1)}$	0.00	0.14	0.15	0.95	$\hat{\alpha}^{(1)}$	-0.02	0.13	0.14	0.94
	$\hat{\beta}^{(2)}$	0.00	0.07	0.07	0.95	$\hat{\alpha}^{(2)}$	0.03	0.09	0.09	0.94
0.5	$\hat{\beta}^{(0)}$	0.00	0.06	0.07	0.96	$\hat{\alpha}^{(0)}$	0.01	0.09	0.09	0.95
	$\hat{\beta}^{(1)}$	0.00	0.12	0.13	0.96	$\hat{\alpha}^{(1)}$	-0.02	0.13	0.14	0.94
	$\hat{\beta}^{(2)}$	0.00	0.06	0.07	0.96	$\hat{\alpha}^{(2)}$	0.03	0.09	0.09	0.94
0.7	$\hat{\beta}^{(0)}$	0.00	0.07	0.08	0.95	$\hat{\alpha}^{(0)}$	0.02	0.09	0.10	0.95
	$\hat{\beta}^{(1)}$	0.00	0.12	0.17	0.97	$\hat{\alpha}^{(1)}$	-0.02	0.13	0.14	0.94
	$\hat{\beta}^{(2)}$	0.00	0.07	0.08	0.97	$\hat{\alpha}^{(2)}$	0.03	0.09	0.09	0.94

when assuming Clayton copula with Kendall's tau= 0.576 and Frank copula with Kendall's tau= $0.26, -0.12$ and -0.33, respectively.

Table 2.3 summarizes the results when we mis-specified the copula function but correctly specified the association parameter, with the dependent censoring rate set to be 30%. Interestingly, the biases are still small and the coverage rates are again close to the nominal level. This suggests that, even with incorrect copula function, we may still obtain unbiased estimation if right knowledge about the degree of association is accessible. In contrast, Figures 2.2 and 2.3 depict the estimated coefficients for T

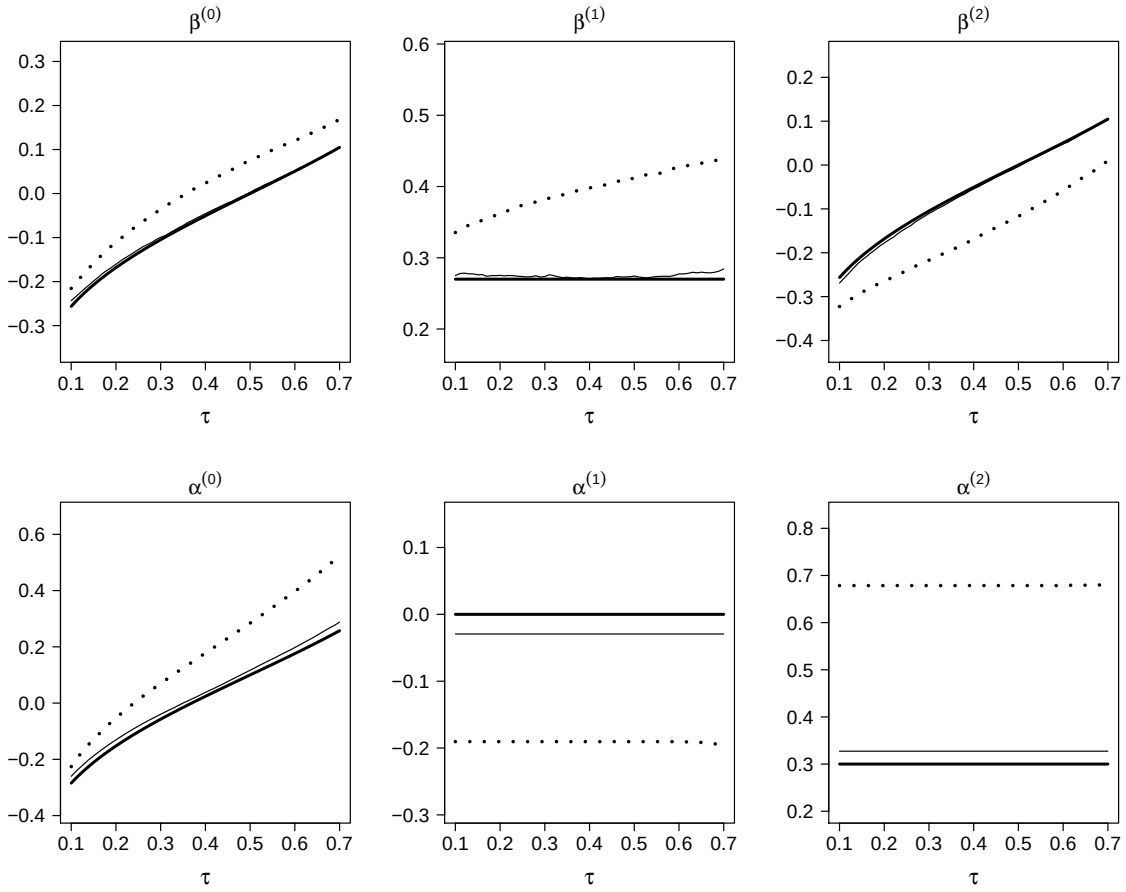

Figure 2.1: Upper Panel: Comparison among True coefficients $\boldsymbol{\beta}_0(\tau)$ (Bold Solid Lines), Mean Estimates for $\boldsymbol{\beta}_0(\tau)$ from the Proposed Method (Solid Lines) under a Correctly Specified Clayton Copula, and Mean Estimates for $\boldsymbol{\beta}_0(\tau)$ from the Naive Approach (Dotted Lines); Lower Panel: Comparison among True coefficients $\boldsymbol{\alpha}_0(\tau)$ (Bold Solid Lines), Mean Estimates for $\boldsymbol{\alpha}_0(\tau)$ from the Proposed Method (Solid Lines) under a Correctly Specified Clayton Copula, and Mean Estimates for $\boldsymbol{\alpha}_0(\tau)$ from the Naive Approach (Dotted Lines).

under correctly specified copula forms with wrong associations. Unsurprisingly, the magnitude of the biases increases with the degree the assumed association deviates from the true value. For example, when the underlying copula was Clayton with Kendall's tau= 0.576, the resulting biases may be moderate (as large as 0.05) for $\boldsymbol{\beta}_0(\tau)$ by assuming Kendall's tau= 0.79 or 0.33, and more pronounced (as large as 0.09) by assuming Kendall's tau= 0.16.

Table 2.3: Simulation Results on Parameter Estimation when the Copula Function is Misspecified. Bias: biases; AvgSD: average estimated resampling-based standard deviations; EmpSD: empirical standard deviations; Cov95: coverage rates of 95% Wald confidence intervals.

τ		Bias	EmpSD	AvgSD	Cov95		Bias	EmpSD	AvgSD	Cov95
Underlying: Clayton, Kendall's tau=0.58; Assumed: Frank, Kendall's tau=0.58										
0.1	$\hat{\beta}^{(0)}$	-0.01	0.09	0.10	0.94	$\hat{\alpha}^{(0)}$	0.01	0.08	0.09	0.97
	$\hat{\beta}^{(1)}$	-0.01	0.17	0.18	0.95	$\hat{\alpha}^{(1)}$	-0.01	0.13	0.14	0.95
	$\hat{\beta}^{(2)}$	0.01	0.09	0.10	0.95	$\hat{\alpha}^{(2)}$	-0.02	0.09	0.10	0.94
0.3	$\hat{\beta}^{(0)}$	-0.01	0.07	0.08	0.95	$\hat{\alpha}^{(0)}$	0.00	0.08	0.09	0.96
	$\hat{\beta}^{(1)}$	-0.01	0.13	0.15	0.96	$\hat{\alpha}^{(1)}$	-0.01	0.13	0.14	0.95
	$\hat{\beta}^{(2)}$	0.02	0.07	0.08	0.95	$\hat{\alpha}^{(2)}$	-0.02	0.09	0.10	0.94
0.5	$\hat{\beta}^{(0)}$	0.00	0.07	0.08	0.95	$\hat{\alpha}^{(0)}$	0.01	0.08	0.10	0.97
	$\hat{\beta}^{(1)}$	0.01	0.13	0.15	0.96	$\hat{\alpha}^{(1)}$	-0.01	0.13	0.14	0.95
	$\hat{\beta}^{(2)}$	0.00	0.07	0.07	0.95	$\hat{\alpha}^{(2)}$	-0.02	0.09	0.10	0.94
0.7	$\hat{\beta}^{(0)}$	0.00	0.07	0.08	0.96	$\hat{\alpha}^{(0)}$	0.07	0.10	0.11	0.94
	$\hat{\beta}^{(1)}$	0.05	0.14	0.18	0.97	$\hat{\alpha}^{(1)}$	-0.01	0.13	0.14	0.95
	$\hat{\beta}^{(2)}$	-0.02	0.07	0.08	0.95	$\hat{\alpha}^{(2)}$	-0.02	0.09	0.10	0.94
Underlying: Frank, Kendall's tau=0.58; Assumed: Clayton, Kendall's tau=0.58										
0.1	$\hat{\beta}^{(0)}$	0.04	0.09	0.10	0.93	$\hat{\alpha}^{(0)}$	0.04	0.09	0.10	0.92
	$\hat{\beta}^{(1)}$	0.03	0.16	0.19	0.96	$\hat{\alpha}^{(1)}$	-0.04	0.14	0.14	0.93
	$\hat{\beta}^{(2)}$	-0.05	0.09	0.10	0.93	$\hat{\alpha}^{(2)}$	0.07	0.08	0.09	0.86
0.3	$\hat{\beta}^{(0)}$	0.03	0.07	0.08	0.93	$\hat{\alpha}^{(0)}$	0.03	0.09	0.09	0.93
	$\hat{\beta}^{(1)}$	0.01	0.13	0.15	0.96	$\hat{\alpha}^{(1)}$	-0.04	0.14	0.14	0.93
	$\hat{\beta}^{(2)}$	-0.02	0.07	0.07	0.96	$\hat{\alpha}^{(2)}$	0.07	0.08	0.09	0.86
0.5	$\hat{\beta}^{(0)}$	0.01	0.07	0.07	0.95	$\hat{\alpha}^{(0)}$	0.01	0.09	0.09	0.95
	$\hat{\beta}^{(1)}$	-0.01	0.12	0.14	0.96	$\hat{\alpha}^{(1)}$	-0.04	0.14	0.14	0.93
	$\hat{\beta}^{(2)}$	0.00	0.06	0.07	0.96	$\hat{\alpha}^{(2)}$	0.07	0.08	0.09	0.86
0.7	$\hat{\beta}^{(0)}$	0.00	0.07	0.08	0.94	$\hat{\alpha}^{(0)}$	-0.01	0.09	0.10	0.95
	$\hat{\beta}^{(1)}$	-0.03	0.13	0.15	0.96	$\hat{\alpha}^{(1)}$	-0.04	0.14	0.14	0.93
	$\hat{\beta}^{(2)}$	0.02	0.07	0.08	0.96	$\hat{\alpha}^{(2)}$	0.07	0.08	0.09	0.86

2.3 The WASID Study Example

We applied the proposed method to the WASID study (Chimowitz et al., 2005), a double-blind and multicenter clinical trial that compared warfarin and aspirin in treating symptomatic intracranial arterial stenosis, an important cause of stroke. In this trial, 569 patients who had stroke or transient ischemic attack resulting from

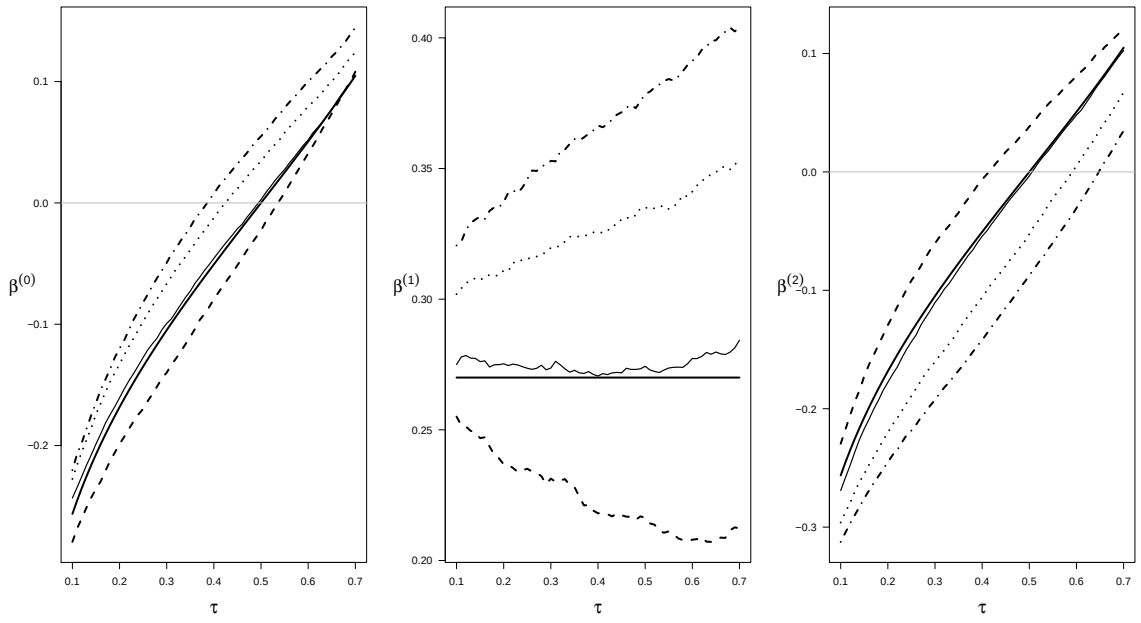

Figure 2.2: Estimates for $\boldsymbol{\beta}_0(\tau)$ under the Correctly Specified Clayton Copula with Misspecified Association Parameters: Kendall's tau= 0.79 (Dashed Lines), Kendall's tau= 0.33 (dotted Lines), Kendall's tau= 0.16 (Dotdash Lines), and the True Association Parameter: Kendall's tau= 0.58 (Solid Lines); and the True Coefficients $\boldsymbol{\beta}_0(\tau)$ (Bold Solid Lines).

stenosis of a major intracranial artery were randomized to receive either warfarin or aspirin. In our analysis, T was defined as time from randomization to ischemic stroke, brain hemorrhage, or death, whichever happened first. Here and hereafter, we refer to this event as the "study endpoint". During an average of 1.8-year follow-up, T was observed for 57 patients treated by warfarin and 60 patients treated by aspirin. Due to various disease related reasons, the study medications were terminated early for 125 patients, among whom 81 were on the warfarin arm and 44 were on the aspirin arm. It is often thought that such discontinuation of treatment is correlated with the underlying disease progression and thus pose dependent censoring to T. We let D denote time from randomization to study withdrawal. In addition, administrative censoring occurred for 146 patients in the warfarin group and 172 patients in the aspirin group. Time to such independent censoring was denoted by C. We considered

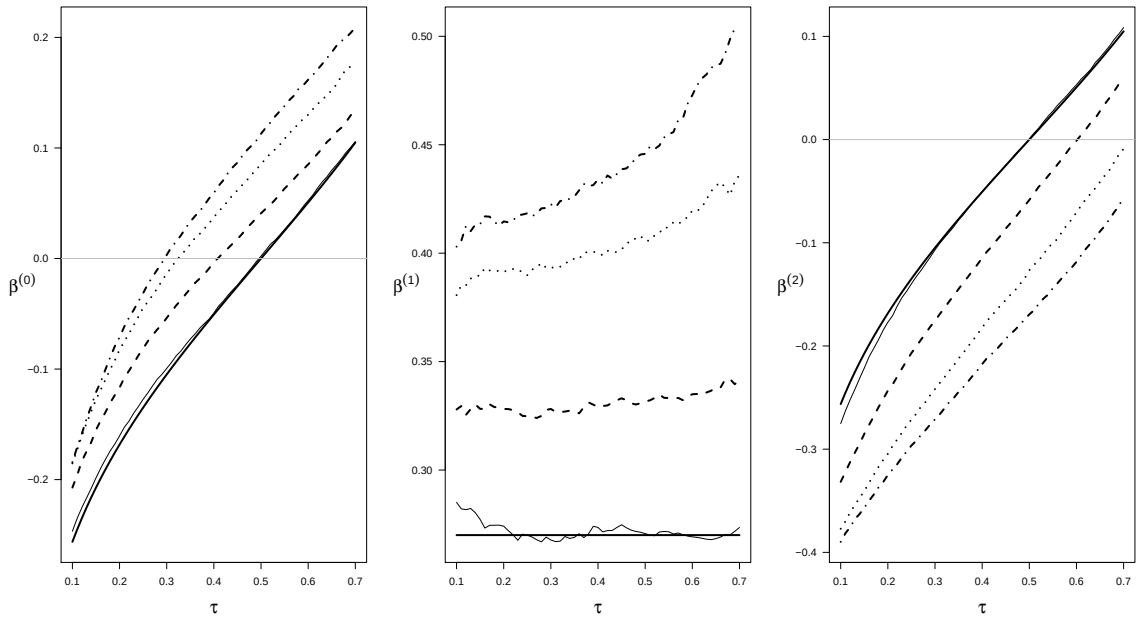

Figure 2.3: Estimates for $\boldsymbol{\beta}_0(\tau)$ under the Correctly Specified Frank Copula with Misspecified Association Parameters: Kendall's tau= 0.26 (Dashed Lines), Kendall's tau= -0.12 (Dotted Lines), Kendall's tau= -0.33 (Dotdash Lines), and the True Association Parameter: Kendall's tau= 0.58 (Solid Lines); and the True Coefficients $\boldsymbol{\beta}_0(\tau)$ (Bold Solid Lines).

three covariates: Treatment, which equals 1 for warfarin and 0 for aspirin; Diabetes, the indicator of having diabetes; Stenosis Percentage, which stands for the percentage of stenosis by central reader.

We first analyzed the WASID data based on some classical approaches, naively treating early drug termination as independent censoring. No treatment effect was detected by the log rank test. Adjusting for Diabetes and Stenosis Percentage, the Cox regression also suggested that there was no significant treatment effect. The hazard ratio of warfarin versus aspirin was 0.91 with p-value=0.63. Stenosis Percentage was not found to be significant in predicting time to the study endpoint either. Having diabetes was found to have a significant negative effect on the progression to the study endpoint. The corresponding hazard ratio and p value were 2.15 and < 0.001 respectively.

We then applied the proposed regression approach adjusting for the same set of covariates considered in the naive analysis. We specified different r values such that the corresponding Kendall's tau were 0.2, 0.4, 0.6, and 0.8, representing the cases where the positive associations between T and D were weak, moderate and strong. The link function was chosen to be $\log(\cdot)$. Due to heavy censoring to D by T or C with the censoring rate around 80%, we adopted an AFT model for D to increase numerical stability. For inference, we performed 300 bootstrap resampling for each scenario. We considered both Clayton copula and Frank copula. Nevertheless we only present the results based on the Clayton copula, since the results under the Frank copula are very similar and thus are omitted.

Figure 2.4 depicts the estimates for $\boldsymbol{\beta}_0(\tau)$ under the Clayton copula, together with the results from a naive application of Peng and Huang (2008) in which D was treated as independent censoring. From Figure 2.4 we observe that, the naive estimate and the proposed estimates for the treatment effect appear to be similar for $\tau < 0.18$ and demonstrate a larger yet moderate divergence for later τs. In all cases, the estimated treatment effects over τ demonstrate a common pattern: being negative at lower quantiles and then decreasing in the magnitudes and becoming stabilized around 0. For Diabetes and Stenosis Percentage, the departure of the estimates that assume dependent censoring from the naive estimate are more noticeable.

In Table 2.4, we summarize the standard errors of the naive estimates and the proposed estimates under different specifications of r with $r > 0$. It can be seen that the proposed estimates have comparable efficiency to the naive estimate obtained by Peng and Huang (2008)'s method. We also performed the second-stage inference procedure on the WASID data. Formal tests on the significance of covariate effects were performed based on the average effects on quantiles of T with τ ranging from 0.05 to 0.25. Results show that the treatment effect was not significant for any choice of r we considered. This is consistent with Chimowitz et al. (2005), which found

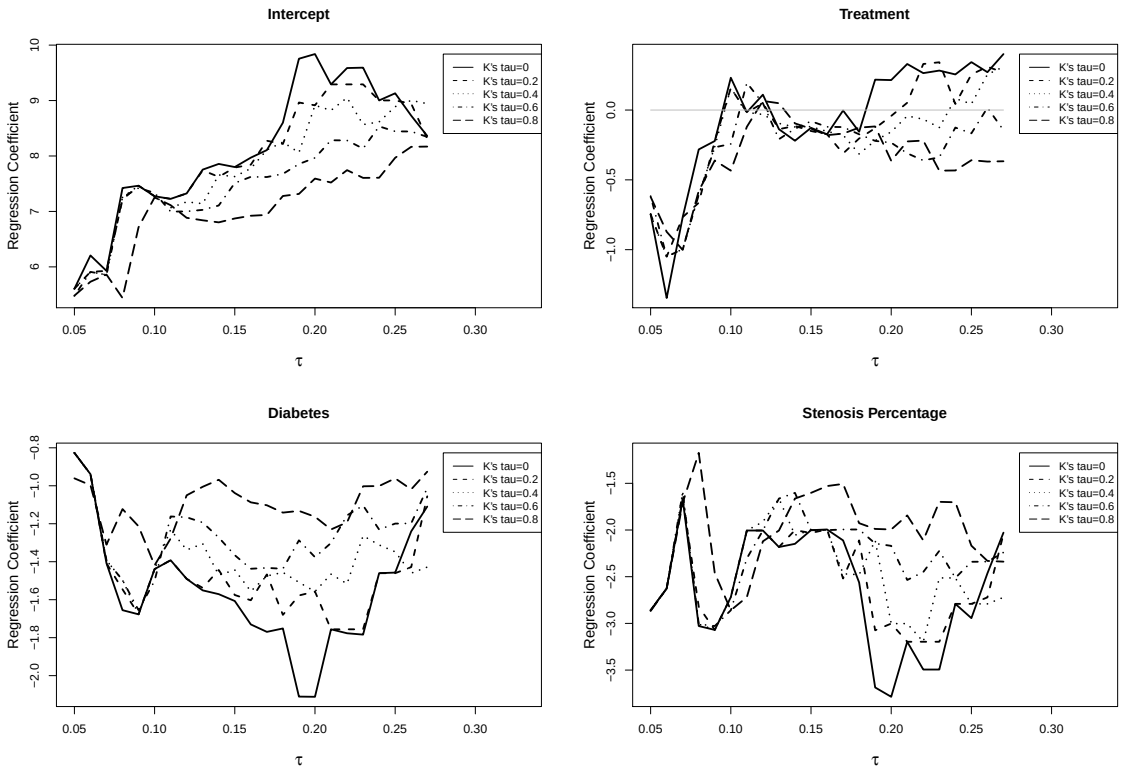

Figure 2.4: Point Estimates of Regression Coefficients for Time to the Primary Endpoint (Ischemic Stroke, Brain Hemorrhage, or Death) under the Clayton Copula with Kendall's tau=0, 0.2, 0.4, 0.6 and 0.8.

no benefit of warfarin over aspirin in the WASID trial. However, we found that Diabetes has significant effects under all choices of r (all p-values < 0.001), with the average effects being -1.58, -1.49, -1.38, -1.28 and -1.11, corresponding to the cases where Kendall's tau $= 0, 0.2, 0.4, 0.6$ and 0.8, respectively. This result suggests the diabetic patients may progress significantly faster to the study endpoint compared to nondiabetic patients. This finding is consistent with the naive Cox regression analysis, but is better endorsed by taking into account the potential dependence between T and D.

To illustrate the impact of adjusting for dependent censoring in a more meaningful way, in Figures 2.5 and 2.6 we plot the estimated quantiles of T and D for each treatment group with and without diabetes, with Stenosis Percentage fixed at

Table 2.4: The WASID Example: Standard Errors under the Clayton Copula with Kendall's tau=0, 0.2, 0.4, 0.6 and 0.8. $\hat{\beta}^{(1)}$, $\hat{\beta}^{(2)}$ and $\hat{\beta}^{(3)}$: estimated coefficients of Treatment, Diabetes and Stenosis Percentage on T, respectively.

τ		K's tau=0	K's tau=0.2	K's tau=0.4	K's tau=0.6	K's tau=0.8
0.05	$\hat{\beta}^{(1)}$	0.79	0.77	0.78	0.76	0.72
	$\hat{\beta}^{(2)}$	0.77	0.75	0.73	0.73	0.69
	$\hat{\beta}^{(3)}$	2.10	2.04	2.04	2.04	1.95
0.10	$\hat{\beta}^{(1)}$	0.69	0.69	0.69	0.71	0.74
	$\hat{\beta}^{(2)}$	0.65	0.65	0.64	0.66	0.69
	$\hat{\beta}^{(3)}$	1.83	1.67	1.79	1.76	1.73
0.15	$\hat{\beta}^{(1)}$	0.43	0.41	0.37	0.36	0.35
	$\hat{\beta}^{(2)}$	0.41	0.38	0.37	0.36	0.39
	$\hat{\beta}^{(3)}$	1.45	1.38	1.33	1.26	1.17
0.20	$\hat{\beta}^{(1)}$	0.51	0.46	0.40	0.39	0.33
	$\hat{\beta}^{(2)}$	0.60	0.48	0.44	0.34	0.32
	$\hat{\beta}^{(3)}$	1.76	1.48	1.42	1.24	1.05
0.25	$\hat{\beta}^{(1)}$	0.53	0.50	0.46	0.41	0.35
	$\hat{\beta}^{(2)}$	0.79	0.66	0.54	0.45	0.33
	$\hat{\beta}^{(3)}$	2.24	1.85	1.67	1.46	1.19

its mean value. From Figure 2.5, it is apparent that the disparity among different estimates is negligible in the diabetes group, but accounting for dependent censoring at different levels can lead to quite dramatically different estimates for $Q_T(\tau|\boldsymbol{Z})$ in the non-diabetic group. One plausible explanation for this is that non-diabetic patients generally progress to the study endpoint slower than diabetic patients and thus are more prone to the "risk" of early termination of study medication. Consequently, adjusting for dependent censoring for the non-diabetic patients makes a bigger influence on the estimated quantiles of T. It is also interesting to note from Figure E3 that assuming independence between patient withdrawal and the study endpoint tends to give more optimistic estimate for $Q_T(\tau|\boldsymbol{Z})$ compared to the other cases where T and D were assumed to be positively associated. This phenomenon is also reasonable. An intuitive explanation may be that an observed D (which means $T > D$ and T is censored) would be suggestive of a smaller T when T and D are believed to be pos-

itively associated than that under the independence between T and D. As a result, the prediction of $Q_T(\tau|\boldsymbol{Z})$ would be more conservative under a positive association assumption. From Figure 2.6 we can see that the warfarin group tends to have smaller D compared to the aspirin group, which means the patients treated by warfarin tend to withdraw earlier than the other group. This is also consistent with Chimowitz et al. (2005), which found a higher rate of adverse events in the warfarin group than in the aspirin group.

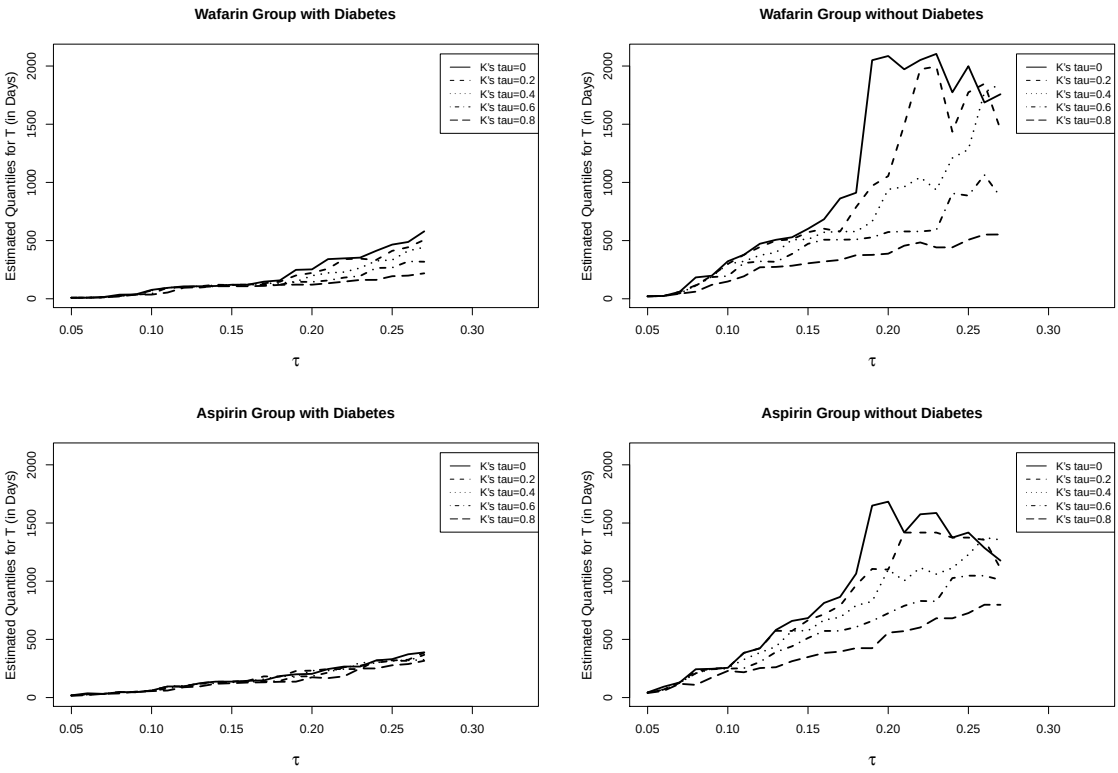

Figure 2.5: Estimated Quantiles of Time to the Primary Endpoint (Ischemic Stroke, Brain Hemorrhage, or Death) under the Clayton Copula with Kendall's tau=0, 0.2, 0.4, 0.6 and 0.8, with the Stenosis Percentage Fixed at Its Mean (63.7%)

In summary, in the WASID example we found no evidence of better clinical efficacy for warfarin compared to aspirin in treating symptomatic intracranial arterial stenosis, which is consistent with previously published results on this trial. In our analysis, we took into account of the dependence between T and D and provided a

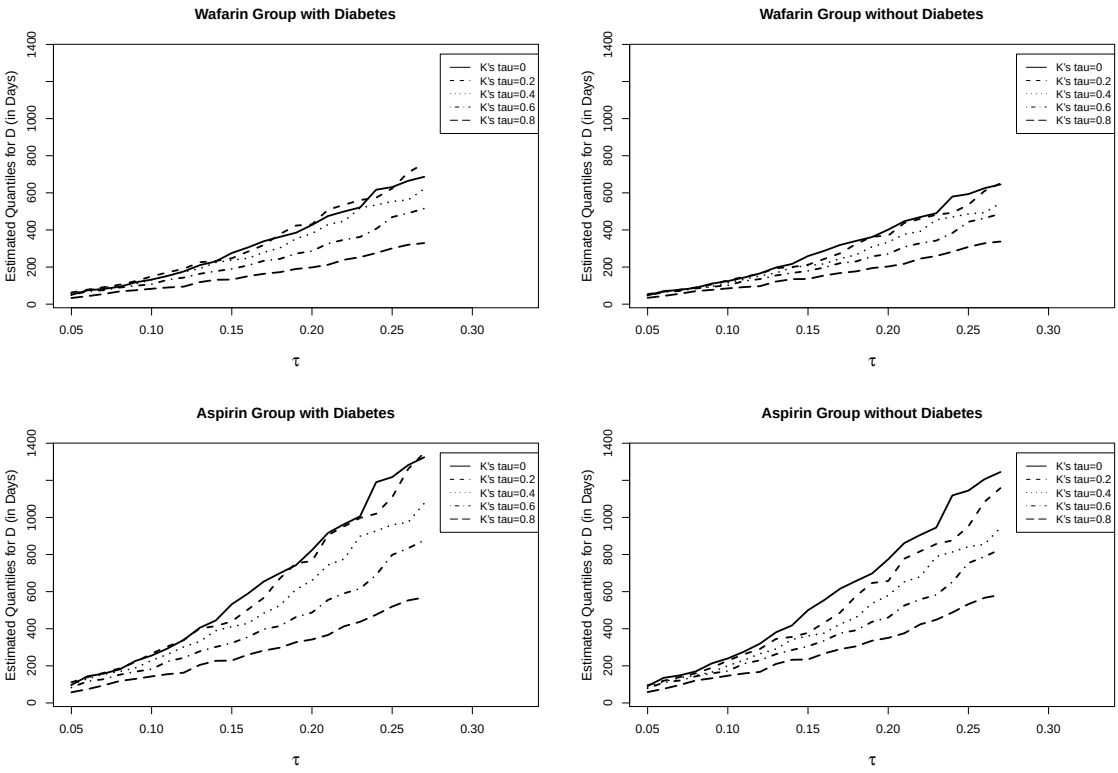

Figure 2.6: Estimated Quantiles of Time to Early Termination of Study Medication under the Clayton Copula with Kendall's tau=0, 0.2, 0.4, 0.6 and 0.8, with the Stenosis Percentage Fixed at Its Mean (63.7%).

comprehensive view of the covariate effects under different specifications of the association. The results we obtained are quite consistent across assumptions of weak, moderate and strong associations between T and D, and therefore more confidence is gained to support the scientific conclusions of Chimowitz et al. (2005) through this new analysis. Our method also enables us to explore dynamic patterns of the covariate effects across different quantiles of T. The predicted conditional quantiles of T provide intuitive and robust prognostic information to physicians and patients in clinical practice.

2.4 Remarks

In this chapter we propose a quantile regression method for survival data subject to dependent censoring. Under the assumed model for the event time of interest, covariate effects are formulated on the quantiles defined on the marginal survival distribution. This type of modeling is sensible when the scenario corresponding the removal of the dependent censoring event is scientifically relevant.

The dependence structure between the event time and the censoring time is specified through a known copula model. This is necessary given the competing risks relationship between the event of interest and dependent censoring. Our numerical results show that the proposed method is quite robust to misspecification of the type of the copula function, provided the strength of association is reasonably specified. When there lacks sufficient information to support a specific choice of the copula parameter, our regression procedure developed based on quantile regression modeling offers a robust sensitive analysis tool for analyzing dependently censored data.

2.5 Proofs

2.5.1 Regularity Conditions

Before stating the regularity conditions, we introduce some necessary notation. For a random variable W, define $F_W(t|\boldsymbol{Z}) = \Pr(W \le t|\boldsymbol{Z})$ and $f_W(t|\boldsymbol{Z}) = dF_W(t|\boldsymbol{Z})/dt$. Define $\tilde{F}_T(t|\boldsymbol{Z}) = \Pr(T \le t, \delta = 1|\boldsymbol{Z})$ and $\tilde{F}_D(t|\boldsymbol{Z}) = \Pr(D \le t, \delta = 1|\boldsymbol{Z})$, and $\tilde{f}_T(t)$ and $\tilde{f}_D(t)$ as the derivatives of $\tilde{F}_T(t|\boldsymbol{Z})$ and $\tilde{F}_D(t|\boldsymbol{Z})$ with respect to t, respectively. For a vector $\boldsymbol{r}$, let $\boldsymbol{r}^{\otimes 2}$ denote $\boldsymbol{r}\boldsymbol{r}^T$ and $\|\boldsymbol{r}\|$ denote the Euclidean norm of $\boldsymbol{r}$. Define $\boldsymbol{s}^{(k)}(\boldsymbol{\beta}, \boldsymbol{\alpha}, \tau) = E\{\boldsymbol{S}_n^{(k)}(\boldsymbol{\beta}, \boldsymbol{\alpha}, \tau)\}$ for $k = 1, 2$, $\boldsymbol{\mu}_1(\boldsymbol{b}) = E[\boldsymbol{Z}N_1\{g(\boldsymbol{Z}^T\boldsymbol{b})\}]$, $\boldsymbol{\mu}_2(\boldsymbol{a}) = E[\boldsymbol{Z}N_2\{g(\boldsymbol{Z}^T\boldsymbol{a})\}]$, $\boldsymbol{J}_1(\boldsymbol{b}) = E[\boldsymbol{Z}^{\otimes 2}\tilde{f}_T\{g(\boldsymbol{Z}^T\boldsymbol{b})|\boldsymbol{Z}\}g'(\boldsymbol{Z}^T\boldsymbol{b})]$, $\boldsymbol{J}_2(\boldsymbol{a}) = E[\boldsymbol{Z}^{\otimes 2}\tilde{f}_D\{g(\boldsymbol{Z}^T\boldsymbol{a})|\boldsymbol{Z}\}$ $g'(\boldsymbol{Z}^T\boldsymbol{a})]$, $\tilde{\boldsymbol{\mu}}_1(\boldsymbol{b}, \boldsymbol{\alpha}) = E[\boldsymbol{Z}Y\{g(\boldsymbol{Z}^T\boldsymbol{b})\} \times \phi_1(1-u, 1-\int_0^{\tau_{U,2}} I\{\boldsymbol{Z}^T\boldsymbol{\alpha}(v) \le \boldsymbol{Z}^T\boldsymbol{b}\}\,\mathrm{d}v)]$, $\tilde{\boldsymbol{\mu}}_2(\boldsymbol{a}, \boldsymbol{\beta}) = E[\boldsymbol{Z}Y\{g(\boldsymbol{Z}^T\boldsymbol{a})\} \times \phi_2(1-\int_0^{\tau_{U,1}} I\{\boldsymbol{Z}^T\boldsymbol{\beta}(v) \le \boldsymbol{Z}^T\boldsymbol{a}\}\,\mathrm{d}v, 1-u)]$, $\tilde{\boldsymbol{J}}_1(\boldsymbol{b}, \boldsymbol{\alpha}) = \partial\tilde{\boldsymbol{\mu}}_1(\boldsymbol{b}, \boldsymbol{\alpha})/\partial\boldsymbol{b}$, $\tilde{\boldsymbol{J}}_2(\boldsymbol{a}, \boldsymbol{\beta}) = \partial\tilde{\boldsymbol{\mu}}_2(\boldsymbol{a}, \boldsymbol{\beta})/\partial\boldsymbol{a}$.

For any constant d, define $\mathscr{B}(d) = \{(\boldsymbol{b}^T, \boldsymbol{a}^T)^T : \boldsymbol{b} \in \mathbb{R}^{p+1}, \boldsymbol{a} \in \mathbb{R}^{p+1}, \inf_{\tau\in(0,\tau_U]} \|\boldsymbol{\mu}_1(\boldsymbol{b}) - \boldsymbol{\mu}_1\{\boldsymbol{\beta}_0(\tau)\}\| \le d, \inf_{\tau\in(0,\tau_U]} \|\boldsymbol{\mu}_2(\boldsymbol{a}) - \boldsymbol{\mu}_2\{\boldsymbol{\alpha}_0(\tau)\}\| \le d\}$. Let $\mathcal{D}$ denote a function space that contains all continuous functions mapping $[0, 1]$ to $\mathbb{R}^{2p+2}$, and $\mathcal{F} = \{c(\boldsymbol{G}_1 - \boldsymbol{G}_2) : c \in \mathbb{R}, \boldsymbol{G}_j \in \mathcal{D}, j = 1, 2\}$.

The regularity conditions are:

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

C2. $f_T(t|\boldsymbol{z})$, $f_D(t|\boldsymbol{z})$, $\tilde{f}_T(t|\boldsymbol{z})$ and $\tilde{f}_D(t|\boldsymbol{z})$ are bounded above uniformly in t and $\boldsymbol{z}$.

C3. (a) $\tilde{f}_T\{g(\boldsymbol{Z}^T\boldsymbol{b})|\boldsymbol{Z}\} > 0$ and $\tilde{f}_D\{g(\boldsymbol{Z}^T\boldsymbol{a})|\boldsymbol{Z}\} > 0$ for all $(\boldsymbol{b}^T, \boldsymbol{a}^T)^T \in \mathscr{B}(d_0)$, where d_0 is a constant; (b) $E(\boldsymbol{Z}^{\otimes 2}) > 0$; (c) $\inf_{\tau\in[\nu_1,\tau_{U,1}]} eigmin(\boldsymbol{J}_1\{\boldsymbol{\beta}_0(\tau)\}) > 0$ and $\inf_{\tau\in[\nu_2,\tau_{U,2}]} eigmin($

$\boldsymbol{J}_2\{\boldsymbol{\alpha}_0(\tau)\}) > 0$ for any $\nu_1 \in (0, \tau_{U,1}]$ and $\nu_2 \in (0, \tau_{U,2}]$, where $eigmin(\cdot)$ denotes the minimum eigenvalue of a matrix.

C4. (a) Each component of $\tilde{\boldsymbol{\mu}}_1\{\boldsymbol{\beta}_0(\tau), \boldsymbol{\alpha}_0\}$ and $\tilde{\boldsymbol{\mu}}_2\{\boldsymbol{\alpha}_0(\tau), \boldsymbol{\beta}_0\}$ is a Lipschitz function of τ; (b) $\phi_1(u, v)$ and $\phi_2(u, v)$ are differentiable with respect to u and v, and

furthermore, each component of $E\big[\boldsymbol{Z}Y\{g(\boldsymbol{Z}^T\boldsymbol{b})\}\times\phi_{12}(1-u,1-\int_0^{\tau_{U,2}}I\{\boldsymbol{Z}^T\boldsymbol{\alpha}_0(v)\le \boldsymbol{Z}^T\boldsymbol{b}\}\,\mathrm{d}v)\big]$ and $E\big[\boldsymbol{Z}Y\{g(\boldsymbol{Z}^T\boldsymbol{a})\}\times\phi_{21}(1-\int_0^{\tau_{U,1}}I\{\boldsymbol{Z}^T\boldsymbol{\beta}_0(v)\le \boldsymbol{Z}^T\boldsymbol{a}\}\,\mathrm{d}v,1-u)\big]$ is bounded uniformly in $(\boldsymbol{b}^T,\boldsymbol{a}^T)^T\in\mathscr{B}(d_0)$, where $\phi_{12}(u,v)=\partial\phi_1(u,v)/\partial v$ and $\phi_{21}(u,v)=\partial\phi_2(u,v)/\partial u$; (c) $\boldsymbol{w}_1(\boldsymbol{b},\boldsymbol{a})=E\big[\boldsymbol{Z}Y\{g(\boldsymbol{Z}^T\boldsymbol{b})\}\times\phi_{12}(1-u,1-\int_0^{\tau_{U,2}}I\{\boldsymbol{Z}^T\boldsymbol{\alpha}_0(v)\le \boldsymbol{Z}^T\boldsymbol{b}\}\,\mathrm{d}v)\times I(\boldsymbol{Z}^T\boldsymbol{b}\ge\boldsymbol{Z}^T\boldsymbol{a})\big]$ and $\boldsymbol{w}_2(\boldsymbol{b},\boldsymbol{a})=E\big[\boldsymbol{Z}Y\{g(\boldsymbol{Z}^T\boldsymbol{a})\}\times\phi_{21}(1-\int_0^{\tau_{U,1}}I\{\boldsymbol{Z}^T\boldsymbol{\beta}_0(v)\le \boldsymbol{Z}^T\boldsymbol{a}\}\,\mathrm{d}v,1-u)\times I(\boldsymbol{Z}^T\boldsymbol{a}\ge\boldsymbol{Z}^T\boldsymbol{b})\big]$ are differentiable with respect to $\boldsymbol{b}$ and $\boldsymbol{a}$, and furthermore, each component of $\boldsymbol{w}_{12}\{\boldsymbol{b},\boldsymbol{\alpha}_0(\tau)\}$ and $\boldsymbol{w}_{21}\{\boldsymbol{\beta}_0(\tau),\boldsymbol{a}\}$ is bounded uniformly in $(\boldsymbol{b}^T,\boldsymbol{a}^T)^T\in\mathscr{B}(d_0)$ and $\tau\in(0,\max\{\tau_{U,1},\tau_{U,2}\}]$, where $\boldsymbol{w}_{12}(\boldsymbol{b},\boldsymbol{a})=\partial\boldsymbol{w}_1(\boldsymbol{b},\boldsymbol{a})/\partial\boldsymbol{a}$ and $\boldsymbol{w}_{21}(\boldsymbol{b},\boldsymbol{a})=\partial\boldsymbol{w}_2(\boldsymbol{b},\boldsymbol{a})/\partial\boldsymbol{b}$.

C5. (a) For any fixed τ, $\boldsymbol{\rho}(\boldsymbol{\theta},\tau)$, as a functional of $\boldsymbol{\theta}(\cdot)$ defined on $\mathcal{D}$, is Gâteaux differentiable at $\boldsymbol{\theta}_0(\cdot)$ with derivative $\boldsymbol{\rho}'_{\boldsymbol{\theta}_0}$, where $\boldsymbol{\theta}_0(u)=\big(\boldsymbol{\mu}_1\{\boldsymbol{\beta}_0(u)\}^T,\boldsymbol{\mu}_2\{\boldsymbol{\alpha}_0(u)\}^T\big)^T$; (b) $\|\boldsymbol{\rho}'_{\boldsymbol{\theta}_0}(\boldsymbol{h})\|>0$ for any $\boldsymbol{h}\in\mathcal{F}$ such that $\sup_{\tau\in(0,1)}\|\boldsymbol{h}(\tau)\|\neq 0$.

2.5.2 Proof of Theorem 2.1.1

For simplicity, we assume $\tau_{U,1}=\tau_{U,2}=\tau_U$. We present the proof of both theorems based on the estimating equation (2.9), which can be adapted to the proof based on the estimating equation (2.10) with minor modification. Let $\boldsymbol{\beta}$, $\boldsymbol{\alpha}$, $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\alpha}}$ be abbreviations for functions $\boldsymbol{\beta}(\cdot)$, $\boldsymbol{\alpha}(\cdot)$, $\hat{\boldsymbol{\beta}}(\cdot)$ and $\hat{\boldsymbol{\alpha}}(\cdot)$. With fixed $\boldsymbol{\alpha}$, define $\hat{\boldsymbol{\beta}}(\boldsymbol{\alpha},\tau)$ and $\tilde{\boldsymbol{\beta}}(\boldsymbol{\alpha},\tau)$ as the solutions for $\boldsymbol{\beta}$ to $\boldsymbol{S}_n^{(1)}(\boldsymbol{\beta},\boldsymbol{\alpha},\tau)=0$ and $\boldsymbol{s}^{(1)}(\boldsymbol{\beta},\boldsymbol{\alpha},\tau)=0$ respectively. Similarly, with fixed $\boldsymbol{\beta}$, define $\hat{\boldsymbol{\alpha}}(\boldsymbol{\beta},\tau)$ and $\tilde{\boldsymbol{\alpha}}(\boldsymbol{\beta},\tau)$ as the solutions for $\boldsymbol{\alpha}$ to $\boldsymbol{S}_n^{(2)}(\boldsymbol{\beta},\boldsymbol{\alpha},\tau)=0$ and $\boldsymbol{s}^{(2)}(\boldsymbol{\beta},\boldsymbol{\alpha},\tau)=0$ respectively. It is easy to see that $\hat{\boldsymbol{\beta}}(\hat{\boldsymbol{\alpha}},\tau)=\hat{\boldsymbol{\beta}}(\tau)$, $\hat{\boldsymbol{\alpha}}(\hat{\boldsymbol{\beta}},\tau)=\hat{\boldsymbol{\alpha}}(\tau)$, $\tilde{\boldsymbol{\beta}}(\boldsymbol{\alpha}_0,\tau)=\boldsymbol{\beta}_0(\tau)$ and $\tilde{\boldsymbol{\alpha}}(\boldsymbol{\beta}_0,\tau)=\boldsymbol{\alpha}_0(\tau)$ for $\tau\in(0,\tau_U]$.

Using the Glivenko-Cantelli Theorem (van der Vaart and Wellner 1996), we can show

$$\sup_{\tau\in(0,\tau_U]}\|\boldsymbol{S}_n^{(k)}(\boldsymbol{\beta},\boldsymbol{\alpha},\tau)-\boldsymbol{s}^{(k)}(\boldsymbol{\beta},\boldsymbol{\alpha},\tau)\|\xrightarrow{p}0,k=1,2. \tag{2.13}$$

Together with the facts that $\boldsymbol{S}_n^{(k)}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\alpha}}, \tau) = 0$ and $\boldsymbol{s}^{(k)}(\boldsymbol{\beta}_0, \boldsymbol{\alpha}_0, \tau) = 0$, (2.13) implies

$$\sup_{\tau \in (0, \tau_U]} \|\boldsymbol{s}^{(1)}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\alpha}}, \tau) - \boldsymbol{s}^{(1)}(\boldsymbol{\beta}_0, \boldsymbol{\alpha}_0, \tau)\| \xrightarrow{p} 0,$$
$$\sup_{\tau \in (0, \tau_U]} \|\boldsymbol{s}^{(2)}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\alpha}}, \tau) - \boldsymbol{s}^{(2)}(\boldsymbol{\beta}_0, \boldsymbol{\alpha}_0, \tau)\| \xrightarrow{p} 0.$$

Let

$$\boldsymbol{r} = \begin{pmatrix} \boldsymbol{b} \\ \boldsymbol{a} \end{pmatrix}, \quad \boldsymbol{\mu}(\boldsymbol{r}) = \begin{pmatrix} \boldsymbol{\mu}_1(\boldsymbol{b}) \\ \boldsymbol{\mu}_2(\boldsymbol{a}) \end{pmatrix} \quad \text{and} \quad \mathscr{A}(d) = \{\boldsymbol{\mu}(\boldsymbol{r}) : \boldsymbol{r} \in \mathscr{B}(d)\}.$$

By condition C3(a)-(b), we can show that $\boldsymbol{\mu}(\cdot)$ is a 1-1 mapping from $\mathscr{B}(d_0)$ to $\mathscr{A}(d_0)$. Hence, there exists an inverse function of $\boldsymbol{\mu}(\cdot)$, denoted by $\boldsymbol{\kappa}(\cdot)$, mapping $\mathscr{A}(d_0)$ to $\mathscr{B}(d_0)$, such that $\boldsymbol{\kappa}\{\boldsymbol{\mu}(\boldsymbol{r})\} = \boldsymbol{r}$ for any $\boldsymbol{r} \in \mathscr{B}(d_0)$.

Consider the following equalities:

$$\begin{aligned}
\boldsymbol{\mu}_1\{\hat{\boldsymbol{\beta}}(\tau)\} - \boldsymbol{\mu}_1\{\boldsymbol{\beta}_0(\tau)\} &= \boldsymbol{\mu}_1\{\hat{\boldsymbol{\beta}}(\hat{\boldsymbol{\alpha}}, \tau)\} - \boldsymbol{\mu}_1\{\tilde{\boldsymbol{\beta}}(\boldsymbol{\alpha}_0, \tau)\} \\
&= \boldsymbol{\mu}_1\{\hat{\boldsymbol{\beta}}(\hat{\boldsymbol{\alpha}}, \tau)\} - \boldsymbol{\mu}_1\{\tilde{\boldsymbol{\beta}}(\hat{\boldsymbol{\alpha}}, \tau)\} + \boldsymbol{\mu}_1\{\tilde{\boldsymbol{\beta}}(\hat{\boldsymbol{\alpha}}, \tau)\} - \boldsymbol{\mu}_1\{\tilde{\boldsymbol{\beta}}(\boldsymbol{\alpha}_0, \tau)\}, \\
\boldsymbol{\mu}_2\{\hat{\boldsymbol{\alpha}}(\tau)\} - \boldsymbol{\mu}_2\{\boldsymbol{\alpha}_0(\tau)\} &= \boldsymbol{\mu}_2\{\hat{\boldsymbol{\alpha}}(\hat{\boldsymbol{\beta}}, \tau)\} - \boldsymbol{\mu}_2\{\tilde{\boldsymbol{\alpha}}(\boldsymbol{\beta}_0, \tau)\} \\
&= \boldsymbol{\mu}_2\{\hat{\boldsymbol{\alpha}}(\hat{\boldsymbol{\beta}}, \tau)\} - \boldsymbol{\mu}_2\{\tilde{\boldsymbol{\alpha}}(\hat{\boldsymbol{\beta}}, \tau)\} + \boldsymbol{\mu}_2\{\tilde{\boldsymbol{\alpha}}(\hat{\boldsymbol{\beta}}, \tau)\} - \boldsymbol{\mu}_2\{\tilde{\boldsymbol{\alpha}}(\boldsymbol{\beta}_0, \tau)\},
\end{aligned} \tag{2.14}$$

Following Peng and Huang (2008), we can show that $\sup_{\tau \in (0, \tau_U]} \|\boldsymbol{\mu}_1\{\hat{\boldsymbol{\beta}}(\hat{\boldsymbol{\alpha}}, \tau)\} - \boldsymbol{\mu}_1\{\tilde{\boldsymbol{\beta}}(\hat{\boldsymbol{\alpha}}, \tau)\}\|$
$= \boldsymbol{o}_p(1)$ and $\sup_{\tau \in (0, \tau_U]} \|\boldsymbol{\mu}_2\{\hat{\boldsymbol{\alpha}}(\hat{\boldsymbol{\beta}}, \tau)\} - \boldsymbol{\mu}_2\{\tilde{\boldsymbol{\alpha}}(\hat{\boldsymbol{\beta}}, \tau)\}\| = \boldsymbol{o}_p(1)$. Then (2.14) can be rewritten as

$$\begin{aligned}
\boldsymbol{\mu}_1\{\hat{\boldsymbol{\beta}}(\tau)\} - \boldsymbol{\mu}_1\{\boldsymbol{\beta}_0(\tau)\} &= \boldsymbol{\mu}_1\{\tilde{\boldsymbol{\beta}}(\hat{\boldsymbol{\alpha}}, \tau)\} - \boldsymbol{\mu}_1\{\tilde{\boldsymbol{\beta}}(\boldsymbol{\alpha}_0, \tau)\} + \boldsymbol{o}_{(0, \tau_U]}(1), \\
\boldsymbol{\mu}_2\{\hat{\boldsymbol{\alpha}}(\tau)\} - \boldsymbol{\mu}_2\{\boldsymbol{\alpha}_0(\tau)\} &= \boldsymbol{\mu}_2\{\tilde{\boldsymbol{\alpha}}(\hat{\boldsymbol{\beta}}, \tau)\} - \boldsymbol{\mu}_2\{\tilde{\boldsymbol{\alpha}}(\boldsymbol{\beta}_0, \tau)\} + \boldsymbol{o}_{(0, \tau_U]}(1),
\end{aligned} \tag{2.15}$$

where $\boldsymbol{o}_I(1)$ denotes a term that converges to 0 in probability uniformly on the interval I.

Define

$$\boldsymbol{\gamma}(\tau) = \begin{pmatrix} \boldsymbol{\beta}(\tau) \\ \boldsymbol{\alpha}(\tau) \end{pmatrix} \text{ and } \tilde{\boldsymbol{g}}(\boldsymbol{\gamma}, \tau) = \begin{pmatrix} \tilde{\boldsymbol{\beta}}(\boldsymbol{\alpha}, \tau) \\ \tilde{\boldsymbol{\alpha}}(\boldsymbol{\beta}, \tau) \end{pmatrix}.$$

Note that $\tilde{\boldsymbol{g}}(\boldsymbol{\gamma}, \tau)$ can be viewed as a functional of $\boldsymbol{\gamma}$ with parameter τ. We further simplify (2.15) as

$$\boldsymbol{\mu}\{\hat{\boldsymbol{\gamma}}(\tau)\} - \boldsymbol{\mu}\{\boldsymbol{\gamma}_0(\tau)\} = \boldsymbol{\mu}\{\tilde{\boldsymbol{g}}(\hat{\boldsymbol{\gamma}}, \tau)\} - \boldsymbol{\mu}\{\tilde{\boldsymbol{g}}(\boldsymbol{\gamma}_0, \tau)\} + \boldsymbol{o}_{(0,\tau_U]}(1). \tag{2.16}$$

Let $\boldsymbol{\theta}(\tau) = \boldsymbol{\mu}\{\boldsymbol{\gamma}(\tau)\}$ and $\boldsymbol{\rho}(\boldsymbol{\theta}, \tau) = \boldsymbol{\theta}(\tau) - \boldsymbol{\mu}\big(\tilde{\boldsymbol{g}}[\boldsymbol{\kappa}\{\boldsymbol{\theta}(\tau)\}]\big)$. Then (2.16) becomes

$$\boldsymbol{\rho}(\hat{\boldsymbol{\theta}}, \tau) - \boldsymbol{\rho}(\boldsymbol{\theta}_0, \tau) = \boldsymbol{o}_{(0,\tau_U]}(1). \tag{2.17}$$

By viewing the parameter τ as fixed and dropping it from the notation of $\boldsymbol{\rho}(\boldsymbol{\theta}, \tau)$ for brevity, we have

$$\boldsymbol{\rho}(\hat{\boldsymbol{\theta}}) - \boldsymbol{\rho}(\boldsymbol{\theta}_0) = \boldsymbol{o}_{(0,\tau_U]}(1). \tag{2.18}$$

Note that $\boldsymbol{\rho}(\boldsymbol{\theta})$ is a functional of $\boldsymbol{\theta}$. By Condition C5(a), $\boldsymbol{\rho}$ is Gâteaux differentiable at $\boldsymbol{\theta}_0$, that is, for any direction $\boldsymbol{h} \in \mathcal{F}$ and $\boldsymbol{\theta}_0 + t\boldsymbol{h} \in \mathcal{D}$, there is a linear map $\boldsymbol{\rho}'_{\boldsymbol{\theta}_0}$ such that

$$\frac{\boldsymbol{\rho}(\boldsymbol{\theta}_0 + t\boldsymbol{h}) - \boldsymbol{\rho}(\boldsymbol{\theta}_0)}{t} \to \boldsymbol{\rho}'_{\boldsymbol{\theta}_0}(\boldsymbol{h}) \text{ as } t \to 0. \tag{2.19}$$

Let $\boldsymbol{h} = (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)/t$. By (2.19) we have

$$\{\boldsymbol{\rho}(\hat{\boldsymbol{\theta}}) - \boldsymbol{\rho}(\boldsymbol{\theta}_0)\} - t\boldsymbol{\rho}'_{\boldsymbol{\theta}_0}\{(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)/t\} \to 0 \text{ as } t \to 0. \tag{2.20}$$

By (2.18), (2.20), and the linearity of $\boldsymbol{\rho}'_{\boldsymbol{\theta}_0}$, we immediately have

$$\boldsymbol{\rho}'_{\boldsymbol{\theta}_0}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) = \boldsymbol{o}_{(0,\tau_U]}(1). \tag{2.21}$$

Since $\boldsymbol{\theta}$ and $\boldsymbol{\theta}_0$ are bounded on $(0, \tau_U]$, by condition C5(b) and a subsequence argument, (2.21) implies

$$\sup_{u\in(0,\tau_U]} \|\hat{\boldsymbol{\theta}}(u) - \boldsymbol{\theta}_0(u)\| = \boldsymbol{o}_p(1). \tag{2.22}$$

Recall that $\boldsymbol{\kappa}\{\hat{\boldsymbol{\theta}}(u)\} = \hat{\boldsymbol{\gamma}}(u)$ and $\boldsymbol{\kappa}\{\boldsymbol{\theta}_0(u)\} = \boldsymbol{\gamma}_0(u)$. By a Taylor expansion of $\boldsymbol{\kappa}\{\hat{\boldsymbol{\theta}}(\tau)\}$ around $\boldsymbol{\theta}_0(\tau)$ for $\tau \in [\nu, \tau_U]$, together with (2.22) and condition C3(c), we can show

$$\sup_{u\in[\nu,\tau_U]} \|\hat{\boldsymbol{\gamma}}(u) - \boldsymbol{\gamma}_0(u)\| = \boldsymbol{o}_p(1),$$

which implies

$$\begin{aligned} \sup_{u\in[\nu,\tau_U]} \|\hat{\boldsymbol{\beta}}(u) - \boldsymbol{\beta}_0(u)\| &= \boldsymbol{o}_p(1), \\ \sup_{u\in[\nu,\tau_U]} \|\hat{\boldsymbol{\alpha}}(u) - \boldsymbol{\alpha}_0(u)\| &= \boldsymbol{o}_p(1), \end{aligned} \tag{2.23}$$

and thus complete the proof for Theorem 2.1.1.

2.5.3 Proof of Theorem 2.1.2

Having the uniform consistency of $\hat{\boldsymbol{\beta}}(\tau)$ and $\hat{\boldsymbol{\alpha}}(\tau)$ on $\tau \in [\nu, \tau_U]$, by following the proof of Lemma B.1. in Peng and Huang (2008), we can show that

$$\sup_{\tau\in[\nu,\tau_U]} n^{1/2} \|\{\boldsymbol{S}_n^{(k)}(\hat{\boldsymbol{\beta}}, \boldsymbol{\alpha}_0, \tau) - \boldsymbol{S}_n^{(k)}(\boldsymbol{\beta}_0, \boldsymbol{\alpha}_0, \tau)\} - \{\boldsymbol{s}^{(k)}(\hat{\boldsymbol{\beta}}, \boldsymbol{\alpha}_0, \tau) - \boldsymbol{s}^{(k)}(\boldsymbol{\beta}_0, \boldsymbol{\alpha}_0, \tau)\}\| \xrightarrow{p} 0,$$

and

$$\sup_{\tau\in[\nu,\tau_U]} n^{1/2} \|\{\boldsymbol{S}_n^{(k)}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\alpha}}, \tau) - \boldsymbol{S}_n^{(k)}(\hat{\boldsymbol{\beta}}, \boldsymbol{\alpha}_0, \tau)\} - \{\boldsymbol{s}^{(k)}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\alpha}}, \tau) - \boldsymbol{s}^{(k)}(\hat{\boldsymbol{\beta}}, \boldsymbol{\alpha}_0, \tau)\}\| \xrightarrow{p} 0. \tag{2.24}$$

From the above results, we get

$$\begin{aligned}
&\sup_{\tau\in[\nu,\tau_U]} n^{1/2}\|\{\boldsymbol{S}_n^{(k)}(\hat{\boldsymbol{\beta}},\hat{\boldsymbol{\alpha}},\tau)-\boldsymbol{S}_n^{(k)}(\boldsymbol{\beta}_0,\boldsymbol{\alpha}_0,\tau)\}-\{\boldsymbol{s}^{(k)}(\hat{\boldsymbol{\beta}},\hat{\boldsymbol{\alpha}},\tau)-\boldsymbol{s}^{(k)}(\boldsymbol{\beta}_0,\boldsymbol{\alpha}_0,\tau)\}\| \\
&\le \sup_{\tau\in[\nu,\tau_U]} n^{1/2}\|\{\boldsymbol{S}_n^{(k)}(\hat{\boldsymbol{\beta}},\boldsymbol{\alpha}_0,\tau)-\boldsymbol{S}_n^{(k)}(\boldsymbol{\beta}_0,\boldsymbol{\alpha}_0,\tau)\}-\{\boldsymbol{s}^{(k)}(\hat{\boldsymbol{\beta}},\boldsymbol{\alpha}_0,\tau)-\boldsymbol{s}^{(k)}(\boldsymbol{\beta}_0,\boldsymbol{\alpha}_0,\tau)\}\| \\
&+ \sup_{\tau\in[\nu,\tau_U]} n^{1/2}\|\{\boldsymbol{S}_n^{(k)}(\hat{\boldsymbol{\beta}},\hat{\boldsymbol{\alpha}},\tau)-\boldsymbol{S}_n^{(k)}(\hat{\boldsymbol{\beta}},\boldsymbol{\alpha}_0,\tau)\}-\{\boldsymbol{s}^{(k)}(\hat{\boldsymbol{\beta}},\hat{\boldsymbol{\alpha}},\tau)-\boldsymbol{s}^{(k)}(\hat{\boldsymbol{\beta}},\boldsymbol{\alpha}_0,\tau)\}\| \xrightarrow{p} 0.
\end{aligned} \tag{2.25}$$

Simple algebra shows that

$$\begin{aligned}
\boldsymbol{s}^{(1)}(\hat{\boldsymbol{\beta}},\hat{\boldsymbol{\alpha}},\tau)-\boldsymbol{s}^{(1)}(\boldsymbol{\beta}_0,\boldsymbol{\alpha}_0,\tau) = \boldsymbol{\mu}_1(\hat{\boldsymbol{\beta}},\tau)-\boldsymbol{\mu}_1(\boldsymbol{\beta}_0,\tau) - \int_0^{\tau}\Big([\tilde{\boldsymbol{\mu}}_1\{\hat{\boldsymbol{\beta}}(u),\hat{\boldsymbol{\alpha}}\}-\tilde{\boldsymbol{\mu}}_1\{\hat{\boldsymbol{\beta}}(u),\boldsymbol{\alpha}_0\}] \\
+[\tilde{\boldsymbol{\mu}}_1\{\hat{\boldsymbol{\beta}}(u),\boldsymbol{\alpha}_0\}-\tilde{\boldsymbol{\mu}}_1\{\boldsymbol{\beta}_0(u),\boldsymbol{\alpha}_0\}]\Big)\,\mathrm{d}u.
\end{aligned} \tag{2.26}$$

For any $\nu\in(0,\tau_U]$ and any fixed $u\in[\nu,\tau_U]$, given the uniform consistency of $\hat{\boldsymbol{\alpha}}(\cdot)$ and condition C4(b)-(c), we have

$$\begin{aligned}
&\tilde{\boldsymbol{\mu}}_1\{\hat{\boldsymbol{\beta}}(u),\hat{\boldsymbol{\alpha}}\}-\tilde{\boldsymbol{\mu}}_1\{\hat{\boldsymbol{\beta}}(u),\boldsymbol{\alpha}_0\} \\
&= E\Big(\boldsymbol{Z}Y[g\{\boldsymbol{Z}^T\hat{\boldsymbol{\beta}}(u)\}][\phi_1(1-u,1-\int_0^{\tau_U} I\{\boldsymbol{Z}^T\hat{\boldsymbol{\alpha}}(v)\le\boldsymbol{Z}^T\hat{\boldsymbol{\beta}}(u)\}\,\mathrm{d}v) \\
&\qquad -\phi_1(1-u,1-\int_0^{\tau_U} I\{\boldsymbol{Z}^T\boldsymbol{\alpha}_0(v)\le\boldsymbol{Z}^T\hat{\boldsymbol{\beta}}(u)\}\,\mathrm{d}v)]\Big) \\
&\approx E\Big(\boldsymbol{Z}Y[g\{\boldsymbol{Z}^T\hat{\boldsymbol{\beta}}(u)\}]\phi_{12}(1-u,1-\int_0^{\tau_U} I\{\boldsymbol{Z}^T\boldsymbol{\alpha}_0(v)\le\boldsymbol{Z}^T\hat{\boldsymbol{\beta}}(u)\}\,\mathrm{d}v) \\
&\qquad\times\int_0^{\tau_U}[I\{\boldsymbol{Z}^T\boldsymbol{\alpha}_0(v)\le\boldsymbol{Z}^T\hat{\boldsymbol{\beta}}(u)\}-I\{\boldsymbol{Z}^T\hat{\boldsymbol{\alpha}}(v)\le\boldsymbol{Z}^T\hat{\boldsymbol{\beta}}(u)\}]\,\mathrm{d}v\Big) \\
&= \int_0^{\tau_U}[\boldsymbol{w}_1\{\hat{\boldsymbol{\beta}}(u),\boldsymbol{\alpha}_0(v)\}-\boldsymbol{w}_1\{\hat{\boldsymbol{\beta}}(u),\hat{\boldsymbol{\alpha}}(v)\}]\,\mathrm{d}v \\
&\approx \int_0^{\tau_U} -\boldsymbol{w}_{12}\{\boldsymbol{\beta}_0(u),\boldsymbol{\alpha}_0(v)\}\{\hat{\boldsymbol{\alpha}}(v)-\boldsymbol{\alpha}_0(v)\}\,\mathrm{d}v,
\end{aligned} \tag{2.27}$$

where $\approx$ indicates that the difference converges uniformly to 0 for on $[\nu,\tau_U]$.

With Taylor expansion,we can show that

$$\tilde{\boldsymbol{\mu}}_1\{\hat{\boldsymbol{\beta}}(u), \boldsymbol{\alpha}_0\} - \tilde{\boldsymbol{\mu}}_1\{\boldsymbol{\beta}_0(u), \boldsymbol{\alpha}_0\} \approx \tilde{\boldsymbol{J}}_1\{\boldsymbol{\beta}_0(u), \boldsymbol{\alpha}_0\}\{\hat{\boldsymbol{\beta}}(u) - \boldsymbol{\beta}_0(u)\}. \tag{2.28}$$

and we also have

$$\boldsymbol{\mu}_1\{\hat{\boldsymbol{\beta}}(\tau)\} - \boldsymbol{\mu}_1\{\boldsymbol{\beta}_0(\tau)\} \approx \boldsymbol{J}_1\{\boldsymbol{\beta}_0(\tau)\}\{\hat{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}_0(\tau)\}. \tag{2.29}$$

Let $\boldsymbol{\psi}_1(\tau) = \hat{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}_0(\tau)$ and $\boldsymbol{\psi}_2(\tau) = \hat{\boldsymbol{\alpha}}(\tau) - \boldsymbol{\alpha}_0(\tau)$. From (2.26),(2.27),(2.28) and (2.29) we can see that

$$\begin{aligned}\boldsymbol{s}^{(1)}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\alpha}}, \tau) - \boldsymbol{s}^{(1)}(\boldsymbol{\beta}_0, \boldsymbol{\alpha}_0, \tau) \approx \boldsymbol{A}_{01}(\tau)\boldsymbol{\psi}_1(\tau) &- \int_0^{\tau}\int_0^{\tau_U} \boldsymbol{B}_{11}(u, v)\boldsymbol{\psi}_2(v)\,\mathrm{d}v\,\mathrm{d}u \\ &- \int_0^{\tau_U} \boldsymbol{B}_{21}(v)\boldsymbol{\psi}_1(v)\,\mathrm{d}v, \end{aligned} \tag{2.30}$$

where $\boldsymbol{A}_{01}(\tau) = \boldsymbol{J}_1\{\boldsymbol{\beta}_0(\tau)\}$, $\boldsymbol{B}_{11}(u, v) = -\boldsymbol{w}_{12}\{\boldsymbol{\beta}_0(u), \boldsymbol{\alpha}_0(v)\}$ and $\boldsymbol{B}_{21}(v) = \tilde{\boldsymbol{J}}_1\{\boldsymbol{\beta}_0(v), \boldsymbol{\alpha}_0\}$.

Similarly, we can show that

$$\begin{aligned}\boldsymbol{s}^{(2)}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\alpha}}, \tau) - \boldsymbol{s}^{(2)}(\boldsymbol{\beta}_0, \boldsymbol{\alpha}_0, \tau) \approx \boldsymbol{A}_{02}(\tau)\boldsymbol{\psi}_2(\tau) &- \int_0^{\tau}\int_0^{\tau_U} \boldsymbol{B}_{12}(u, v)\boldsymbol{\psi}_1(v)\,\mathrm{d}v\,\mathrm{d}u \\ &- \int_0^{\tau_U} \boldsymbol{B}_{22}(v)\boldsymbol{\psi}_2(v)\,\mathrm{d}v, \end{aligned} \tag{2.31}$$

where $\boldsymbol{A}_{02}(\tau) = \boldsymbol{J}_2\{\boldsymbol{\alpha}_0(\tau)\}$, $\boldsymbol{B}_{12}(u, v) = -\boldsymbol{w}_{21}\{\boldsymbol{\beta}_0(v), \boldsymbol{\alpha}_0(u)\}$ and $\boldsymbol{B}_{22}(v) = \tilde{\boldsymbol{J}}_2\{\boldsymbol{\alpha}_0(v), \boldsymbol{\beta}_0\}$.

Let

$$\boldsymbol{\omega}(\tau) = \begin{pmatrix} \boldsymbol{s}^{(1)}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\alpha}}, \tau) - \boldsymbol{s}^{(1)}(\boldsymbol{\beta}_0, \boldsymbol{\alpha}_0, \tau) \\ \boldsymbol{s}^{(2)}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\alpha}}, \tau) - \boldsymbol{s}^{(2)}(\boldsymbol{\beta}_0, \boldsymbol{\alpha}_0, \tau) \end{pmatrix}, \quad \boldsymbol{\psi}(\tau) = \begin{pmatrix} \boldsymbol{\psi}_1(\tau) \\ \boldsymbol{\psi}_2(\tau) \end{pmatrix},$$

then from (2.30) and (2.31) we have

$$\boldsymbol{\omega}(\tau) \approx \boldsymbol{A}_0(\tau)\boldsymbol{\psi}(\tau) - \int_0^{\tau_U} \boldsymbol{A}_1(\tau, v)\boldsymbol{\psi}(v)\,\mathrm{d}v, \tag{2.32}$$

where

$$\boldsymbol{A}_0(\tau) = \begin{bmatrix} \boldsymbol{A}_{01}(\tau) & 0 \\ 0 & \boldsymbol{A}_{02}(\tau) \end{bmatrix} \text{ and } \boldsymbol{A}_1(\tau, v) = \begin{bmatrix} \boldsymbol{B}_{21}(v) & \int_0^{\tau} \boldsymbol{B}_{11}(u, v)\,\mathrm{d}u \\ \int_0^{\tau} \boldsymbol{B}_{12}(u, v)\,\mathrm{d}u & \boldsymbol{B}_{22}(v) \end{bmatrix},$$

Let

$$\boldsymbol{S}_n(\boldsymbol{b}, \boldsymbol{a}, \tau) = \begin{pmatrix} \boldsymbol{S}_n^{(1)}(\boldsymbol{b}, \boldsymbol{a}, \tau) \\ \boldsymbol{S}_n^{(2)}(\boldsymbol{b}, \boldsymbol{a}, \tau) \end{pmatrix},$$

from (2.25) and (2.32) we have

$$\begin{aligned} -n^{1/2}\boldsymbol{S}_n(\boldsymbol{\beta}_0, \boldsymbol{\alpha}_0, \tau) = {} & \{\boldsymbol{A}_0(\tau) + \boldsymbol{o}_{[\nu,\tau_U]}(1)\} \times n^{1/2}\boldsymbol{\psi}(\tau) \\ & - \int_0^{\tau_U} \{\boldsymbol{A}_1(\tau, v) + \boldsymbol{o}_{[\nu,\tau_U]}(1)\} \times n^{1/2}\boldsymbol{\psi}(v)\,\mathrm{d}v. \end{aligned} \tag{2.33}$$

Equation (2.33) can be viewed as a stochastic differential equation for $n^{1/2}\boldsymbol{\psi}(\tau)$. Specifically, it is a Fredholm integral equation of the second kind and the solution can be presented in the following form (Polyanin and Manzhirov, 2008):

$$n^{1/2}\boldsymbol{\psi}(\tau) = -\boldsymbol{A}_0(\tau)^{-1}\{n^{1/2}\boldsymbol{S}_n(\boldsymbol{\beta}_0, \boldsymbol{\alpha}_0, \tau) - \int_0^{\tau_U} \boldsymbol{R}(\tau, v) \times n^{1/2}\boldsymbol{S}_n(\boldsymbol{\beta}_0, \boldsymbol{\alpha}_0, v)\,\mathrm{d}v\} + \boldsymbol{o}_{[\nu,\tau_U]}(1), \tag{2.34}$$

where $\boldsymbol{R}(\tau, v)$ is determined by $\boldsymbol{A}_1(\tau, v)$ and $\boldsymbol{A}_0(\tau)$, and independent of $\boldsymbol{\psi}(\tau)$. The detailed solution can be found in Polyanin and Manzhirov (2008) and thus omitted here.

By observing (2.34) we can see that, to show the weak convergence of $n^{1/2}\{\hat{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}_0(\tau)\}$ and $n^{1/2}\{\hat{\boldsymbol{\alpha}}(\tau) - \boldsymbol{\alpha}_0(\tau)\}$, it suffices to show the weak convergence of $-n^{1/2}\boldsymbol{S}_n(\boldsymbol{\beta}_0, \boldsymbol{\alpha}_0, \tau)$.

Let

$$\boldsymbol{N}_i(\boldsymbol{\beta},\boldsymbol{\alpha},\tau)=\begin{pmatrix} N_{1i}[g\{\boldsymbol{Z}_i^T\boldsymbol{\beta}(\tau)\}] \\ N_{2i}[g\{\boldsymbol{Z}_i^T\boldsymbol{\alpha}(\tau)\}] \end{pmatrix}$$

and

$$\boldsymbol{K}_i(\boldsymbol{\beta},\boldsymbol{\alpha},\tau)=\begin{pmatrix} \int_0^\tau Y_i[g\{\boldsymbol{Z}_i^T\boldsymbol{\beta}(u)\}]\times\phi_1(1-u,1-\int_0^{\tau_U} I\{\boldsymbol{Z}_i^T\boldsymbol{\alpha}(v)\le\boldsymbol{Z}_i^T\boldsymbol{\beta}(u)\}\,\mathrm{d}v)\,\mathrm{d}u \\ \int_0^\tau Y_i[g\{\boldsymbol{Z}_i^T\boldsymbol{\alpha}(u)\}]\times\phi_2(1-\int_0^{\tau_U} I\{\boldsymbol{Z}_i^T\boldsymbol{\beta}(v)\le\boldsymbol{Z}_i^T\boldsymbol{\alpha}(u)\}\,\mathrm{d}v,1-u)\,\mathrm{d}u \end{pmatrix}.$$

First we note that $\{\boldsymbol{Z}_i\boldsymbol{N}_i(\boldsymbol{\beta}_0,\boldsymbol{\alpha}_0,\tau),\tau\in(0,\tau_U]\}$ is a VC-class (van der Vaart and Wellner 1996), and $\boldsymbol{K}_i(\boldsymbol{\beta}_0,\boldsymbol{\alpha}_0,\tau)$ is Lipschitz in τ. It then follows that $\Big\{\boldsymbol{Z}_i\{\boldsymbol{N}_i(\boldsymbol{\beta}_0,\boldsymbol{\alpha}_0,\tau)-\boldsymbol{K}_i(\boldsymbol{\beta}_0,\boldsymbol{\alpha}_0,\tau)\},\tau\in(0,\tau_U]\Big\}$ is a Donsker class by the permanence properties. By the Donsker theorem, $-n^{1/2}\boldsymbol{S}_n(\boldsymbol{\beta}_0,\boldsymbol{\alpha}_0,\tau)$ converges weakly to a Gaussian process, namely $\boldsymbol{G}(\tau)$, with mean 0 and covariance $\boldsymbol{\Sigma}(s,t)$ for $s,t\in(0,\tau_U]$, where $\boldsymbol{\Sigma}(s,t)=E\{\boldsymbol{\iota}_j(s)\boldsymbol{\iota}_j(t)^T\}$ with $\boldsymbol{\iota}_j(\tau)=\boldsymbol{Z}_j\{\boldsymbol{N}_j(\boldsymbol{\beta}_0,\boldsymbol{\alpha}_0,\tau)-\boldsymbol{K}_j(\boldsymbol{\beta}_0,\boldsymbol{\alpha}_0,\tau)\},\tau\in(0,\tau_U]$. By this fact, coupled with (2.34) and condition C3(a), we can see that $n^{1/2}\{\hat{\boldsymbol{\beta}}(\tau)-\boldsymbol{\beta}_0(\tau)\}$ converges weakly to a Guassian process for $\tau\in[\nu_1,\tau_U]$, and $n^{1/2}\{\hat{\boldsymbol{\alpha}}(\tau)-\boldsymbol{\alpha}_0(\tau)\}$ also converges weakly to a Guassian process for $\tau\in[\nu_2,\tau_U]$, where $0<\nu_1,\nu_2<\tau_U$.

2.6 Convergence Criteria

2.6.1 Convergence Criteria for Computing Algorithms

In this section we provide detailed convergence criteria which is shared by the two algorithms A and B proposed in Section 2.1. First, we set the maximum number of iterations, denoted by M_0, and the tolerance level, denoted by *tol*. At the m–th iteration ($m \geq 1$), define $d_{\beta,q}^{[m]} = \frac{1}{\tau_{U,1}-\nu_1}\int_{\nu_1}^{\tau_{U,1}}[\beta^{[m+1]^{(q)}}(\tau) - \beta^{[m]^{(q)}}(\tau)]\,\mathrm{d}\tau$ and $\tilde{d}_{\beta,q}^{[m]} = \frac{1}{\tau_{U,1}-\nu_1}\int_{\nu_1}^{\tau_{U,1}}[\beta^{[m+1]^{(q)}}(\tau) - \beta^{[m-1]^{(q)}}(\tau)]\,\mathrm{d}\tau$ for $q = 0, \cdots, p$, recalling that $\beta^{(0)}(\tau)$ is the intercept and $\beta^{(q)}(\tau)$ is the coefficient corresponding to the q–th element of $\tilde{\boldsymbol{Z}}$ for $q = 1, \cdots, p$. Let $d_{\beta}^{[m]} = \max_q\{abs(d_{\beta,q}^{[m]})\}$ and $\tilde{d}_{\beta}^{[m]} = \max_q\{abs(\tilde{d}_{\beta,q}^{[m]})\}$, where $abs(\cdot)$ stands for the absolute value function. We also define $d_{\alpha}^{[m]}$ and $\tilde{d}_{\alpha}^{[m]}$ in the similar fashion. At the end of Step A2 (B2) of the m–th iteration, we carry out the following steps:

Step 0. Compare m with M_0. If $m < M_0$ then continue to the next step, otherwise stop and claim non-convergence.

Step 1. If $\max\{d_{\beta}^{[m]}, d_{\alpha}^{[m]}\} < tol$, then announce convergence. Let $\hat{\boldsymbol{\beta}}(\tau) = \hat{\boldsymbol{\beta}}^{[m+1]}(\tau)$ for $\tau \in (0, \tau_{U,1}]$ and $\hat{\boldsymbol{\alpha}}(\tau) = \hat{\boldsymbol{\alpha}}^{[m+1]}(\tau)$ for $\tau \in (0, \tau_{U,2}]$ and stop. Otherwise continue to the next step.

Step 2. If $\max\{\tilde{d}_{\beta}^{[m]}, \tilde{d}_{\alpha}^{[m]}\} < tol$, then announce convergence. Let $\hat{\boldsymbol{\beta}}(\tau) = \{\hat{\boldsymbol{\beta}}^{[m+1]}(\tau) + \hat{\boldsymbol{\beta}}^{[m-1]}(\tau)\}/2$ for $\tau \in (0, \tau_{U,1}]$ and $\hat{\boldsymbol{\alpha}}(\tau) = \{\hat{\boldsymbol{\alpha}}^{[m+1]}(\tau) + \hat{\boldsymbol{\alpha}}^{[m-1]}(\tau)\}/2$ for $\tau \in (0, \tau_{U,2}]$ and stop. Otherwise continue to Step A3 (B3).

In the simulation studies and data analysis reported in Sections 2.2 and 2.3, we chose $M_0 = 10$ and $tol = 10^{-2}$.

Chapter 3

Quantile Regression for Doubly Censored Data with Known Left Censoring Times

3.1 Quantile Regression Procedures

3.1.1 Data and Model

Let T denote the event time of interest, L be the left censoring time, U be the right censoring time, and $\tilde{\boldsymbol{Z}}$ be the $p\times 1$ vector of recorded covariates. Define $\boldsymbol{Z} = (1, \tilde{\boldsymbol{Z}}^T)^T$ and $X = \max\{L, \min(T, U)\}$. The censoring indicator δ is defined as

$$\delta = \begin{cases} 1, & \text{if } L < T \leq U; \\ 2, & \text{if } T \leq L; \\ 3, & \text{if } T > U. \end{cases}$$

We assume that $(L, U) \perp T$ given $\boldsymbol{Z}$. The observed data consists of n i.i.d. replicates of $(X, \delta, \boldsymbol{Z}, L)$, denoted by $\{(X_i, \delta_i, \boldsymbol{Z}_i, L_i)\}_{i=1}^n$.

Define the conditional τ–th quantile of T given $\boldsymbol{Z}$ by $Q_T(\tau|\boldsymbol{Z}) = \inf\{t : F_T(t|\boldsymbol{Z}) \geq \tau\}$, where $F_T(t|\boldsymbol{Z}) = \Pr(T \leq t|\boldsymbol{Z})$. We consider the quantile regression model taking the form,

$$Q_T(\tau|\boldsymbol{Z}) = g\{\boldsymbol{Z}^T\boldsymbol{\beta_0}(\tau)\}, \quad \tau \in (0, 1), \tag{3.1}$$

where $g(\cdot)$ is a known monotone link function, and $\boldsymbol{\beta}_0(\tau)$ is a vector of unknown coefficients representing covariate effects on $Q_T(\tau|\boldsymbol{Z})$.

3.1.2 Estimation Procedure

To estimate $\boldsymbol{\beta_0}(\tau)$ in model (3.1), our basic idea is to determine an appropriate martingale process which allows us to construct an unbiased stochastic integral estimating equation by using Peng and Huang (2008)'s technique.

Following this line, we consider $M(t) = N(t) - \int_0^t R(u)\,\mathrm{d}\Lambda_T(u|\boldsymbol{Z})$, where $N(t) = I(X \leq t, \delta = 1)$, $R(t) = I(L < t \leq X)$, and $\Lambda_T(t|\boldsymbol{Z})$ denotes the cumulative hazard function of T given $\boldsymbol{Z}$. Let $N_i(t)$, $R_i(t)$, $\Lambda_T(t|\boldsymbol{Z}_i)$, and $M_i(t)$ be sample analogues

of $N(t)$, $R(t)$, $\Lambda_T(t|\boldsymbol{Z})$, and $M(t)$. Denote the filtration $\sigma\{N_i(u), R_i(u+), \boldsymbol{Z}_i : i = 1, \cdots, n; 0 \le u \le t\}$ by $\mathscr{F}_t$. Provided $(L_i, U_i) \perp T_i$ given $\boldsymbol{Z}_i$,

$$\begin{aligned} E[dN_i(t)|\mathscr{F}_{t-}] &= Pr[t \le T_i < t + dt, R_i(t) = 1|\mathscr{F}_{t-}] \\ &= R_i(t)Pr[t \le T_i < t + dt \mid T_i \ge t, L_i < t, U_i \ge t, \boldsymbol{Z}_i] = R_i(t)d\Lambda_T(t|\boldsymbol{Z}_i). \end{aligned}$$

This shows that $M_i(t)$ is a martingale (Fleming and Harrington 1991) and

$$E\{M_i(t)|\boldsymbol{Z}_i\} = 0, \forall t \ge 0. \tag{3.2}$$

Under model (3.1), which implies $\Lambda_T(g\{\boldsymbol{Z}_i^T\boldsymbol{\beta}_0(\tau)\}|\boldsymbol{Z}_i) = -\log(1-\tau)$, it follows that

$$M_i[g\{\boldsymbol{Z}_i^T\boldsymbol{\beta}_0(\tau)\}] = N_i[g\{\boldsymbol{Z}_i^T\boldsymbol{\beta}_0(\tau)\}] - \int_0^\tau I[L_i < g\{\boldsymbol{Z}_i^T\boldsymbol{\beta}_0(v)\} \le X_i]\,\mathrm{d}H(v)$$

from a use of variable transformation within the integral, where $H(x) = -\log(1-x)$. This equality and (3.2) naturally lead to an estimating equation for $\boldsymbol{\beta}_0(\cdot)$ given by

$$n^{1/2}\boldsymbol{S}_n(\boldsymbol{\beta}, \tau) = 0, \tag{3.3}$$

where

$$\boldsymbol{S}_n(\boldsymbol{\beta}, \tau) = \frac{1}{n}\sum_{i=1}^n \boldsymbol{Z}_i\bigg(N_i[g\{\boldsymbol{Z}^T\boldsymbol{\beta}(\tau)\}] - \int_0^\tau I[L_i < g\{\boldsymbol{Z}^T\boldsymbol{\beta}(v)\} \le X_i]\,\mathrm{d}H(v)\bigg).$$

It is easy to see that $E\{\boldsymbol{S}_n(\boldsymbol{\beta}_0, \tau)\} = 0$. With all $L_i = 0$, equation (3.3) reduces to Peng and Huang (2008)'s estimating equation for randomly right censored data.

The stochastic integral representation of $S_n(\boldsymbol{\beta}, \tau)$ suggests a grid-based procedure to obtain an estimator of $\boldsymbol{\beta}_0(\cdot)$ based on equation (3.3). Specifically, define the estimator $\hat{\boldsymbol{\beta}}(\tau)$ as a right-continuous step function that jumps only on a prespecified grid, $\mathscr{G}_{L_n} = \{0 = \tau_0 < \tau_1 < \cdots < \tau_{L_n} = \tau_U < 1\}$, where τ_U is a prespecified constant

subject to certain theoretical constraint. Because the definition of $Q_T(\cdot|\boldsymbol{Z}_0)$ and model (3.1) imply $g\{\boldsymbol{Z}^T\boldsymbol{\beta}_0(0)\} = 0$, we always set $g\{\boldsymbol{Z}^T\hat{\boldsymbol{\beta}}(0)\} = 0$. We propose to obtain $\hat{\boldsymbol{\beta}}(\tau_j)$ $(j = 1, \cdots, L_n)$ by sequentially solving the following equation for $\boldsymbol{\beta}(\tau_j)$:

$$n^{-\frac{1}{2}} \sum_{i=1}^{n} \boldsymbol{Z}_i (N_i[g\{\boldsymbol{Z}^T\boldsymbol{\beta}(\tau_j)\}] - \sum_{k=1}^{j-1} I[L_i < g\{\boldsymbol{Z}^T\hat{\boldsymbol{\beta}}(\tau_k)\} \leq X_i] \times \{H(\tau_{k+1}) - H(\tau_k)\}) = 0. \tag{3.4}$$

Note that equation (3.4) is a monotone estimating equation (Fygenson and Ritov, 1994), and the left hand side of (3.4) equals $2n^{-\frac{1}{2}}$ times the gradient of the following L_1–type convex function,

$$\begin{aligned} l_j(\boldsymbol{h}) = & \sum_{i=1}^{n} |I(\delta_i = 1)g^{-1}(X_i) - \boldsymbol{h}^T I(\delta_i = 1)\boldsymbol{Z}_i| + |R^* - \boldsymbol{h}^T \sum_{l=1}^{n} \{-I(\delta_l = 1)\boldsymbol{Z}_l\}| \\ & + |R^* - \boldsymbol{h}^T \sum_{r=1}^{n} (2\boldsymbol{Z}_r \sum_{k=1}^{j-1} I[L_r < g\{\boldsymbol{Z}_r^T\hat{\boldsymbol{\beta}}(\tau_k)\} \leq X_r] \times \{H(\tau_{k+1}) - H(\tau_k)\})|, \end{aligned} \tag{3.5}$$

where R^* is a very large number. As a result, $\hat{\boldsymbol{\beta}}(\tau_j)$ can be alternatively obtained as the minimizer of $l_j(\boldsymbol{h})$. This L_1–minimization problem can be readily solved, for example, by using the Barrodale-Roberts algorithm (Barrodale and Roberts, 1974) implemented in standard statistical software, such as $l1fit()$ function in S-PLUS and $rq()$ function in R.

3.1.3 Asymptotic Results

Asymptotic studies of the proposed estimator are facilitated by the stochastic integral representation of our estimating function. We introduce necessary notation before stating the regularity conditions and theorems.

Let $F_X(\cdot|\boldsymbol{Z})$ and $\bar{F}_X(\cdot|\boldsymbol{Z})$ be the distribution function and survival function of

X given $\boldsymbol{Z}$, respectively. Define $\tilde{F}_{X,\delta}(t|\boldsymbol{Z}) = \Pr(X \le t, \delta = 1|\boldsymbol{Z})$ and $\bar{F}_{X,L}(t|\boldsymbol{Z}) = \Pr(X \ge t, L \ge t|\boldsymbol{Z})$. Let $f_T(\cdot|\boldsymbol{Z})$, $f_X(\cdot|\boldsymbol{Z})$, $\bar{f}_X(\cdot|\boldsymbol{Z})$, $\tilde{f}_{X,\delta}(\cdot|\boldsymbol{Z})$, $\bar{f}_{X,L}(\cdot|\boldsymbol{Z})$ and $g'(\cdot)$ denote the first order derivatives of $F_T(\cdot|\boldsymbol{Z})$, $F_X(\cdot|\boldsymbol{Z})$, $\bar{F}_X(\cdot|\boldsymbol{Z})$, $\tilde{F}_{X,\delta}(\cdot|\boldsymbol{Z})$, $\bar{F}_{X,L}(\cdot|\boldsymbol{Z})$ and $g(\cdot)$, respectively. Let $\|\mathscr{G}_{L_n}\| = \max\{|\tau_j - \tau_{j-1}|,\ j = 1, \ldots, L_n\}$.

Define $\boldsymbol{\mu}(\boldsymbol{b}) = E[\boldsymbol{Z}N\{g(\boldsymbol{Z}^T\boldsymbol{b})\}]$, $\boldsymbol{B}(\boldsymbol{b}) = E[\boldsymbol{Z}^{\otimes 2}\tilde{f}_{X,\delta}\{g(\boldsymbol{Z}^T\boldsymbol{b})|\boldsymbol{Z}\}g'(\boldsymbol{Z}^T\boldsymbol{b})]$, $\boldsymbol{v}_n(\boldsymbol{b}) = n^{-1}\sum_{i=1}^n \boldsymbol{Z}_i N_i\{g(\boldsymbol{Z}_i^T\boldsymbol{b})\} - \boldsymbol{\mu}(\boldsymbol{b})$, $\tilde{\boldsymbol{\mu}}(\boldsymbol{b}) = E[\boldsymbol{Z}I\{L < g(\boldsymbol{Z}^T\boldsymbol{b}) \le X\}]$, $\boldsymbol{J}(\boldsymbol{b}) = E[\boldsymbol{Z}^{\otimes 2}(\bar{f}_X\{g(\boldsymbol{Z}^T\boldsymbol{b})|\boldsymbol{Z}\} - \bar{f}_{X,L}\{g(\boldsymbol{Z}^T\boldsymbol{b})|\boldsymbol{Z}\})g'(\boldsymbol{Z}^T\boldsymbol{b})]$, and $\tilde{\boldsymbol{v}}_n(\boldsymbol{b}) = n^{-1}\sum_{i=1}^n \boldsymbol{Z}_i I\{L_i < g(\boldsymbol{Z}_i^T\boldsymbol{b}) \le X_i\} - \tilde{\boldsymbol{\mu}}(\boldsymbol{b})$.

The regularity conditions include:

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

C2. (a) Each component of $E(\boldsymbol{Z}N[g\{\boldsymbol{Z}^T\boldsymbol{\beta}_0(\tau)\}])$ is a Lipschitz function of τ; (b) $\tilde{f}_{X,\delta}(t|\boldsymbol{Z})$ and $f_X(t|\boldsymbol{z})$ are bounded above uniformly in t and $\boldsymbol{z}$.

C3. (a) $\tilde{f}_{X,\delta}\{g(\boldsymbol{Z}^T\boldsymbol{b})|\boldsymbol{Z}\} > 0$ for all $\boldsymbol{b} \in \mathscr{B}(d_0)$; (b) $E(\boldsymbol{Z}^{\otimes 2}) > 0$; (c) each component of $\boldsymbol{J}(\boldsymbol{b})\boldsymbol{B}(\boldsymbol{b})^{-1}$ is uniformly bounded in $\boldsymbol{b} \in \mathscr{B}(d_0)$, where $\mathscr{B}(d_0)$ is a neighborhood containing $\{\boldsymbol{\beta}_0(\tau), \tau \in (0, \tau_U)\}$, defined as $\mathscr{B}(d_0) = \{\boldsymbol{b} \in \mathbb{R}^p : \inf_{\tau\in(0,\tau_U]}\|\boldsymbol{\mu}(\boldsymbol{b}) - \boldsymbol{\mu}(\boldsymbol{\beta}_0(\tau))\| \le d_0\}$.

C4. $\inf_{\tau\in[\nu,\tau_U]}\text{eigmin}\boldsymbol{B}(\boldsymbol{\beta}_0(\tau)) > 0$ for any $\nu \in (0, \tau_U)$, where eigmin$(\cdot)$ denotes the minimal eigenvalue of a matrix.

We establish the uniform consistency and weak convergence of $\hat{\boldsymbol{\beta}}(\tau)$ stated in the following theorems.

Theorem 3.1.1. *Assuming conditions C1-C4 hold and* $\lim_{n\to\infty}\|\mathscr{G}_{L_n}\| = 0$, *then* $sup_{\tau\in[\nu,\tau_U]}\|\hat{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}_0(\tau)\| \to_p 0$, *where* $0 < \nu < \tau_U$.

Theorem 3.1.2. *Assuming conditions C1-C4 hold and* $\lim_{n\to\infty} n^{1/2}\|\mathscr{G}_{L_n}\| = 0$, *then* $n^{1/2}\{\hat{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}_0(\tau)\}$ *converges weakly to a Gaussian process for* $\tau \in [\nu, \tau_U]$, *where* $0 < \nu < \tau_U$.

The regularity conditions and the proofs of Theorems 3.1.1 and 3.1.2 bear similarity with those in Peng and Huang (2008). Specifically, providing $E\{\boldsymbol{S}_n(\boldsymbol{\beta}_0, \tau)\} = 0$

and $\boldsymbol{\mu}\{\hat{\boldsymbol{\beta}}(0)\} = 0$ shown in Section 3.1.2, we can establish the uniform consistency of the proposed estimator by straightforwardly adapting the arguments in Peng and Huang (2008, Web Appendices). The proof of Theorem 3.1.1 is thus omitted. We sketch the proof of Theorem 3.1.2 in Section 3.7, which provides more concrete information on the asymptotic distribution of $n^{1/2}\{\hat{\boldsymbol{\beta}}(\cdot) - \boldsymbol{\beta}_0(\cdot)\}$.

We would also like to remark that the regularity condition C4 is the crucial constraint that ensures the identifiability of $\{\boldsymbol{\beta}_0(\tau), \tau \in (0, \tau_U]\}$. In the simple one-sample case, this condition is equivalent to

$$f_T[g\{\boldsymbol{\beta}_0(\tau)\}] \cdot \Pr[L < g\{\boldsymbol{\beta}_0(\tau)\} \leq U] > 0, \forall \tau \in (\nu, \tau_U],$$

for any $\nu \in (0, \tau_U)$. Assuming $f_T(\cdot)$ is bounded away from 0, this condition further reduces to $\Pr\{L < Q_T(\tau) \leq U\} > 0, \forall \tau \in (\nu, \tau_U]$, implying $\tau_U \leq F_T(U^+)$ and $\nu \geq F_T(L^-)$. Here and hereafter, for a random variable Y, we use Y^+ and Y^- to denote the upper bound and the lower bound of its support respectively. This requirement concurs with Chang and Yang (1987)'s identifiability condition proposed for estimating the distribution function of doubly censored data in the one-sample case. Since $\nu > 0$ can be chosen arbitrarily, it implies $L^- \leq T^-$. In the general regression setting, however, C4 only renders implicit conditions on L and τ_U to guarantee the identifiability of $\boldsymbol{\beta}_0(\cdot)$.

3.1.4 Inferences

For inferences on $\boldsymbol{\beta}_0(\cdot)$, we propose resampling-based approaches, given the complexity in the asymptotic distribution of $\hat{\boldsymbol{\beta}}(\cdot)$ shown in the proof of Theorem 3.1.2.

More specifically, to estimate the variance of $\hat{\boldsymbol{\beta}}(\tau)$, we perturb the objective function (3.5) by $\xi_1, \cdots, \xi_n$, a set of i.i.d. variates from a nonnegative known distribution

with mean 1 and variance 1, for example, $Exp(1)$. The resulting objective function is

$$\tilde{l}_j(\boldsymbol{h}) = \sum_{i=1}^{n} |\xi_i I(\delta_i = 1) g^{-1}(X_i) - \boldsymbol{h}^T \xi_i I(\delta_i = 1) \boldsymbol{Z}_i| + |R^* - \boldsymbol{h}^T \sum_{l=1}^{n} \{-\xi_l I(\delta_l = 1) \boldsymbol{Z}_l\}|$$
$$+|R^* - \boldsymbol{h}^T \sum_{r=1}^{n} (2\xi_r \boldsymbol{Z}_r \sum_{k=1}^{j-1} I[L_r \le g\{\boldsymbol{Z}_r^T \boldsymbol{\beta}^*(\tau_k)\} \le X_r] \times \{H(\tau_{k+1}) - H(\tau_k)\}) \quad (3.6)$$

for $j = 1, \cdots, L_n$, where $\boldsymbol{\beta}^*(\tau_j)$ is defined as the minimizer of $\tilde{l}_j(\boldsymbol{h})$ and can be obtained sequentially using the same procedure as that taken to compute $\hat{\boldsymbol{\beta}}(\tau_j)$. For a fixed τ_*, we can approximate the variance of $\hat{\boldsymbol{\beta}}(\tau_*)$ by repeatedly generating the variates set $\{\xi_1, \cdots, \xi_n\}$ for B times and obtaining the corresponding $\{\boldsymbol{\beta}_k^*(\tau_*)\}_{k=1}^B$. The confidence interval for $\boldsymbol{\beta}_0(\tau_*)$ can be constructed using $\{\boldsymbol{\beta}_k^*(\tau_*)\}_{k=1}^B$ based on a normal approximation.

Second-stage inferences can also be conducted. First, we consider the general hypothesis $H_0 : \psi\{\boldsymbol{\beta_0}(\tau)\} = r_0(\tau), \tau \in [l, u]$, where $\psi(\cdot)$ is a known function and $r_0(\tau)$ is a hypothesized value for $\psi\{\boldsymbol{\beta_0}(\tau)\}$. Let $\psi(\mathbf{x}) = \mathbf{x^{(q)}}$ and $r_0(\tau) = 0$, where $u^{(l)}$ denotes the l–th component of vector $\mathbf{u}$ and $2 \le q \le p+1$. Testing H_0 is equivalent to assess whether the effect of $Z^{(q)}$ is significant for $\tau \in [l, u]$. One natural test may take the form, $\Gamma = n^{1/2} \int_l^u \{\psi\{\hat{\boldsymbol{\beta}}(v)\} - r_0(v)\}\Theta(v)\, \mathrm{d}v$, where $\Theta(\cdot)$ is a nonnegative weight function. It can be shown that the limit distribution of Γ under H_0 is a mean zero normal distribution, which may be approximated using the empirical distribution of $\Gamma^* = n^{1/2} \int_l^u [\psi\{\boldsymbol{\beta}^*(v)\} - \psi\{\hat{\boldsymbol{\beta}}(v)\}]\Theta(v)\, \mathrm{d}v$ given the observed data.

Another second-stage hypothesis of interest is given by $\tilde{H}_0 : \tilde{\psi}\{\boldsymbol{\beta}(\tau)\} = \eta_0, \tau \in [l, u]$, where $\tilde{\psi}(\cdot)$ is a known function and η_0 is an unspecified constant. With $\tilde{\psi}(\mathbf{x}) = \mathbf{x^{(q)}}$, $\tilde{H}_0$ depicts the scenario where the effect of $\boldsymbol{Z}^{(q)}$ is constant over $\tau \in [l, u]$. To test $\tilde{H}_0$, one may adopt the test statistic $\tilde{\Gamma} = n^{1/2} \int_l^u \{\tilde{\psi}\{\hat{\boldsymbol{\beta}}(v)\} - \hat{\rho}\}\tilde{\Theta}(v)\, \mathrm{d}v$, where $\tilde{\Theta}(\cdot)$ is a nonconstant weight function and $\hat{\rho} = \int_l^u \tilde{\psi}\{\hat{\boldsymbol{\beta}}(v)\}\, \mathrm{d}v/(u-l)$. We can show that the distribution of $\tilde{\Gamma}$ under $\tilde{H}_0$ is equivalent to the conditional distribution of $\tilde{\Gamma}^* = n^{1/2} \int_l^u ([\tilde{\psi}\{\boldsymbol{\beta}^*(v)\} - \tilde{\psi}\{\hat{\boldsymbol{\beta}}(v)\}] - (\rho^* - \hat{\rho}))\tilde{\Theta}(v)\, \mathrm{d}v$ given the observed data, where

ρ^* is $\hat{\rho}$ with $\hat{\boldsymbol{\beta}}(\cdot)$ replaced by $\boldsymbol{\beta}^*(\cdot)$.Therefore, we may reject $\tilde{H}_0$ when $\tilde{\Gamma}$ is greater than the $(1-\alpha/2)$th quantile or less than the $(\alpha/2)$th empirical quantile of $\tilde{\Gamma}^*$.

A useful by-product of the hypothesis testing for $\tilde{H}_0$ is $\hat{\rho}$ in the special case with $\tilde{\psi}(\mathbf{x}) = \mathbf{x}^{(\mathbf{q})}$ and $\tilde{\Theta}(v) = 1$, denoted by $\hat{\eta}$. Let $\eta_0 = \int_l^u \boldsymbol{\beta}_0^{(q)}(v)dv$. This quantity may be interpreted as the average quantile effect of $\boldsymbol{Z}^{(q)}$ across $\tau \in [l, u]$. We can show that $\hat{\eta}$ may be a consistent estimator of η_0. The asymptotic distribution of $n^{1/2}(\hat{\eta}-\eta_0)$ may be approximated by the empirical distribution of $\eta^* = \int_l^u \{\boldsymbol{\beta}^{*(q)}(v) - \hat{\boldsymbol{\beta}}^{(q)}(v)\}dv$ given the observed data. Inference on η_0 may be made accordingly.

3.2 A Conditional Version of Quantile Regression

As pointed out in Section 3.1.3, certain conditions are required for the identifiability of $\boldsymbol{\beta}_0(\tau)$ and may not be satisfied in some real datasets, for example, when $L^- > T^-$. It is of practical interest to propose some remedies when data fail to identify some part of $\boldsymbol{\beta}_0(\cdot)$, most likely to be $\boldsymbol{\beta}_0(\tau)$ with small or large τ. In the presence of only random right censoring, one practical solution is to adaptively imposing an upper bound, τ_U, on the τ-range in which $\boldsymbol{\beta}_0(\tau)$ is estimated (Peng and Huang, 2008). This action has little impact on the estimation because the estimating equations for $\{\boldsymbol{\beta}_0(\tau),\ \ 0 < \tau < \tau_U\}$ stand alone without involving estimates for $\boldsymbol{\beta}_0(\tau)$ with $\tau > \tau_U$.

Dealing with the identifiability issue in the double censoring case is more challenging because non-identifiability can occur on both tails of regression quantiles. One perceivable difficulty in view of equation (3.3) is that the sequential procedure presented in Section 3.1.2 would fail to estimate $\{\boldsymbol{\beta}_0(\tau) :\ \tau \in [\tau_L, \tau_U]\}$, which, say, are identifiable with available data, without any good estimate for $\boldsymbol{\beta}_0(\tau)$ with $\tau < \tau_L$, which may be not be identifiable.

When the identifiability of the lower tail of $\boldsymbol{\beta}_0(\cdot)$ is precluded by doubly censored

data, we propose a conditional version of model (3.1), which takes the form,

$$Q_T(\tau|\boldsymbol{Z}, T > t_0) = g\{\boldsymbol{Z}^T\boldsymbol{\alpha_0}(\tau)\} + t_0, \quad \tau \in (0,1), \tag{3.7}$$

where $Q_T(\tau|\boldsymbol{Z}, T > t_0) = \inf\{t \geq t_0 : \Pr(T \leq t|\boldsymbol{Z}, T > t_0) \geq \tau\}$, and $t_0 > 0$ is a prespecified constant subject to certain theoretical and practical constraints. A more detailed discussion on how to choose t_0 is relegated to the end of this section. In model (3.7), the unknown coefficients in $\boldsymbol{\alpha}_0(\tau)$ represent the effects of covariates on the τ-th conditional quantile of T provided $T > t_0$.

The reasoning for adopting this conditional version of quantile regression is similar to that for estimating a conditional survival function of left truncated and right censored data when the unconditional one is not identifiable (Tsai et al., 1987). Model (3.7) essentially imposes a lower bound of t_0 for regression quantiles. As elaborated later, doing so helps circumvent the difficulty associated with non-identifiable lower tail of $\boldsymbol{\beta}_0(\tau)$ at the cost of estimating a quantity which may slightly deviate from the primary interest.

Model (3.7) necessitates a different estimation procedure from that of the unconditional model (3.1). Estimating equation (3.3) can not be directly borrowed without modification. The critical step in the adaptation of equation (3.3) is to identify an analogue of $M(t)$ when model (3.7) is assumed instead of model (3.1). Along this line, we propose a natural substitute of $M(t)$, given by $\check{M}(t) = \check{N}(t) - \int_0^t \check{R}(u)\, \mathrm{d}\Lambda_T(u|\boldsymbol{Z}, T > t_0)$, where $\check{N}(t) = I(t_0 < X \leq t, \delta = 1)$, $\check{R}(t) = I(t_0 \vee L < t \leq X)$, and $\Lambda_T(u|\boldsymbol{Z}, T > t_0)$ is the cumulative hazard function of T conditional on $\boldsymbol{Z}$ and $T > t_0$, namely, $-\log \Pr(T > t|\boldsymbol{Z}, T > t_0)$. Here $\vee$ is the maximum operator. Note that $\check{M}(t)$ resembles the standard martingale for T truncated by $t_0 \vee L$ except for the conditional hazard function involved. The use of the conditional hazard $\mathrm{d}\Lambda_T(u|\boldsymbol{Z}, T > t_0)$ in place of $\mathrm{d}\Lambda_T(u|\boldsymbol{Z})$ in $\check{M}(t)$ is in tune with the assumed conditional model (3.7).

There are two key facts that need to be verified before we adapt the estimation procedure in Section 3.1.2. First, we need to show $E\{\check{M}(t)|\boldsymbol{Z}\} = 0$ for $\forall t \geq t_0$. To this end, standard arguments based on martingale however may not be directly applicable, and we instead prove this by examining the connection between $M(t)$ and $\check{M}(t)$. A detailed proof is provided in Section 3.7. Secondly, we need to have $g\{\boldsymbol{Z}^T\boldsymbol{\alpha_0}(0)\} = 0$ as the boundary condition. Like in the unconditional case, this easily follows from the definition of $Q_T(\tau|\boldsymbol{Z}, T > t_0)$, and model assumption (3.7).

The estimating equation, motivated by $E\{\check{M}(g\{\boldsymbol{Z}^T\boldsymbol{\alpha}(\tau)\} + t_0)|\boldsymbol{Z}\} = 0$, is then given by

$$n^{1/2}\check{\boldsymbol{S}}_n(\boldsymbol{\alpha}, \tau) = 0, \tag{3.8}$$

where

$$\check{\boldsymbol{S}}_n(\boldsymbol{\alpha}, \tau) = \sum_{i=1}^{n} \boldsymbol{Z}_i \bigg(\check{N}_i[g\{\boldsymbol{Z}_i^T\boldsymbol{\alpha}(\tau)\} + t_0] - \int_0^{\tau} \check{R}_i[g\{\boldsymbol{Z}_i^T\boldsymbol{\alpha}(v)\} + t_0] \mathrm{d}H(v) \bigg),$$

where $\check{N}_i(\cdot)$ and $\check{R}_i(\cdot)$ are sample analogues of $\check{N}(\cdot)$ and $\check{R}(\cdot)$.

An estimator of $\boldsymbol{\alpha}_0(\tau)$, denoted by $\hat{\boldsymbol{\alpha}}(\cdot)$, can be easily obtained based on equation (3.8) by slightly modifying the algorithm presented in Section 3.1.2. The key is to note that $\check{\boldsymbol{S}}_n(\boldsymbol{\alpha}, \tau)$ can be rewritten as

$$\check{\boldsymbol{S}}_n(\boldsymbol{\alpha}, \tau) = \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{Z}_i^* \bigg\{ N_i^*[g\{\boldsymbol{Z}_i^{*T}\boldsymbol{\alpha}(\tau)\}] - \int_0^{\tau} I(L_i^* < g\{\boldsymbol{Z}_i^{*T}\boldsymbol{\alpha}(v)\} \leq X_i^*) \, \mathrm{d}H(v) \bigg\},$$

where $\boldsymbol{Z}_i^* = \boldsymbol{Z}_i I(X_i > t_0)$, $X_i^* = X_i - t_0$, $L_i^* = L_i - t_0$, and $N_i^*(t) = I(X_i^* \leq t, \delta = 1)$. This shows that we can simply replace X_i, L_i, $\boldsymbol{Z}_i$ in the proposed estimation procedure for $\boldsymbol{\beta}_0(\tau)$ by X_i^*, L_i^*, and $\boldsymbol{Z}_i^*$ respectively to compute $\hat{\boldsymbol{\alpha}}(\tau)$. The same strategy can be used to adapt the resampling-based procedures in Section 3.1.4 to make inferences on $\boldsymbol{\alpha}_0(\tau)$.

The analogy in estimating equation also suggests the similarity in asymptotic

studies between the unconditional and conditional cases. Specifically, under regularity conditions C1′–C4′, which are C1–C4 defined based on transformed data (X^*, δ, $\boldsymbol{Z}^*$, L^*), we have the following theorems:

Theorem 3.2.1. *Assuming conditions C1′-C4′ hold and* $\lim_{n\to\infty} \|\mathscr{G}_{L_n}\| = 0$*, then*

$$sup_{\tau\in[\nu,\tau_U]}\|\hat{\boldsymbol{\alpha}}(\tau) - \boldsymbol{\alpha}_0(\tau)\| \to_p 0,$$

where $0 < \nu < \tau_U$.

Theorem 3.2.2. *Assuming conditions C1′-C4′ hold and* $\lim_{n\to\infty} \|n^{1/2}\mathscr{G}_{L_n}\| = 0$*, then* $n^{1/2}\|\hat{\boldsymbol{\alpha}}(\tau) - \boldsymbol{\alpha}_0(\tau)\|$ *converges weakly to a Gaussian process for* $\tau \in [\nu, \tau_U]$*, where* $0 < \nu < \tau_U$.

Similar to the unconditional case, the regularity condition C4′ is concerned with the identifiability of $\{\boldsymbol{\alpha}_0(\tau) : \tau \in (0, \tau_U)\}$ and implicitly impose the theoretical requirement on t_0 and τ_U. In the one-sample case, it becomes

$$f_T[g\{\boldsymbol{\beta}_0(\tau)\} + t_0] \cdot \Pr[L < g\{\boldsymbol{\beta}_0(\tau)\} + t_0 \le U] > 0, \forall \tau \in (\nu, \tau_U]$$

for $\nu \in (0, \tau_U)$. It is easy to see that, with t_0 chosen to be greater than L^-, $L < g\{\boldsymbol{\beta}_0(\tau)\} + t_0$ would hold with a positive probability and thus the above condition would only impose constraints on τ_U. This finding in the one-sample case is suggestive of the diminished identifiability issue with the lower tail of $\boldsymbol{\alpha}_0(\cdot)$ when the conditional version of quantile regression model is adopted. In practice, t_0 may be chosen as a constant which is greater than the observed lower bound of L and also produces converged estimates for model (3.7) with τ close to 0. The final selection of τ may be further adjusted according to the scientific interest regarding the quantity $Q_T(\tau|\boldsymbol{Z}, T > t_0)$, thereby yielding more meaningful conditional inference.

3.3 Extension to Handle Left Truncation

In this section, we present an extension to scenarios where left truncation is present, as frequently occurs in observational studies. Following the notation in Section 3.1.1, when X is subject to left truncation by event time A, the observed data include n i.i.d. replicates of $(X', L', A', \delta', \boldsymbol{Z})$, denoted by $\{(X', L', A', \delta', \boldsymbol{Z})\}_{i=1}^{n}$, where $\{X', L', A', \delta', \boldsymbol{Z}'\}$ follows the conditional distribution of $\{X, L, A, \delta, \boldsymbol{Z}\}$ given $X \geq A$. It is assumed that (L, U, A) is independent of T given $\boldsymbol{Z}$. Such data can be referred to as doubly censored data with left truncation. With all $L_i = 0$, the data reduce to the usual left truncated right censored data.

Adopting the same idea for doubly censored data studied in the previous sections, we construct an estimating equation for model (3.1) by utilizing the martingale structure associated with the observed truncated data described above. In the current setting, we define the at-risk process as $R'(t) = I(L' \vee A' < t \leq X')$, and can show that $M'(t) = N(t) - \int_0^t R'(s) d\Lambda_T(s)$ is a martingale process. By this fact, only minor changes to equation (3.3) may be needed in order to accommodate the presence of left truncation. The proposed estimating equation for $\boldsymbol{\beta}_0(\cdot)$ is given by

$$n^{1/2} \boldsymbol{K}_n(\boldsymbol{\beta}, \tau) = 0, \tag{3.9}$$

where

$$\boldsymbol{K}_n(\boldsymbol{\beta}, \tau) = n^{-1} \sum_{i=1}^{n} \boldsymbol{Z}_i \left(N_i[g\{\boldsymbol{Z}_i^\intercal \boldsymbol{\beta}(\tau)\}] - \int_0^\tau I[L_i' \vee A_i' < g\{\boldsymbol{Z}_i^\intercal \boldsymbol{\beta}(v)\} \leq X_i'] \, \mathrm{d}H(v) \right).$$

The estimation and inference procedures can be developed based on equation (3.9) similarly to those described in Section 3.1. Following the same reasoning, one may estimate $\boldsymbol{\alpha}_0(\cdot)$ in model (3.7) by solving the equation, $n^{1/2} \check{K}_n(\boldsymbol{\alpha}, \tau) = 0$, with $\check{K}_n(\boldsymbol{\alpha}, \tau) = \sum_{i=1}^{n} \boldsymbol{Z}_i \left(\check{N}_i[g\{\boldsymbol{Z}_i^\intercal \boldsymbol{\alpha}(\tau)\} + t_0] - \int_0^\tau \check{R}_i'[g\{\boldsymbol{Z}_i^\intercal \boldsymbol{\alpha}(v)\} + t_0] \mathrm{d}H(v) \right)$, where $\check{R}_i'(t) = I(L' \vee$

$A' \vee t_0 < t \leq X'$). Theoretical justifications for these extensions are expected to be very similar to those for the doubly censoring setting discussed in Section 3.2.

3.4 Simulation Studies

We studied the finite-sample performance of the proposed methods through Monte-Carlo simulations. We first generated event times from an AFT model with i.i.d. errors:

$$\log T = b_1 Z_1 + b_2 Z_2 + \epsilon,$$

where ϵ followed the extreme value distribution. The covariates Z_1 and Z_2 were generated from $Unif(0,1)$ and $Bernoulli(0.5)$ respectively. We obtained the right censoring time U and left censoring time L by generating U and L respectively from $Unif(0.1 \cdot I(Z_2 = 1), c_u)$ and $Unif(0, c_l) \cdot W$ until $L \leq U$ with Z_2 fixed, where W was a $Bernoulli(1 - p_0)$ variate. It is easy to show that under this set-up, both model (3.1) and model (3.7) hold with $g(\cdot) = \exp(\cdot)$, $\boldsymbol{\beta}_0(\tau) = \boldsymbol{\alpha}_0(\tau) = \{Q_\epsilon(\tau), b_1, b_2\}^T$. By setting $p_0 = 0.2$, L had a probability mass of 0.2 at zero and rendered a scenario that the lower tail of $\boldsymbol{\beta}_0(\cdot)$ was identifiable. In this case, model (3.1) was considered. Choosing $b_1 = 0$, $b_2 = -0.5$, $c_l = 0.5$, $c_u = 3.8$ resulted in 20% right censoring and 20% left censoring. With $p_0 = 0$, we studied model (3.7) with $t_0 = 0.16$. We set $p_0 = 0$, $b_1 = 0$, $b_2 = -1.0$, $c_l = 0.3$, $c_u = 4.5$. The resulting proportions of right censoring and left censoring are 15% and 20% respectively. Under each configuration, we generated 1000 data sets of sample size $n = 200$. We set $B = 200$ in the resampling procedures with $\{\zeta_i\}_{i=1}^B$ generated from $Exp(1)$. An equally spaced grid on τ with size 0.01 was adopted when estimating $\boldsymbol{\beta}_0(\cdot)$ or $\boldsymbol{\alpha}_0(\cdot)$. We also carried out tests on the overall significance and the constant effect hypotheses for each covariate. In the latter test we adopted the weight function $\Theta(v) = I\{v \geq (l+u)/2\}$. We set $l = 0.1$ and $u = 0.7$.

Table 3.1 presents the results from estimating model (3.1) and model (3.7) in the AFT model setting. We report absolute values of biases (Bias), empirical standard deviations (EmpSD), and average estimated resampling-based standard deviations (AvgSD) of $\hat{\boldsymbol{\beta}}(\tau)$ and $\hat{\boldsymbol{\alpha}}(\tau)$, and coverage rates of 95% Wald confidence intervals of $\boldsymbol{\beta}_0(\tau)$ and $\boldsymbol{\alpha}_0(\tau)$ with $\tau = 0.1, 0.3, 0.5$ and 0.7. It is observed from Table 3.1 that in either unconditional or conditional case, biases are small, the resampling-based standard deviation estimates agree well with the empirical ones, and the coverage rates are in general close to the nominal level.

Table 3.1: Simulation Results under AFT Models. Bias: absolute biases; AvgSD: average estimated resampling-based standard deviations; EmpSD: empirical standard deviations; Cov95: coverage rates of 95% Wald confidence intervals.

		Unconditional Case b_1=0, b_2=−0.5					Conditional Case b_1=0, b_2=−1			
τ		Bias	AvgSD	EmpSD	Cov95		Bias	AvgSD	EmpSD	Cov95
0.1	$\hat{\beta}_0$	0.07	0.93	0.96	0.87	$\hat{\alpha}_0$	0.02	0.77	0.72	0.90
	$\hat{\beta}_1$	0.01	1.36	1.49	0.89	$\hat{\alpha}_1$	0.06	1.18	1.15	0.92
	$\hat{\beta}_2$	0.00	0.92	0.86	0.92	$\hat{\alpha}_2$	0.01	0.75	0.66	0.95
0.3	$\hat{\beta}_0$	0.04	0.55	0.49	0.94	$\hat{\alpha}_0$	0.02	0.41	0.38	0.94
	$\hat{\beta}_1$	0.03	0.82	0.77	0.95	$\hat{\alpha}_1$	0.04	0.66	0.61	0.95
	$\hat{\beta}_2$	0.02	0.51	0.43	0.97	$\hat{\alpha}_2$	0.02	0.39	0.36	0.96
0.5	$\hat{\beta}_0$	0.01	0.35	0.31	0.96	$\hat{\alpha}_0$	0.01	0.31	0.28	0.95
	$\hat{\beta}_1$	0.03	0.55	0.50	0.95	$\hat{\alpha}_1$	0.02	0.49	0.45	0.96
	$\hat{\beta}_2$	0.01	0.32	0.29	0.95	$\hat{\alpha}_2$	0.01	0.29	0.27	0.95
0.7	$\hat{\beta}_0$	0.01	0.28	0.24	0.96	$\hat{\alpha}_0$	0.03	0.28	0.26	0.95
	$\hat{\beta}_1$	0.02	0.43	0.39	0.96	$\hat{\alpha}_1$	0.02	0.43	0.39	0.96
	$\hat{\beta}_2$	0.00	0.25	0.23	0.95	$\hat{\alpha}_2$	0.01	0.25	0.23	0.96

Table 3.2 presents the hypothesis testing results. The empirical rejection rates (ERR) for both tests at level 0.05 are reported, together with the estimated average effects (AvgEst), empirical standard deviations of the average effects, and average resampling-based standard deviation estimates of the average effects. We see that the type I errors are close to the nominal level 0.05. The estimated average covariate effects of Z_1 and Z_2 are close to the true values. The resampling-based standard

deviation estimates for the average covariate effect estimates agree well with the empirical standard deviations.

Table 3.2: Simulation Results on Hypothesis Testing and Second-Stage Inference under AFT Models. ERR: empirical rejection rates; AvgEst: estimated average effects; AvgSD: average estimated resampling-based standard deviations; EmpSD: empirical standard deviations.

		Unconditional Case			
		b_1=0, b_2=−0.5			
		H_0: $\beta(\tau) = 0, l \le \tau \le u$			H_0: $\beta(\tau) = \eta_0, l \le \tau \le u$
	ERR	AvgEst	AvgSD	EmpSD	ERR
$\hat{\beta}_0$	0.60	−0.82	0.40	0.39	0.89
$\hat{\beta}_1$	0.07	−0.03	0.60	0.63	0.05
$\hat{\beta}_2$	0.29	−0.49	0.37	0.34	0.04
		Conditional Case			
		b_1=0, b_2=−1			
		H_0: $\alpha(\tau) = 0, l \le \tau \le u$			H_0: $\alpha(\tau) = \eta_0, l \le \tau \le u$
	ERR	AvgEst	AvgSD	EmpSD	ERR
$\hat{\alpha}_0$	0.76	−0.79	0.31	0.30	0.96
$\hat{\alpha}_1$	0.06	−0.03	0.49	0.49	0.05
$\hat{\alpha}_2$	0.93	−1.01	0.30	0.29	0.03

In addition to the AFT setting, we considered a log linear model with heteroscedastic errors. That is, event times were generated from the model:

$$\log T = b_1 Z_1 + b_2 Z_2 \xi + \epsilon,$$

where Z_1 followed $Unif(0,1)$, Z_2 followed $Bernoulli(0.5)$, and ξ followed $Exp(1)$. The right censoring time U and left censoring time L were generated in the same way as in the AFT setting. We considered two different configurations: (a) ϵ was a $N(0,1)$ variate, $p_0 = 0.2, b_1 = 0, b_2 = -1.5, c_l = 0.7, c_u = 4.8$; (b) ϵ was an extreme value variate, $p_0 = 0, t_0 = 0.03, b_1 = 0, b_2 = -4.5, c_l = 0.05, c_u = 4.0$. One can verify that model (3.1) holds under configuration (a) with $g(\cdot) = \exp(\cdot)$ and $\boldsymbol{\beta}_0(\tau) = \{\beta_0(\tau), \beta_1(\tau), \beta_2(\tau)\}^T$, where $\beta_0(\tau) = Q_\epsilon(\tau)$, $\beta_1(\tau) = 0$, and $\beta_2(\tau) = Q_{b_2\xi+\epsilon}(\tau) - \beta_0(\tau)$. Under configuration (b), model (3.7) was satisfied with $g(\cdot) = \exp(\cdot)$, $\alpha_0(\tau) = Q_\epsilon(\tau) =$

$\log\{-\log(1-\tau)\}$, $\alpha_1(\tau) = 0$, and $\alpha_2(\tau) = \log[Q_{\exp(b_2\xi+\epsilon)}(1 - \Pr\{\exp(b_2\xi + \epsilon) > t_0\}(1-\tau)) - t_0] - \alpha_0(\tau)$. Here, for a random variable Y, $Q_Y(\tau)$ denotes its τ–th quantile. Note that unlike in the AFT settings, the effects of Z_2 were not constant and took a complicated analytic form. We approximated the true coefficients for Z_2 by using bootstrapping. In (a), there were 25% right censoring and 20% left censoring. In (b), the rates of right censoring and left censoring were 15% and 25% respectively.

Tables 3.3–3.4 present estimation results and hypothesis testing results under the heteroscedastic setting . We can see that in the presence of varying covariate effects, our proposed method also performs well. In both unconditional and conditional cases, the regression quantile estimates are virtually unbiased with standard deviations accurately estimated the proposed resampling method. The estimates for average effects also have small biases with estimated standard deviations agreeing well with empirical ones. The proposed test for the overall significance and the test for the constancy of covariate effect appear to have the right sizes and also reasonable power.

Table 3.3: Simulation Results under Log-Linear Models with Heteroscedastic Errors. Bias: absolute biases; AvgSD: average estimated resampling-based standard deviations; EmpSD: empirical standard deviations; Cov95: coverage rates of 95% Wald confidence intervals.

		Unconditional Case b_1=0, b_2=−1.5					Conditional Case b_1=0, b_2=−4.5			
τ		Bias	AvgSD	EmpSD	Cov95		Bias	AvgSD	EmpSD	Cov95
0.1	$\hat{\beta}_0$	0.06	0.64	0.60	0.91	$\hat{\alpha}_0$	0.05	0.71	0.62	0.94
	$\hat{\beta}_1$	0.05	1.11	1.09	0.91	$\hat{\alpha}_1$	0.01	1.15	1.04	0.94
	$\hat{\beta}_2$	0.09	1.09	1.02	0.85	$\hat{\alpha}_2$	0.02	0.82	0.75	0.94
0.3	$\hat{\beta}_0$	0.04	0.41	0.37	0.94	$\hat{\alpha}_0$	0.01	0.40	0.36	0.95
	$\hat{\beta}_1$	0.02	0.70	0.63	0.95	$\hat{\alpha}_1$	0.01	0.67	0.62	0.95
	$\hat{\beta}_2$	0.00	0.65	0.53	0.95	$\hat{\alpha}_2$	0.01	0.49	0.44	0.95
0.5	$\hat{\beta}_0$	0.01	0.33	0.31	0.94	$\hat{\alpha}_0$	0.01	0.32	0.28	0.96
	$\hat{\beta}_1$	0.00	0.56	0.52	0.95	$\hat{\alpha}_1$	0.03	0.54	0.49	0.95
	$\hat{\beta}_2$	0.01	0.42	0.37	0.95	$\hat{\alpha}_2$	0.00	0.39	0.36	0.95
0.7	$\hat{\beta}_0$	0.02	0.32	0.29	0.96	$\hat{\alpha}_0$	0.03	0.30	0.27	0.95
	$\hat{\beta}_1$	0.00	0.53	0.49	0.95	$\hat{\alpha}_1$	0.02	0.51	0.46	0.96
	$\hat{\beta}_2$	0.02	0.35	0.31	0.96	$\hat{\alpha}_2$	0.01	0.36	0.33	0.95

Table 3.4: Simulation Results on Hypothesis Testing and Second-Stage Inference under Log-Linear Models with Heteroscedastic Errors. ERR: empirical rejection rates; AvgEst: estimated average effects; AvgSD: average estimated resampling-based standard deviations; EmpSD: empirical standard deviations.

Unconditional Case					
b_1=0, b_2=-1.5					
	H_0: $\beta(\tau) = 0, l \leq \tau \leq u$				H_0: $\beta(\tau) = \eta_0, l \leq \tau \leq u$
	ERR	AvgEst	AvgSD	EmpSD	ERR
$\hat{\beta}_0$	0.17	-0.32	0.32	0.31	0.90
$\hat{\beta}_1$	0.06	0.00	0.54	0.53	0.03
$\hat{\beta}_2$	0.96	-1.50	0.45	0.42	0.23
Conditional Case					
b_1=0, b_2=-4.5					
	H_0: $\alpha(\tau) = 0, l \leq \tau \leq u$				H_0: $\alpha(\tau) = \eta_0, l \leq \tau \leq u$
	ERR	AvgEst	AvgSD	EmpSD	ERR
$\hat{\alpha}_0$	0.79	-0.79	0.30	0.28	0.98
$\hat{\alpha}_1$	0.05	-0.02	0.50	0.48	0.04
$\hat{\alpha}_2$	0.99	-1.66	0.37	0.36	0.14

We also evaluated the performance of our method in doubly censored data with left truncation. The set-ups and true parameter values were chosen to be the same as for the conditional model investigated before, except that we now imposed an additional left truncation time, A, which was generated from $Unif(0, 0.5)$ for the AFT model case and $Unif(0, 0.03)$ for the heteroscedastic model case. The resulting truncation portions were around 15% for both cases. t_0 was still set to be 0.16 and 0.03 respectively. Tables 3.5 and 3.6 present simulation results for the left truncation scenario. It can be seen that our method also performs well under this setting.

Simulations were also conducted to compare our approach with a naive approach that simply discards all left-censored subjects. Data were generated from the same configurations as the unconditional cases in Table 3.1 and Table 3.2. We evaluated the estimation of model (3.1). Figure 3.1 displays the mean estimated coefficients from the proposed approach and those from the naive approach along with the true coefficients. This figure shows that the proposed estimators are virtually unbiased while the naive approach can produce substantial biases particularly in the estimation

of non-zero coefficients. This finding shows that it is important to account for double censoring arising from practical situations.

Table 3.5: Simulation Results on Parameter Estimation for Doubly Censored Data with Left Truncation. Bias: absolute biases; AvgSD: average estimated resampling-based standard deviations; EmpSD: empirical standard deviations; Cov95: coverage rates of 95% Wald confidence intervals.

		AFT Case b_1=0, b_2=−0.5					Heteroscedastic Case b_1=0, b_2=−1			
τ		Bias	AvgSD	EmpSD	Cov95		Bias	AvgSD	EmpSD	Cov95
0.1	$\hat{\alpha}_0$	0.06	0.62	0.57	0.92	$\hat{\alpha}_0$	0.01	0.59	0.55	0.92
	$\hat{\alpha}_1$	0.02	0.95	0.91	0.93	$\hat{\alpha}_1$	0.01	0.96	0.89	0.94
	$\hat{\alpha}_2$	0.05	0.61	0.54	0.95	$\hat{\alpha}_2$	0.06	0.66	0.59	0.95
0.3	$\hat{\alpha}_0$	0.02	0.29	0.26	0.95	$\hat{\alpha}_0$	0.01	0.36	0.34	0.94
	$\hat{\alpha}_1$	0.00	0.45	0.43	0.96	$\hat{\alpha}_1$	0.02	0.59	0.56	0.95
	$\hat{\alpha}_2$	0.00	0.28	0.27	0.96	$\hat{\alpha}_2$	0.03	0.41	0.40	0.94
0.5	$\hat{\alpha}_0$	0.00	0.20	0.20	0.94	$\hat{\alpha}_0$	0.01	0.29	0.28	0.95
	$\hat{\alpha}_1$	0.00	0.32	0.31	0.95	$\hat{\alpha}_1$	0.01	0.49	0.46	0.95
	$\hat{\alpha}_2$	0.00	0.19	0.19	0.95	$\hat{\alpha}_2$	0.03	0.35	0.34	0.93
0.7	$\hat{\alpha}_0$	0.01	0.18	0.16	0.95	$\hat{\alpha}_0$	0.03	0.28	0.25	0.95
	$\hat{\alpha}_1$	0.00	0.28	0.27	0.95	$\hat{\alpha}_1$	0.01	0.47	0.42	0.96
	$\hat{\alpha}_2$	0.01	0.16	0.16	0.95	$\hat{\alpha}_2$	0.02	0.33	0.31	0.95

3.5 The CFFPR Data Example

We apply the proposed quantile regression method to the CFFPR data discussed in Section 2.1. Cystic Fibrosis (CF) is one of the most common and life-shortening genetic disorders affecting the lungs and digestive systems of about 30,000 children and adults in the United States and 70,000 worldwide (Cystic Fibrosis Foundation 2010). Pseudomonas aeruginosa (PA), the predominant bacterial pathogen infecting 80% of CF patients under age 18, accelerates decline in lung function (Kosorok et al. 2001) and serves as an important predictor of mortality in CF (Retsch-Bogart et al. 2008). In our analysis, we used the CFFPR data collected during 1986-2005 to investigate the association between onset ages of the first detected PA infection and several risk factors in CF patients diagnosed by age 10. Similar data were analyzed

Table 3.6: Simulation Results on Hypothesis Testing and Second-Stage Inference for Doubly Censored Data with Left Truncation. ERR: empirical rejection rates; AvgEst: estimated average effects; AvgSD: average estimated resampling-based standard deviations; EmpSD: empirical standard deviations.

AFT Case					
b_1=0, b_2=−1					
	H_0: $\beta(\tau) = 0, l \le \tau \le u$				H_0: $\beta(\tau) = \eta_0, l \le \tau \le u$
	ERR	AvgEst	AvgSD	EmpSD	ERR
$\hat{\alpha}_0$	0.97	−0.81	0.22	0.22	0.99
$\hat{\alpha}_1$	0.05	−0.00	0.36	0.35	0.05
$\hat{\alpha}_2$	0.98	−1.00	0.22	0.22	0.05
Heteroscedastic Case					
b_1=0, b_2=−4.5					
	H_0: $\alpha(\tau) = 0, l \le \tau \le u$				H_0: $\alpha(\tau) = \eta_0, l \le \tau \le u$
	ERR	AvgEst	AvgSD	EmpSD	ERR
$\hat{\alpha}_0$	0.87	−0.78	0.26	0.26	0.99
$\hat{\alpha}_1$	0.06	−0.02	0.43	0.44	0.04
$\hat{\alpha}_2$	0.99	−1.46	0.32	0.31	0.13

by Lai et al. (2004) under the Cox model and Yan et al. (2009) based on temporal process regression.

In the CFFPR, a patient's age at the first PA infection, which is the event time of interest T, is subject to both left and right censoring. Among 12,818 CF patients diagnosed between 1986 and 2000, 3,343 (26.1%) patients had PA infection at study entry (i.e. T is left censored by the patient's age at the first CFFPR record L) and 2,213 (17.3%) patients had no PA infection documented by December 2005 (i.e. T is right censored by age at the last follow-up before the cut-off date U). To avoid the complication with delayed entry, we restricted the study population to subjects who were diagnosed before age 10 years during 1986–2000 and alive at age 10 years. The first restriction was imposed because the first 10 years were known to have greatest potential to take advantage of early diagnosis (Campbell and White, 2005; Grosse et al., 2006). The second restriction was imposed to avoid left truncation due to mortality prior to CF diagnosis. Since the mortality rate before age 10 years was very low, about 1.5% (Grosse et al., 2006), we expect excluding patients who died

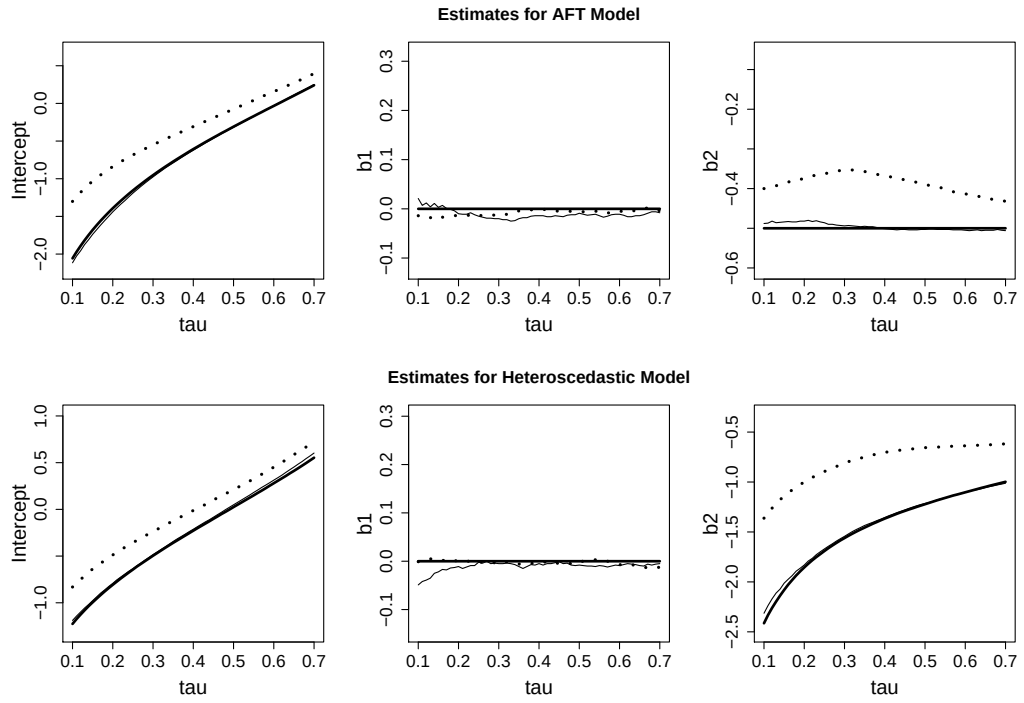

Figure 3.1: Comparison among True coefficients (Bold Solid Lines), Mean Estimated Coefficients from the Proposed Method (Solid Lines), and Mean Estimated Coefficients from the Naive Approach (Dotted Lines).

before age 10 years would only result in a small deviation from the general young CF population. The restricted sample contains 11,179 patients with 23.7% left censoring and 16.2% right censoring.

We applied the proposed quantile regression method to this doubly-censored restricted CFFPR sample. Since the support of left censoring time L in this dataset appears to have a lower bound approaching 0, suggesting the lower tail identifiability, we fit the data with the unconditional model (3.1), choosing $g(\cdot)$ as the identity function. The same set of covariates examined in Yan et al. (2009) was considered, including gender (1 for females and 0 for males), diagnosis mode (denoted by "factor") and diagnosis year (denoted by "dx"). Diagnosis mode was defined according to common clinical practices that identify CF prior to 2005, which includes four categories: diagnosis at birth due to meconium ileus (MI), diagnosis shortly after birth by neonatal/prenatal screening (SCR), diagnosis at variable ages because of family

history (FH), and diagnosis at variable ages from various symptoms (SYMP) other than MI (Lai et al., 2004; Yan et al., 2009). Diagnosis year was classified into three periods, i.e., 1986-1989 (dx86), 1990-1993 (dx90), and 1994-2000 (dx94), that coincided with the major therapeutic breakthroughs in CF, which are, Pulmozyme in 1994 (Fuchs et al. 1994; Ramsey and Dorkin 1994) and TOBI in 1999 (Ramsey et al. 1999). Boy patients who were diagnosed between 1986 and 1989 (dx86) by symptoms other than MI (SYMP) was chosen as the reference comparison group.

Figure 3.2 displays the proposed coefficient estimates for $\boldsymbol{\beta}_0(\tau)$ in bold solid lines along with their 95% pointwise Wald confidence intervals in bold dashed lines for $\tau \in [l, u]$ with l=0.10 and u=0.65. The naive estimates obtained from the subset excluding left-censored observations and the corresponding 95% confidence intervals are plotted in dot-dash lines and dotted lines respectively. The estimates for $\boldsymbol{\beta}_0(\tau)$ with τ close to 0 (not shown), though exhibiting rather large variability, are all converged solutions to the L_1–minimization problem (3.5). This further suggests that the lower tail identifiability of regression quantiles may be of little concern in this example. As shown in Figure 3.2 (panel A), with all observations included in the analysis, the estimated intercept indicates that about 10% of male patients with CF diagnosed during 1986-1989 by SYMP acquired their first PA infection by age 2 years and approximately 65% of them had their first PA infection by age 9 years.

With regard to the gender effect (Figure 3.2, panel B), the regression coefficient exhibited a cross-over pattern, that is, it is first positive and then flips the sign around $\tau = 0.35$. This indicates that girls acquired their first PA infection earlier than boys except for patients below 35% quantiles among each gender. More importantly, the gender difference is more pronounced at larger τ's, which correspond to patients who acquired first PA infection at older ages. Similarly, many other covariates show non-constant effects across the quantile range of 0.1 to 0.65. For example, the effect of "FH diagnosis" (Figure 3.2, panel D) increases with τ, while the coefficient estimates

for both "diagnosis 1990-1993" and "diagnosis 1994-2000"(Figure 3.2, panel F and G) decrease with τ. Such varying effects would not have been identified by traditional Cox regression or classic linear regression that only models the mean.

Our result on the beneficial effect of "FH diagnosis" (i.e., later first PA infection compared to the "SYMP diagnosis") appears to be new. This effect would have been masked by an opposite effect if left-censored cases were excluded from the analysis. Panel D of Figure 3.2 shows that even the pointwise 95% confidence intervals do not overlap across the entire range of τ, leading to an erroneous conclusion on the effect of "FH diagnosis". Substantial discrepancy is also noted on the magnitude of the regression estimates of covariates "MI diagnosis" (panel C), "diagnosis 1990-1993" (panel F) and "diagnosis 1994-2000" (panel G) between the proposed method and the naive method. These systematic differences are also observed in our simulation studies, and provide strong evidence to support the importance of appropriately accommodating left censored observations when investigating PA infections in CF.

Formal tests on the significance of covariate effects were performed based on the average quantile effects across τ ranging from 0.10 to 0.65. The delaying effect of FH diagnosis and the accelerating effects of more recent diagnosis cohorts (1990-1993 and 1994-2000) on first PA infections observed in Figure 3.2 (panels D, F, and G) are confirmed by very significant p-values, which are all < 0.001. We paid particular attention to testing the gender effect due to its cross-over pattern as shown in Figure 3.2, and assessed the aggregated gender effect in two $\tau-$intervals:$\tau \in [0.1, 0.35)$ and $\tau \in [0.35, 0.65]$. Our test is significant in neither interval, with both p-values around 0.2 (data not shown).

In the view of rather monotone patterns of the coefficients for these covariates, we conducted the proposed constancy tests using the weight function $\tilde{\Theta}(t) = I[t < (l+u)/2]$. These analyses show that the effects of gender, FH, dx90 and dx94 may vary across τ, and the corresponding p-values are 0.01, 0.001, < 0.001, and < 0.001

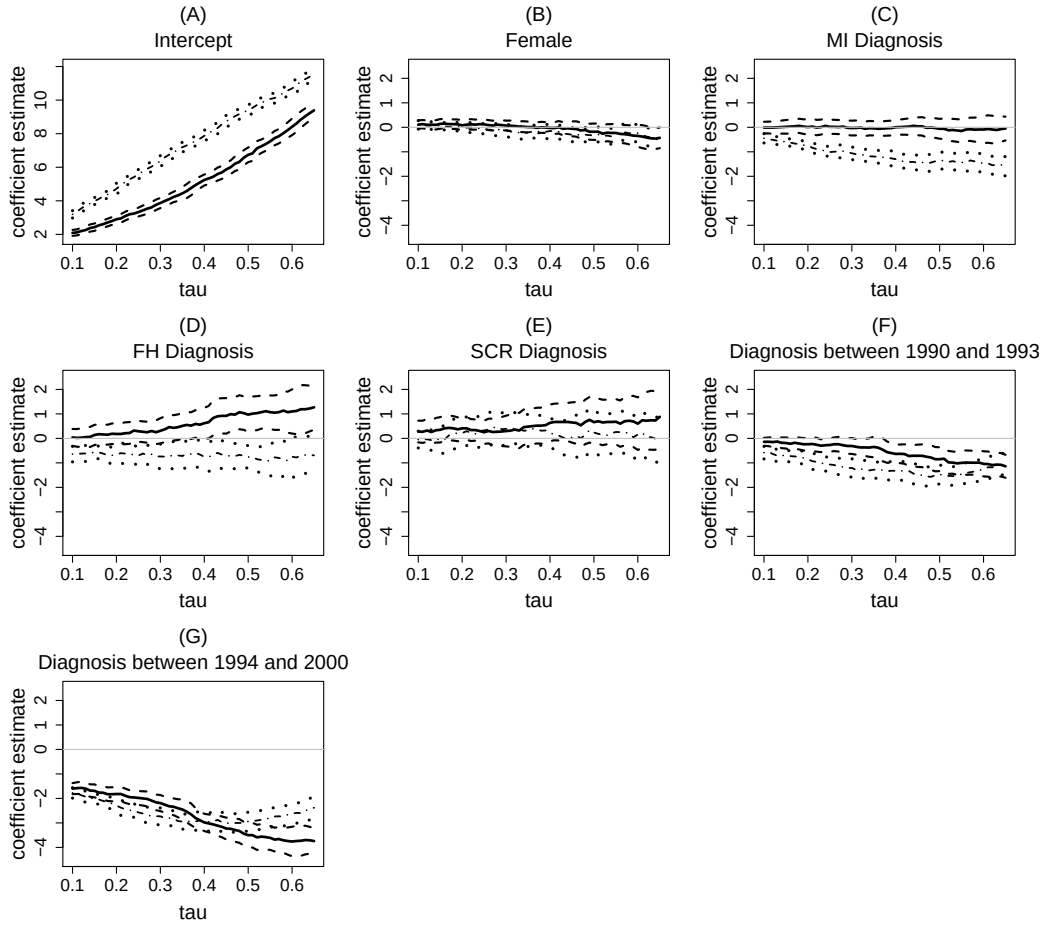

Figure 3.2: Coefficient Estimates (Bold Solid Lines) and 95% Pointwise Confidence Intervals (Bold Dashed Lines) from the Proposed Method, in Contrast with Coefficient Estimates (Dot-Dash Lines) and 95% Pointwise Confidence Intervals (Dotted Lines) from the Naive Method.

respectively, confirming the visual trends illustrated in Figure 3.2. These results have some interesting clinical implications. The earlier acquisition of first PA infections in females with CF reported repeatedly in the literature (Lai et al., 2004; Yan et al., 2009) is not uniform across the entire female population; this gender effect is smaller among CF patients who are subject to high risk versus low risk of PA infection. The association between more recent diagnosis cohorts and shorter times to first PA infection may be explained by the increased culture frequency in these patients, which may shorten the time to detecting PA infection in patients with late onsets of PA infection in a greater extent, as compared to those who experienced first PA infections at young ages. This may be because patients in the latter group tended to have frequent sick visits in early life, which may offset the benefit of frequent cultures in the detection of PA infection.

3.6 Remarks

In this chapter we propose a quantile regression method for doubly censored data with known left censoring times. The stochastic integration presentation of the proposed estimating equation facilitates asymptotic studies and entails computationally simple implementations. A useful solution to handle the unique identifiability issue with doubly censored regression quantiles is proposed based on conditional inference. We also present an adaptation of our method to settings where left truncation is present.

The double censoring mechanism, $(L, U) \perp C$ given $\boldsymbol{Z}$, is adopted in this work and thus both L and U are allowed to depend on $\boldsymbol{Z}$. In addition, we require L be always observed here, which is often true in registry study settings, as exemplified by the CFFPR data. Such additional information on left censoring time contributes to identifying an appropriate martingale for constructing estimating equations. How to relax the assumption on known L merits future research, and we conduct further

investigation in the next chapter.

3.7 Proofs

3.7.1 Proof of Theorem 3.1.2

Following the proof of *Lemma B.1.* in Peng and Huang (2008) we can show that

$$\sup_{\tau\in(0,\tau_U]}\|n^{-1/2}\sum_{i=1}^{n}\boldsymbol{Z}_i\{N_i(\exp\{\boldsymbol{Z}_i^{\intercal}\hat{\boldsymbol{\beta}}(\tau)\})-N_i(\exp\{\boldsymbol{Z}_i^{\intercal}\boldsymbol{\beta}_0(\tau)\})\}$$
$$-n^{1/2}(\boldsymbol{\mu}\{\hat{\boldsymbol{\beta}}(\tau)\}-\boldsymbol{\mu}\{\boldsymbol{\beta}_0(\tau)\})\|\to_p 0. \tag{3.10}$$

and

$$\sup_{\tau\in(0,\tau_U]}\|n^{-1/2}\sum_{i=1}^{n}\boldsymbol{Z}_i\{I(L_i<\exp\{\boldsymbol{Z}_i^{\intercal}\hat{\boldsymbol{\beta}}(\tau)\}\le X_i)-I(L_i<\exp\{\boldsymbol{Z}_i^{\intercal}\boldsymbol{\beta}_0(\tau)\}\le X_i)\}$$
$$-n^{1/2}(\tilde{\boldsymbol{\mu}}\{\hat{\boldsymbol{\beta}}(\tau)\}-\tilde{\boldsymbol{\mu}}\{\boldsymbol{\beta}_0(\tau)\})\|\to_p 0. \tag{3.11}$$

(3.10) and (3.11) together with the uniform convergence of $\boldsymbol{\mu}\{\hat{\boldsymbol{\beta}}(\tau)\}$ for $\tau\in(0,\tau_U]$ imply a stochastic differential equation as mentioned in Peng and Huang (2008):

$$\begin{aligned}-n^{1/2}\boldsymbol{S}_n(\boldsymbol{\beta}_0,\tau)=&n^{1/2}[\boldsymbol{\mu}\{\hat{\boldsymbol{\beta}}(\tau)\}-\boldsymbol{\mu}\{\boldsymbol{\beta}_0(\tau)\}]\\&-\int_0^{\tau}[\boldsymbol{J}\{\boldsymbol{\beta}_0(u)\}\boldsymbol{B}\{\boldsymbol{\beta}_0(u)\}^{-1}+o_{(0,\tau_U]}(1)]\\&\qquad\times n^{1/2}[\boldsymbol{\mu}\{\hat{\boldsymbol{\beta}}(\tau)\}-\boldsymbol{\mu}\{\boldsymbol{\beta}_0(\tau)\}]\,\mathrm{d}H(u)+o_{(0,\tau_U]}(1).\end{aligned}$$

Using the production integration theory (Gill and Johansen, 1990; Andersen et al., 1998), we have

$$n^{1/2}[\boldsymbol{\mu}\{\hat{\boldsymbol{\beta}}(\tau)\}-\boldsymbol{\mu}\{\boldsymbol{\beta}_0(\tau)\}]=\phi\{-n^{1/2}\boldsymbol{S}_n(\boldsymbol{\beta}_0,\tau)\}+o_{(0,\tau_U]}(1), \tag{3.12}$$

where ϕ is a linear operator. By the Donsker theorem, $-n^{1/2}\boldsymbol{S}_n(\boldsymbol{\beta}_0,\tau)$ converges weakly to a tight Gaussian process $\boldsymbol{G}(\tau)$ for $\tau\in(0,\tau_U]$. Hence $n^{1/2}[\boldsymbol{\mu}\{\hat{\boldsymbol{\beta}}(\tau)\}-$

$\boldsymbol{\mu}\{\boldsymbol{\beta}_0(\tau)\}]$ converges weakly to $\phi\{\boldsymbol{G}(\tau)\}$ which is also a Gaussian process. Using Taylor expansions we immediately have $n^{1/2}\{\hat{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}_0(\tau)\}$ converges weakly to the Gaussian process $\boldsymbol{B}\{\boldsymbol{\beta}_0(\tau)\}^{-1}\phi\{\boldsymbol{G}(\tau)\}$ for $\tau \in [\nu, \tau_U]$, where the lower limit ν ensures $\boldsymbol{B}\{\boldsymbol{\beta}_0(\tau)\}^{-1}$ is uniformly bounded.

3.7.2 Justification for $E\{\check{M}(t)|\boldsymbol{Z}\} = 0$

From the arguments in the unconditional case we know that

$$\begin{aligned} E\{N(t)|\boldsymbol{Z}\} &= \int_0^t I(L < u \leq X)\lambda(u|\boldsymbol{Z})\,\mathrm{d}u \\ &= \int_0^t I(u \leq X)\lambda(u|\boldsymbol{Z})\,\mathrm{d}u - \int_0^t I(u \leq L)\lambda(u|\boldsymbol{Z})\,\mathrm{d}u \\ &= \Lambda_T(t \wedge X|\boldsymbol{Z}) - \Lambda_T(t \wedge L|\boldsymbol{Z}). \end{aligned}$$

Hence, for $t > t_0$, we have

$$\begin{aligned} E\{N(t) - N(t_0)|\boldsymbol{Z}\} &= \{\Lambda_T(t \wedge X|\boldsymbol{Z}) - \Lambda_T(t_0 \wedge X|\boldsymbol{Z})\} \\ &\quad - \{\Lambda_T(t \wedge L|\boldsymbol{Z}) - \Lambda_T(t_0 \wedge L|\boldsymbol{Z})\}. \end{aligned} \tag{3.13}$$

On the other hand, we have

$$\begin{aligned} &\Lambda_T(t \wedge X|\boldsymbol{Z}) - \Lambda_T(t_0 \wedge X|\boldsymbol{Z}) \\ &\quad = I(X > t_0)\{\Lambda_T(t \wedge X|\boldsymbol{Z}) - \Lambda_T(t_0|\boldsymbol{Z})\} \\ &\quad = I(X > t_0)[-\log\{1 - Pr(T \leq t \wedge X|\boldsymbol{Z})\} \\ &\qquad + \log\{1 - Pr(t \leq t_0|\boldsymbol{Z})\}] \\ &\quad = I(X > t_0)[-\log\{1 - \frac{Pr(T \leq t \wedge X|\boldsymbol{Z}) - Pr(T \leq t_0|\boldsymbol{Z})}{1 - Pr(T \leq t_0|\boldsymbol{Z})}\}] \\ &\quad = I(X > t_0)[-\log\{1 - Pr(T \leq t \wedge X|T > t_0, \boldsymbol{Z})\}] \\ &\quad = I(X > t_0)\Lambda_T(t \wedge X|T > t_0, \boldsymbol{Z}). \end{aligned} \tag{3.14}$$

Similarly we can show that

$$\Lambda_T(t \wedge L|\boldsymbol{Z}) - \Lambda_T(t_0 \wedge L|\boldsymbol{Z}) = I(L > t_0)\Lambda_T(t \wedge L|T > t_0, \boldsymbol{Z}). \tag{3.15}$$

Plugging (3.14) and (3.15) into (3.13), we have

$$\begin{aligned} E\{N(t) - N(t_0)|\boldsymbol{Z}\} =& I(X > t_0)\Lambda_T(t \wedge X|T > t_0, \boldsymbol{Z}) \\ & - I(L > t_0)\Lambda_T(t \wedge L|T > t_0, \boldsymbol{Z}) \\ =& \int_0^t \{I(X > t_0)I(u \le X) - I(L > t_0)I(u \le L)\} \\ & \times \lambda_T(u|T > t_0, \boldsymbol{Z})\,\mathrm{d}u \\ =& \int_0^t \{L < u \le X\}\lambda_T(u|T > t_0, \boldsymbol{Z})\,\mathrm{d}u \\ =& \int_0^\tau I(L < F_T^{-1}(v|T > t_0, \boldsymbol{Z}_i) \le X)\,\mathrm{d}H(v). \end{aligned} \tag{3.16}$$

From (3.13) and (3.16) we immediately have $E\{\check{M}(t)|\boldsymbol{Z}\} = 0$.

Chapter 4

Quantile Regression for Doubly Censored Data

4.1 Quantile Regression Procedures

4.1.1 Data and Model

The structure of the data we study in this chapter is very similar to that in Chapter 1, with the only distinction that the left censoring time may not be observed here, while it is always known in Chapter 1. More specifically, the data consists of $(X, \delta, \boldsymbol{Z})$, where $X = \max\{L, \min(T, U)\}$, with L and U denoting the left and right censoring times respectively, $\delta = I(L < T \leq U) + 2I(T \leq L) + 3I(T > U)$, and $\boldsymbol{Z} = (1, Z^{(1)}, Z^{(2)}, \cdots, Z^{(p)})^T$ is the covariate vector. We observe n i.i.d. replicates of $(X, \delta, \boldsymbol{Z})$, denoted by $\{X_i, \delta_i, \boldsymbol{Z}_i\}_{i=1}^n$.

We consider the quantile regression model

$$Q_T(\tau|\boldsymbol{Z}) = g\{\boldsymbol{Z}^T\boldsymbol{\beta}_0(\tau)\}, \quad \tau \in (0, 1), \tag{4.1}$$

where $Q_T(\tau|\boldsymbol{Z}) = \inf\{t : \Pr(T \leq t|\boldsymbol{Z}) \geq \tau\}$, $g(\cdot)$ is a known monotone link function, and $\boldsymbol{\beta}_0(\tau)$ is a vector of unknown coefficients.

4.1.2 Estimation Procedure

For any random variable W, define $F_W(t) = \Pr(W \leq t)$. In the one sample setting, Turnbull (1974) suggested a self-consistent estimator for $F_T(t)$, which can essentially be solved from the following estimating equation:

$$F_T(t) = \frac{1}{n}\sum_{i=1}^{n}\Big\{I(X_i \leq t, \delta_i = 1) + I(X_i \leq t, \delta_i = 2) + I(X_i > t, \delta_i = 2)\frac{F_T(t)}{F_T(X_i)} + I(X_i \leq t, \delta_i = 3)\frac{F_T(t) - F_T(X_i)}{1 - F_T(X_i)}\Big\}. \tag{4.2}$$

To see the self-consistency of equation (4.2), it is equivalent to show $E(\phi) = 0$,

where ϕ is defined as follows:

$$\begin{aligned}
\phi \equiv & \{I(X \le t) - F_T(t)\} I(\delta = 1) + \{I(X \le t) + I(X > t) \cdot \frac{F_T(t)}{F_T(X)} - F_T(t)\} I(\delta = 2) \\
& + \{I(X \le t) \cdot \frac{F_T(t) - F_T(X)}{1 - F_T(X)} - F_T(t)\} I(\delta = 3) \\
= & \{I(T \le t) - F_T(t)\} \qquad (4.3) \\
& + \big\{I(L \le t) + I(L > t) \cdot \frac{F_T(t)}{F_T(L)} - I(T \le t)\big\} I(T \le L) \qquad (4.4) \\
& + \big\{I(U \le t) \cdot \frac{F_T(t) - F_T(U)}{1 - F_T(U)} - I(T \le t)\big\} I(T > U). \qquad (4.5)
\end{aligned}$$

For simplicity, we denote the expressions in (4.3), (4.4) and (4.5) by ϕ_1, ϕ_2 and ϕ_3, respectively. It is obvious that $E(\phi_1) = 0$, and we also have:

$$\begin{aligned}
E(\phi_2) &= E\big\{E\big[I(L \le t) + I(L > t) \cdot \frac{F_T(t)}{F_T(L)} - I(T \le t)\big] I(T \le L) | L\big\} \\
&= E\big\{E\big[I(L \le t) I(T \le L) + I(L > t) \cdot \frac{F_T(t)}{F_T(L)} \cdot I(T \le L) \\
&\qquad - \{I(L \le t) I(T \le L) + I(L > t) I(T \le t)\} | L\big]\big\} \\
&= E\big[I(L \le t) F_T(L) + I(L > t) F_T(t) - \{I(L \le t) F_T(L) + I(L > t) F_T(t)\}\big] \\
&= 0, \\
E(\phi_3) &= E\big\{E\big[I(U \le t) \cdot \frac{F_T(t) - F_T(U)}{1 - F_T(U)} - I(T \le t)\big] I(T > U) | U\big\} \\
&= E\big\{E\big[I(U \le t) \cdot \frac{F_T(t) - F_T(U)}{1 - F_T(U)} \cdot I(T > U) - I(U < T \le t) | U\big]\big\} \\
&= E\big[I(U \le t) \cdot \{F_T(t) - F_T(U)\} - I(U \le t) \cdot \{F_T(t) - F_T(U)\}\big] \\
&= 0.
\end{aligned}$$

Thus we have justified the estimating equation (4.2).

Let $\tilde{T} = g^{-1}(T)$, $\tilde{X} = g^{-1}(X)$, $\tilde{N}(t) = I(\tilde{X} \le t, \delta = 1)$, $\tilde{L}(t) = I(\tilde{X} \le t, \delta = 2)$ and $\tilde{R}(t) = I(\tilde{X} \le t, \delta = 3)$. We can easily obtain an estimating equation for $F_{\tilde{T}}(t)$

from (4.2). Further, we can rewrite it as a stochastic integral:

$$F_{\tilde{T}}(t) = \frac{1}{n}\sum_{i=1}^{n}\Big\{\tilde{N}_i(t) + \tilde{L}_i(t) + \int_t^{\infty}\frac{F_{\tilde{T}}(t)}{F_{\tilde{T}}(u)}\,\mathrm{d}\tilde{L}_i(u) + \tilde{R}_i(t)\int_0^t \frac{F_{\tilde{T}}(t) - F_{\tilde{T}}(u)}{1 - F_{\tilde{T}}(u)}\,\mathrm{d}\tilde{R}_i(u)\Big\}. \tag{4.6}$$

Assuming the continuity of $F_{\tilde{T}}(u)$ and applying integral by parts to the right hand side of (4.6), we have

$$\begin{aligned} F_{\tilde{T}}(t) = \frac{1}{n}\sum_{i=1}^{n}\Big\{\tilde{N}_i(t) + F_{\tilde{T}}(t)\tilde{L}_i(\infty) + F_{\tilde{T}}(t)\int_t^{\infty}\frac{\tilde{L}_i(u)}{\{F_{\tilde{T}}(u)\}^2}\,\mathrm{d}F_{\tilde{T}}(u) \\ + (1 - F_{\tilde{T}}(t))\int_0^t \frac{\tilde{R}_i(u)}{\{1 - F_{\tilde{T}}(u)\}^2}\,\mathrm{d}F_{\tilde{T}}(u). \end{aligned} \tag{4.7}$$

The above equation can easily accommodate covariates and thus naturally extend to the regression setting under model (4.1). With t replaced by $\boldsymbol{Z}_i\boldsymbol{\beta}(\tau)$, equation (4.7) can be adapted to be an estimating equation for the regression coefficient $\boldsymbol{\beta}_0(\cdot)$:

$$n^{1/2}\boldsymbol{S}_n(\boldsymbol{\beta}, \tau) = 0, \tag{4.8}$$

where

$$\begin{aligned} \boldsymbol{S}_n(\beta, \tau) = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{Z}_i\Big[\tilde{N}_i\{\boldsymbol{Z}_i^T\boldsymbol{\beta}(\tau)\} + \tau\tilde{L}_i(\infty) + \tau\int_{\tau}^{1}\frac{\tilde{L}_i\{\boldsymbol{Z}_i^T\boldsymbol{\beta}(\tau)\}}{u^2}\,\mathrm{d}u \\ + (1-\tau)\int_0^{\tau}\frac{\tilde{R}_i\{\boldsymbol{Z}_i^T\boldsymbol{\beta}(\tau)\}}{(1-u)^2}\,\mathrm{d}u - \tau\Big]. \end{aligned}$$

The proposed estimating equation (4.8) bears similar spirits with that of Lin et al. (2012) in the self-consistency property, but distinguishes itself by the stochastic integral representation. This new formulation of the self-consistent estimating equation offers two major advantages. First, it evokes a clearly-defined estimation procedure, which is discussed in details in Section 4.1.3. More importantly, it may greatly facilitate understanding of the large sample properties, as can be seen in the work of

Peng and Huang (2008) and Peng (2011) on the random right censoring case. Note that under the random right censorship, equation (4.8) reduces to the one proposed in Peng (2011), which renders an estimator closely related to that of Portnoy (2003), the special case of Lin et al. (2012) under the right censoring setting.

4.1.3 Computing Algorithm

The estimating equation (4.8), unlike the one in the previous chapter, does not allow for a sequential estimation procedure. Instead, it suggests an iterative algorithm, which is commonly adopted for self-consistent estimators. The procedure is described below:

Step 0. Choose the initial value $\hat{\boldsymbol{\beta}}^{(0)}(\tau), \tau \in (0,1]$. Set $m = 0$.

Step 1. Solve $\boldsymbol{S}_n^*(\hat{\boldsymbol{\beta}}^{(m+1)}; \hat{\boldsymbol{\beta}}^{(m)}, \tau) = 0$ for $\hat{\boldsymbol{\beta}}^{(m+1)}(\tau), \tau \in (0,1]$, where

$$\boldsymbol{S}_n^*(\boldsymbol{\beta}; \boldsymbol{b}, \tau) = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{Z}_i \Big[\tilde{N}_i\{\boldsymbol{Z}_i^T \boldsymbol{\beta}(\tau)\} + \tau \tilde{L}_i(\infty) + \tau \int_{\tau}^{1} \frac{\tilde{L}_i\{\boldsymbol{Z}_i^T \boldsymbol{b}(\tau)\}}{u^2}\, \mathrm{d}u + (1-\tau)\int_{0}^{\tau} \frac{\tilde{R}_i\{\boldsymbol{Z}_i^T \boldsymbol{b}(\tau)\}}{(1-u)^2}\, \mathrm{d}u - \tau \Big]. \tag{4.9}$$

Step 2. Let $m = m + 1$. Repeat Step 1 until certain convergence criteria are met.

At Step 0, we may obtain the initial estimate by ignoring left censoring and fitting model (4.1) using existing quantile regression methods which assume right censoring only, for example, Peng and Huang (2008)'s method.

At Step 1, we may assume $\hat{\boldsymbol{\beta}}(\tau)$ to be a cadlag function that jumps only on a prespecified grid, $\mathscr{G}_{L_n} = \{0 = \tau_0 < \tau_1 < \cdots < \tau_{L_n} < \tau_{L_n+1} = 1\}$. According to model (4.1), we may set $g\{\boldsymbol{Z}_i^T \hat{\boldsymbol{\beta}}^{(m+1)}(\tau_0)\} = 0$, and obtain $\hat{\boldsymbol{\beta}}^{(m+1)}(\tau_j)$ for $j = 1, \cdots, L_n$,

which is the minimizer of the L_1-type convex function

$$l_j(\boldsymbol{h}) = \sum_{i=1}^{n} |I(\delta_i = 1) g^{-1}(X_i) - I(\delta_i = 1)\boldsymbol{h}^T \boldsymbol{Z}_i| + |R^* - \boldsymbol{h}^T \sum_{l=1}^{n} -I(\delta_l = 1)\boldsymbol{Z}_l|$$
$$+ \Big| R^* - \boldsymbol{h}^T \sum_{r=1}^{n} 2\boldsymbol{Z}_r \Big\{ \tau_j \tilde{L}_i(\infty) + \tau_j \sum_{k=j}^{L_n} \tilde{L}_r\{\boldsymbol{Z}_r^T \hat{\boldsymbol{\beta}}^{(m)}(\tau_k)\} \Big(\frac{1}{\tau_k} - \frac{1}{\tau_{k+1}} \Big)$$
$$+ (1-\tau_j) \sum_{k=0}^{j-1} \tilde{R}_r\{\boldsymbol{Z}_r^T \hat{\boldsymbol{\beta}}^{(m)}(\tau_k)\} \Big(\frac{1}{1-\tau_{k+1}} - \frac{1}{\tau_k} \Big) - \tau_j \Big\} \Big|, \tag{4.10}$$

where R^* is a very large number.

Due to censoring on both the lower and upper tails of T, we anticipate that $\boldsymbol{\beta}_0(\tau)$ may be unidentifiable for lower and higher τ's, and therefore, $\hat{\boldsymbol{\beta}}(\tau)$ may not be stably estimated for certain τ ranges, say, $\tau(0, \tau_L)$ and $(\tau_U, 1)$, where $0 < \tau_L < \tau_U < 1$. We can obtain reasonable approximates for τ_L and τ_U based on the rationale that, given a fine grid $\mathscr{G}_L$, $\hat{\boldsymbol{\beta}}(\tau_j)$ and $\hat{\boldsymbol{\beta}}(\tau_{j+1})$ are expected to be close in the identifiable region for $\hat{\boldsymbol{\beta}}(\cdot)$. For example, we may set τ_L to be the smallest j such that $\|\hat{\boldsymbol{\beta}}(\tau_{j+1}) - \hat{\boldsymbol{\beta}}(\tau_j)\| < d$, and set τ_U to be the largest j such that $\|\hat{\boldsymbol{\beta}}(\tau_j) - \hat{\boldsymbol{\beta}}(\tau_{j-1})\| < d$, where d is a pre-specified positive constant. The choice of d can be quite flexible as long as we avoid values that are too small (e.g., less than 10^{-2}) or too large (e.g., greater than 10^2). We set $d = 1$ in our numerical studies.

Despite the fact that $\boldsymbol{\beta}_0(\tau)$ may only be estimable for $\tau \in [\tau_L, \tau_U]$, the whole process of $\hat{\boldsymbol{\beta}}(\tau)$ for $\tau \in (0, 1)$ is needed in the iterative procedure. Having decided the values of τ_L and τ_U, we may employ a simple solution for this difficulty, which shares the spirit of the Last-Observation-Carried-Forward (LOCF) approach. More specifically, we can impute $\hat{\boldsymbol{\beta}}(\tau)$ with $\hat{\boldsymbol{\beta}}(\tau_L)$ for $\tau \in (0, \tau_L)$ and $\hat{\boldsymbol{\beta}}(\tau_U)$ for $\tau \in (\tau_U, 1)$ at Step 1, and carry the imputed $\hat{\boldsymbol{\beta}}(\cdot)$ over to Step 2.

4.1.4 Inferences

To make inference on $\boldsymbol{\beta}_0(\tau)$, we utilize the bootstrap approach (Efron, 1979). First we generate B data sets from boostrapping the original data set, and then estimate $\boldsymbol{\beta}_0(\tau)$ for each of these data sets. We denote the resulting estimates by $\{\hat{\boldsymbol{\beta}}_b^*(\tau)\}_{b=1}^B$. For each fixed τ, we may approximate the variance of $\hat{\boldsymbol{\beta}}(\tau)$ using the sample variance of $\{\hat{\boldsymbol{\beta}}_b^*(\tau)\}_{b=1}^B$, and then construct confidence intervals by normal approximation.

It is often of practical interest to conduct hypothesis testing on the covariate effects. Denote the coefficient corresponding to $Z^{(q)}$ by $\boldsymbol{\beta}_0^{(q)}(\tau)$ for $q = 1, \cdots, p$. Here we provide the procedures for performing two important tests on $\boldsymbol{\beta}_0^{(q)}(\tau)$.

The first test concerns the overall significance of $\boldsymbol{\beta}_0^{(q)}(\tau)$ across $\tau \in [l, u]$, where l and u are pre-specified constants satisfying $\tau_L < l < u < \tau_U$. We may formulate this test as $H_0 : \boldsymbol{\beta}_0^{(q)}(\tau) = 0, \tau \in [l, u]$. A test statistic can be given by $\hat{\eta}_q = \int_l^u \hat{\beta}^{(q)}(v)\Theta(v)dv/(u-l)$, where $\Theta(v)$ is a non-negative weight function and can be properly chosen to emphasize the departure from H_0. Define $\eta_q^* = \int_l^u \hat{\beta}^{*(q)}(v)\Theta(v)dv/(u-l)$. It can be shown that, under H_0, the limiting distribution of $\hat{\eta}_q$ is a normal distribution with mean 0. The standard error, namely SE_q, can be approximated from the bootstrap realizations $\{\eta_{b,q}^*\}_{b=1}^B$, where $\eta_{b,q}^* = \int_l^u \beta_b^{*(q)}(v)\Theta(v)dv/(u-l)$. Subsequently, we may reject H_0 at α level if $|\hat{\eta}_q/SE_q| > Z_{1-\alpha/2}$, where $Z_{1-\alpha/2}$ denotes the Z score for the $1-\alpha/2$ percentile point.

To assess the constancy of $\boldsymbol{\beta}_0^{(q)}(\tau)$ over $[l, u]$, we may formulate the test as $\tilde{H}_0 : \beta_0^{(q)}(\tau) = \rho_0, \tau \in [l, u]$, where ρ_0 is an unknown constant. To test $\tilde{H}_0$, we may adopt the test statistic $\tilde{\Gamma} = \int_l^u \{\hat{\beta}^{(q)}(v) - \hat{\eta}_q\}\tilde{\Theta}(v)\,\mathrm{d}v/(u-l)$, where $\tilde{\Theta}(v)$ is a non-negative weight function. Under the observed data, we can approximate the distribution of $\tilde{\Gamma}$ by $\{\tilde{\Gamma}_b^*\}_{b=1}^B$ obtained from the bootstrapped sample, where $\tilde{\Gamma}_b^* = n^{1/2}\int_l^u \{\hat{\beta}_b^{*(q)}(v) - \eta_{b,q}^*\}\tilde{\Theta}(v)\,\mathrm{d}v/(u-l)$. This fact naturally leads to a Wald-type of percentile-based test for $\tilde{H}_0$, the rejection of which indicating varying covariate effect for $Z^{(q)}$.

4.2 Numerical Studies

4.2.1 Simulation Studies

We assessed the finite-sample performance of the proposed method through Monte-Carlo simulations. First we considered a linear model with i.i.d errors, following Lin et al. (2012):

$$T = b_1 Z_1 + \epsilon,$$

where $\epsilon \sim N(0,1)$, $Z_1 \sim Unif(0,1)$, and ϵ and Z_1 are independent. For the censoring mechanism, we considered two cases: (I) the censoring times are unconditionally independent with the survival time, and (II) the censoring times are dependent on the covariates, and conditionally independent with the survival time given the covariates. More specifically, in case (I) we set $L \sim Unif(-7,3)$ and $U \sim L + Unif(4.5, 11.5)$. In case (II) we set $L \sim Exp(4Z_1 + 4)$ and $U \sim L + Unif(1.5, 8)$. For both case (I) and case (II), we set $b_1 = 5$, which leads to 20% right censoring and 10% left censoring.

We also examined our method under an AFT model with heteroscedastic errors:

$$\log(T) = b_1 Z_1 + b_2 Z_2 \xi + \epsilon,$$

where ϵ follows the extreme distribution, Z_1 follows $Unif(0,1)$, Z_2 follows $Bernoulli(0.5)$, and ξ follows $Exp(1)$. L was generated from $Unif(0, 0.25)$, and U was generated from $L + Unif(0.1 \cdot I(Z_2 = 1), 4.5)$. Under this set-up, namely case (III), model (1) holds with $g(\cdot) = \exp(\cdot)$ and $\boldsymbol{\beta}_0(\tau) = \{\beta_0^{(0)}(\tau), \beta_0^{(1)}(\tau), \beta_0^{(2)}(\tau)\}^T$, where $\beta_0^{(0)}(\tau) = Q_\epsilon(\tau)$, $\beta_0^{(1)}(\tau) = b_1$, and $\beta_0^{(2)}(\tau) = Q_{b_2\xi+\epsilon}(\tau) - \beta_0^{(0)}(\tau)$. We set $b_1 = 0$ and $b_2 = -0.5$ to maintain 15% left censoring and 15% right censoring.

Under each scenario, we generated 1000 data sets of sample size $n = 200$ and adopted an equally spaced grid of τ of size 0.01. For the estimation, the convergence rates are 100%, 98.7% and 99.8% for cases (I), (II) and (III) respectively, reached after

4, 7.7 and 4.7 iterations on average. We set $B = 200$ in the bootstrap procedure.

Results for cases (I) and (II) are summarized in Table 4.1, and results for case (III) are presented in Table 4.2. In both tables, we report the biases (Bias), empirical standard deviations (EmpSD), average estimated standard deviations from resampling (AvgSD) for $\hat{\boldsymbol{\beta}}(\tau)$, and coverage rates of 95% confidence intervals of $\boldsymbol{\beta}_0(\tau)$ using normal approximation, with τ ranging from 0.2 to 0.8. It can be seen that the estimation and inference procedures perform well. The biases are small, the resampling-based standard error estimates generally agree well with the empirical ones, and the coverage rates are close to the nominal level of 95%. Our results for cases (I) and (II) agree well with that reported in Lin et al. (2012).

Table 4.1: Simulation Results under Models with Constant Covariate Effects. Bias: absolute biases; EmpSD: empirical standard deviations; AvgSD: average estimated resampling-based standard deviations; Cov95: coverage rates of 95% Wald confidence intervals.

Case I

	$\hat{\beta}^{(0)}$				$\hat{\beta}^{(1)}$			
τ	Bias	EmpSD	AvgSD	Cov95	Bias	EmpSD	AvgSD	Cov95
0.2	0.000	0.246	0.266	0.939	0.012	0.425	0.453	0.948
0.3	0.017	0.218	0.238	0.938	-0.010	0.377	0.413	0.946
0.4	0.011	0.205	0.224	0.947	-0.003	0.357	0.391	0.953
0.5	0.018	0.202	0.217	0.943	-0.020	0.359	0.385	0.944
0.6	0.008	0.202	0.218	0.947	-0.004	0.362	0.392	0.948
0.7	0.009	0.219	0.229	0.934	-0.001	0.393	0.413	0.939
0.8	0.008	0.229	0.248	0.935	-0.006	0.421	0.452	0.941

Case II

	$\hat{\beta}^{(0)}$				$\hat{\beta}^{(1)}$			
τ	Bias	EmpSD	AvgSD	Cov95	Bias	EmpSD	AvgSD	Cov95
0.2	-0.024	0.305	0.381	0.958	0.037	0.503	0.587	0.956
0.3	0.002	0.231	0.267	0.947	0.010	0.404	0.456	0.952
0.4	-0.002	0.198	0.220	0.943	0.014	0.364	0.394	0.954
0.5	-0.002	0.188	0.205	0.949	0.013	0.356	0.381	0.942
0.6	0.000	0.190	0.203	0.942	0.005	0.361	0.383	0.942
0.7	0.000	0.196	0.210	0.937	0.015	0.376	0.405	0.948
0.8	0.005	0.216	0.231	0.943	0.000	0.420	0.454	0.943

Table 4.2: Simulation Results under Model with Varying Covariate Effects. Bias: absolute biases; EmpSD: empirical standard deviations; AvgSD: average estimated resampling-based standard deviations; Cov95: coverage rates of 95% Wald confidence intervals.

	Case III			
	$\hat{\beta}^{(0)}$			
τ	Bias	EmpSD	AvgSD	Cov95
0.2	-0.018	0.456	0.464	0.951
0.3	-0.012	0.317	0.333	0.935
0.4	0.000	0.265	0.289	0.954
0.5	-0.005	0.234	0.261	0.958
0.6	-0.007	0.219	0.250	0.959
0.7	-0.001	0.220	0.244	0.948
0.8	-0.004	0.222	0.256	0.954
	$\hat{\beta}^{(1)}$			
τ	Bias	EmpSD	AvgSD	Cov95
0.2	0.021	0.700	0.724	0.959
0.3	0.005	0.499	0.525	0.951
0.4	-0.004	0.420	0.452	0.949
0.5	-0.004	0.365	0.413	0.957
0.6	0.003	0.344	0.392	0.962
0.7	-0.001	0.344	0.383	0.957
0.8	0.005	0.349	0.399	0.962
	$\hat{\beta}^{(2)}$			
τ	Bias	EmpSD	AvgSD	Cov95
0.2	-0.008	0.369	0.421	0.959
0.3	0.010	0.279	0.306	0.961
0.4	0.003	0.237	0.258	0.949
0.5	0.011	0.220	0.235	0.948
0.6	0.007	0.211	0.223	0.946
0.7	0.006	0.204	0.219	0.954
0.8	-0.001	0.203	0.227	0.957

4.2.2 Data Analysis

In this section we apply the proposed methods to the CFFPR study, with the same objective as in Chapter 3, that is, we investigate the association between onset ages of the first detected PA infection and several risk factors in young CF patients who were diagnosed by age 10 years. The same set of covariates were considered: gender, diagnosis mode and diagnosis year. Diagnosis mode was classified into four categories: diagnosis at birth due to meconium ileus (MI), diagnosis shortly after birth by neonatal/prenatal screening (SCR), diagnosis at variable ages because of family history (FH), and diagnosis at variable ages from various symptoms (SYMP) other than MI. Diagnosis year was categorized into three periods, 1986-1989 (dx86), 1990-1993 (dx90), and 1994-2000 (dx94). The reference group was chosen to be boy patients diagnosed with SYMP between 1986 and 1989. The analyzed sample contains 11,179 patients, among whom 23.7% were left censored and 16.2% right censored.

We fit model (4.1) using the identity link. Unlike in Chapter 3, here we do not utilize the information on L, although it is always observed in the CFFPR study. By the proposed methods we were able to obtain $\hat{\boldsymbol{\beta}}(\tau)$ for $\tau \in [0.2, 0.9]$. Figure 4.1 depicts $\hat{\boldsymbol{\beta}}(\tau)$ in bold lines and their 95% Wald confidence intervals in shades for $\tau \in [0.2, 0.75]$. The naive estimates from Peng and Huang (2008)'s approach, ignoring all left censored observations, are plotted in dotted lines in the same figure. The obvious departure displayed by the naive estimates from the proposed estimates suggests the importance of properly handling left censoring. The estimated intercept (Figure 4.1, panel A) indicates that about 20% of boy patients with CF diagnosed during 1986-1989 by SYMP acquired their first PA infection by age 2 years and approximately 75% of them had their first PA infection by age 11 years.

From Figure 4.1 we observe that, the gender effect (panel A) and the effect of "dx90" (panel G) appear to be always negative for $\tau \in [0.2, 0.75]$. The effects of MI diagnosis (panel C) and "dx94" (panel F) are also negative for most of the quantile

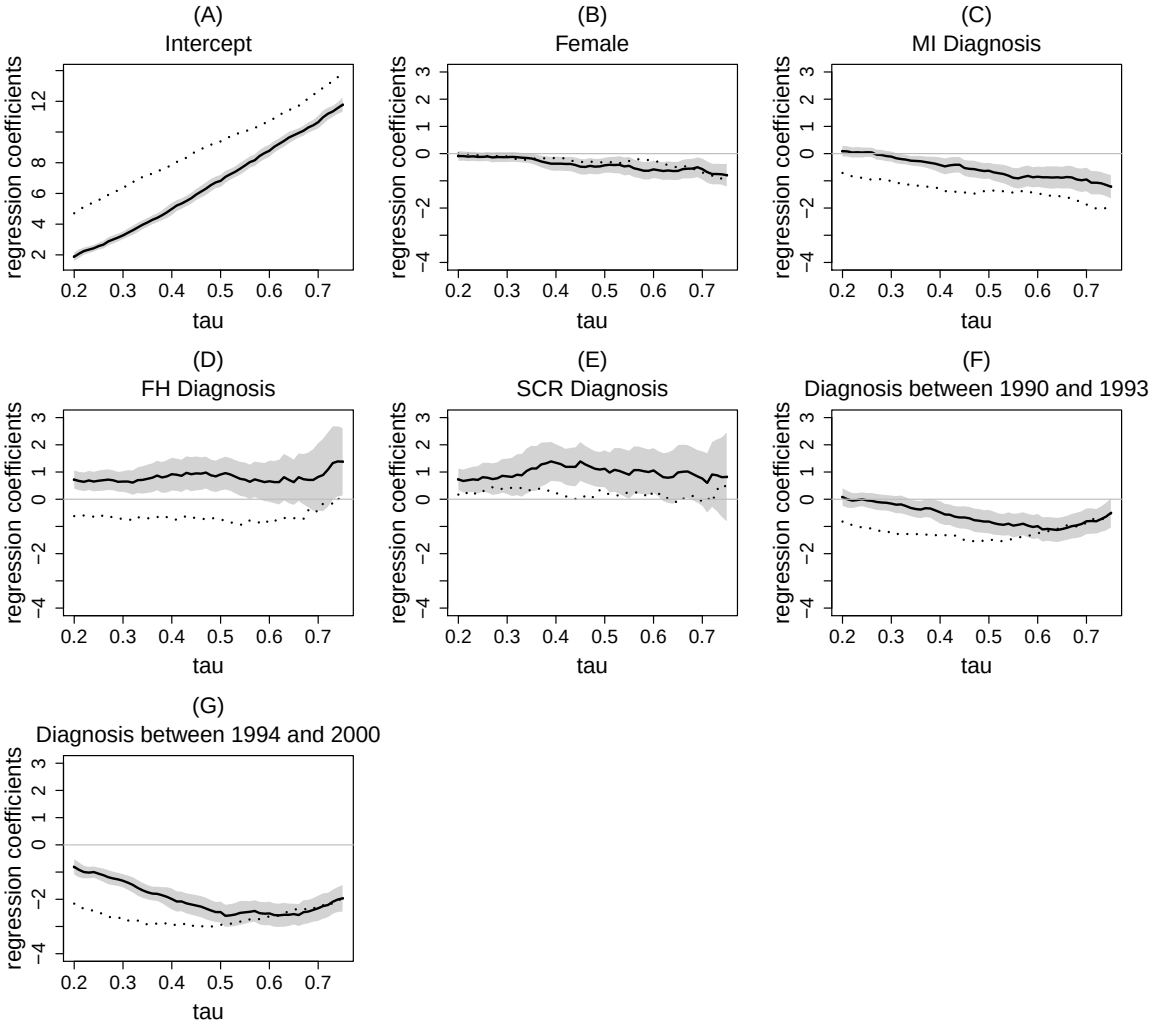

Figure 4.1: Coefficient Estimates (Bold Solid Lines) and 95% Pointwise Confidence Intervals (Shaded Areas) from the Proposed Method, in Contrast with Naive Coefficient Estimates Ignoring Left Censored Observations (Dotted Lines).

range. In contrast, the effects of FH diagnosis and SCR diagnosis (panels D and E) seem to be always positive for $\tau \in [0.2, 0.75]$. In addition, we can see that the effects of gender, MI diagnosis, "dx90" and "dx94" exhibit non-constant patterns with increasing magnitudes at higher τs.

We conducted formal tests on the significance of the covariates based on the average quantile effects over $\tau \in [0.2, 0.75]$. These tests confirmed the accelerating effects of female gender, MI diagnosis, "dx90" and "dx94" (p-values=0.03, 0.01, 0.01, and < 0.001, respectively), and the delaying effect of SCR diagnosis (p-value=0.04). The

effect of FH diagnosis is marginally significant with a p-value of 0.06. These results indicate that, girl patients, newer diagnosis cohorts (patients diagnosed between 1994 and 2000, or between 1990 and 1993), and patients diagnosed by MI tend to acquire their first PA infection earlier, while the situation may be opposite for patients diagnosed by FH and SCR.

The varying patterns of the effects of gender, MI diagnosis, "dx90" and "dx94" were also confirmed by statistical tests (all p-values< 0.01). The delaying effects of these covariates are even more pronounced at higher quantiles, which correspond to patients who acquired their first PA infection at older ages. Such inhomogeneity in the covariate effects across quantiles would not have been captured by a traditional AFT model or Cox model.

In Figure 4.2, we plotted the estimated quantiles of age at first PA infection (in bold lines) along with the 95% Wald confidence intervals (in shades) for 7 subgroups of patients. Panel A corresponds to the reference group, boy patients who were diagnosed by SYMP between 1986 and 1989. Panels B through G represent groups that differed from the reference group by each covariate: gender (panel B), diagnosis mode (panels B, C, D), and diagnosis year (panels E, F). Also shown in Figure 4.2 are the estimated quantiles from Chang and Yang (1987)'s nonparametric method (in dot-dash lines), given that all covariates are binary. We observe that, the estimates from the proposed methods are generally close to those from Chang and Yang (1987)'s method, which suggests the validity of our model.

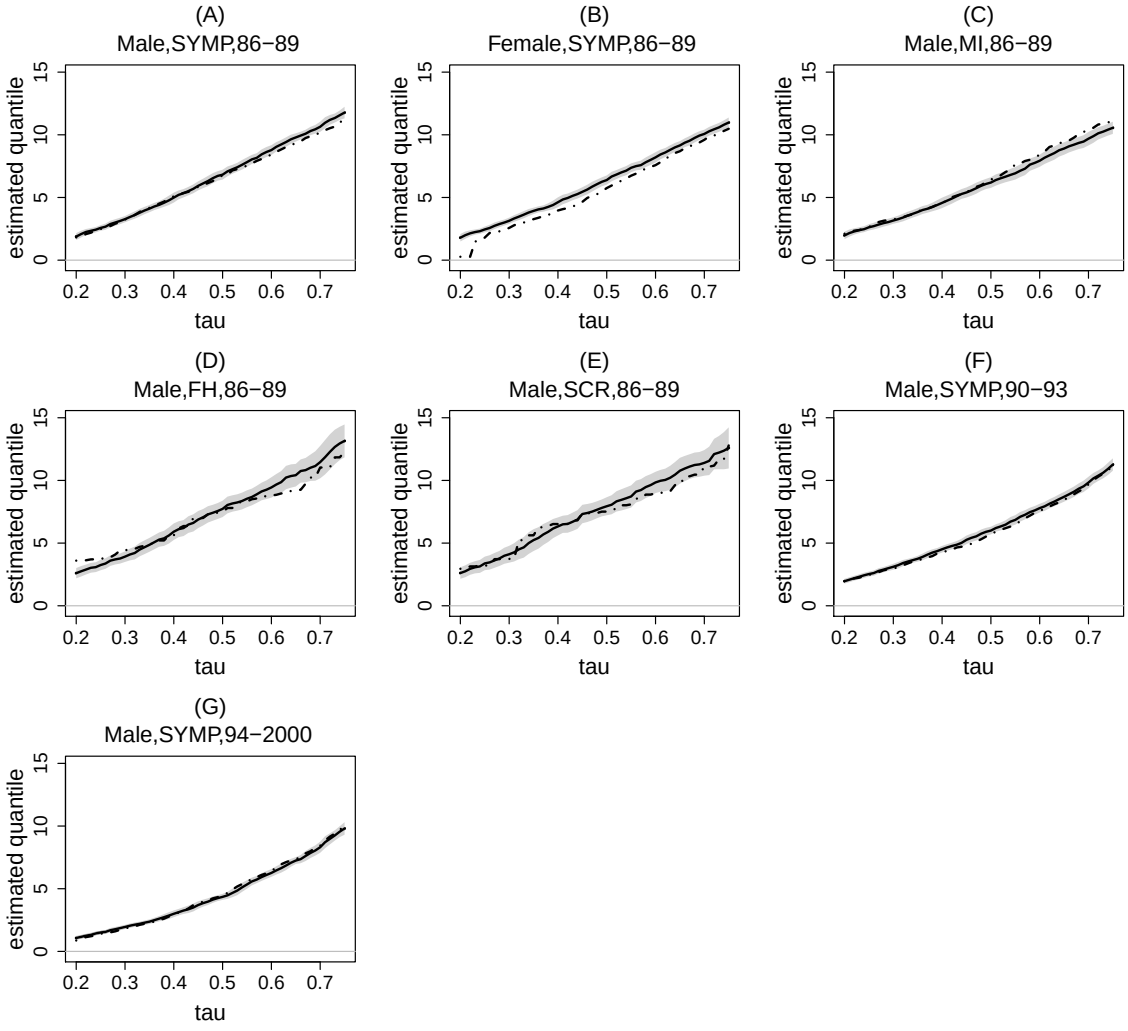

Figure 4.2: Estimated Quantiles of Age at First PA infection (Bold Solid Lines) and 95% Pointwise Confidence Intervals (Shaded Areas) from the Proposed Method, in Contrast with Estimated Quantiles from Chang and Yang (1987) (Dot-Dash Lines).

Chapter 5

Summary and Future Work

5.1 Summary

In this dissertation we investigated two complex censoring schemes, dependent censoring and double censoring. We developed quantile regression methods which can accommodate these complications.

We first studied survival data subject to dependent censoring, focusing on the scenario where the marginal quantiles of survival outcome is of interest. We proposed unbiased estimating equations with efficient and stable algorithms, and established uniform consistency and weak convergence of the proposed estimators based on theory in empirical processes and stochastic differential equations. We conducted extensive simulation studies, which demonstrated satisfactory performance of the proposed quantile regression procedure with moderate sample size. We applied the proposed method to the WASID study to demonstrate its practical utility.

We then developed a new quantile regression method for doubly censored data with known left censoring times (while the right censoring times may not always be observable). We also proposed conditional inference to address the special identifiability issues attached to the double censoring setting. The conditional inference can be easily adapted to handle left truncation. The proposed inference procedures well utilize the embedded martingale structure, which facilitates the establishment of the asymptotic properties for the proposed estimators. In addition, our methods can be conveniently implemented in standard software. Simulation studies show satisfactory finite-sample performance of our method. An application to the CFFPR data revealed interesting scientific findings.

We further considered a more challenging double censoring problem, where the random left censoring times are not required to be always observed. As there is no clear martingale structure attached to this scenario, we developed a self-consistent estimating equation along with an efficient iterative algorithm. Our simulation studies suggest the validity of the proposed procedure. We also conducted data analysis for

the CFFPR study using this new method.

5.2 Future Work

We plan to complete the ongoing work on doubly censored data in the near future. First, we will establish the asymptotic properties, including uniform consistency and weak convergence, of the proposed estimators. We will further polish our algorithm to accommodate potential identifiability issues arising from double censoring, and conduct additional simulation studies to test the new proposals. Finally, we plan to apply the proposed methods to data sets other than the CFFPR data.

In what follows we describe some possible topics for future work. One direction is to extend our work on dependent censoring to more complicated scenarios with a mixture of competing risks and semi-competing risks. For example, in the WASID study, multiple events were of interest. The primary endpoint is a terminating endpoint, since the follow-up of a patient would be stopped if any single component of the primary endpoint was observed. A secondary event is nonfatal myocardial infarction, which is a nonterminating event since its occurrence did not prevent subsequent observations of other events. This poses a scenario consisting of both competing risks and semi-competing risks, which is often encountered in other biomedical studies. It would be very desirable to develop sensible quantile regression methods for such a general case while properly disentangling the complex censoring relationship and handling the dependence among the multiple events.

It is also worthwhile to notice that, in both projects presented in this proposal, we employed a grid-based estimation procedure, for which sufficiently small grid size is warranted for nice asymptotic properties. It may be interesting to develop a grid-free approach, following Huang (2010) for example. This may also merit future research.

Bibliography

Andersen, P., Borgan, Ø., Gill, R., and Keiding, N. (1998). *Statistical Models Based on Counting Processes (2nd ed.).* New York: Springer-Verlag.

Barrodale, I. and Roberts, F. (1974). Solution of an overdetermined system of equations in the l_1 norm. *Communications of the ACM* **17,** 319–320.

Cai, T. and Cheng, S. (2004). Semiparametric regression analysis for doubly censored data. *Biometrika* **91,** 277–290.

Campbell, P. r. and White, T. (2005). Newborn screening for cystic fibrosis: an opportunity to improve care and outcomes. *The Journal of Pediatrics* **147,** S2–S5.

Carey, J. R., Liedo, P., Orozco, D., Tatar, M., and Vaupel, J. W. (1995). A male-female longevity paradox in medfly cohorts. *The Journal of Animal Ecology* **64,** 107–116.

Chang, M. (1990). Weak Convergence of a Self-Consistent Estimator of the Survival Function with Doubly Censored Data. *The Annals of Statistics* **18,** 391–404.

Chang, M. N. and Yang, G. L. (1987). Strong consistency of a nonparametric estimator of the survival function with doubly censored data. *The Annals of Statistics* **15,** 1536–1547.

Chen, Y.-H. (2010). Semiparametric marginal regression analysis for dependent com-

peting risks under an assumed copula. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)* **72,** 235–251.

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

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.

Dickson, E., Grambsch, P., Fleming, T., Fisher, L., and Langworthy, A. (1989). Prognosis in primary biliary cirrhosis: Model for decision making. *Hepatology* **10,** 1–7.

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,** 496–509.

Fine, J. P., Yan, J., and Kosorok, M. R. (2004). Temporal process regression. *Biometrika* **91,** 683–703.

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

Gehan, E. A. (1965). A generalized two-sample wilcoxon test for doubly censored data. *Biometrika* **52,** 650–653.

Genest, C. (1987). Frank's family of bivariate distributions. *Biometrika* **74,** 549–555.

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

Gray, R. (1988). A class of k-sample tests for comparing the cumulative incidence of a competing risk. *The Annals of Statistics* **16,** 1141–1154.

Grosse, S., Rosenfeld, M., Devine, O., Lai, H., and PM, F. (2006). Potential impact of newborn screening for cystic fibrosis on child survival: a systematic review and analysis. *The Journal of Pediatrics* **149,** 362–366.

Gu, M. and Zhang, C.-H. (1993). Asymptotic Properties of Self-Consistent Estimators Based on Doubly Censored Data. *The Annals of Statistics* **21,** 611–624.

Honore, B., Khan, S., and Powell, J. (2002). Quantile regression under random censoring. *Journal of Econometrics* **109,** 67–105.

Huang, X. and Zhang, N. (2008). Regression survival analysis with an assumed copula for dependent censoring: a Sensitivity analysis approach. *Biometrics* **64,** 1090–1099.

Huang, Y. (2010). Quantile calculus and censored regression. *The Annals of Statistics* **38,** 1607–1637.

Jensen, G., Torp-Pedersen, C., Hildebrandt, P., Kober, L., Nielsen, F., Melchior, T., Joen, T., and Andersen, P. (1997). Does in-hosipital fibrillation affect prognosis after myocardial infarction? *European Heart Journal* **18,** 919–924.

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

Kalbfleisch, J. and Prentice, R. (2002). *The Statistical Analysis of Failure Time Data, 2nd ed.* Wiley.

Kaplan, E. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association* **53,** 457–481.

Kaslow, R., Ostrow, D., Detels, R., Phair, J., Polk, B., and Rinaldo, C. (1987). The multicenter aids cohort study: rationale, organization and selected characteristics of the participants. *American Journal of Epidemiology* **126,** 310–318.

Klein, J. and Moeschberger, M. (1988). Bounds on net survival probabilities for dependent competing risks. *Biometrics* **44,** 529–538.

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

Kosorok, M., Zeng, L., West, S., Rock, M., Splaingard, M., Laxova, A., Green, C., Collins, J., and Farrell, P. (2001). Acceleration of lung disease in children with cystic fibrosis after Pseudomonas aeruginosa acquisition. *Pediatric Pulmonology* **32,** 277–287.

Lai, H., Cheng, Y., Cho, H., Kosorok, M., and Farrell, P. (2004). Association between initial disease presentation, lung disease outcomes, and survival in patients with cystic fibrosis. *American Journal of Epidemiology* **159,** 537–546.

Lin, D. (1997). Non-parametric inference for cumulative incidence functions in competing risks studies. *Statistics in Medicine* **16,** 901–910.

Lin, G., He, X., and Pornoy, S. (2012). Quantile regression with doubly censored data. *Computational Statistics and Data Analysis* **56,** 797–812.

Link, W. (1989). A model for informative censoring. *Journal of the American Statistical Association* **84,** 749–752.

Mantel, N. (1966). Evaluation of survival data and two new rank order statistics arising in its consideration. *Cancer Chemother Rep.* **50,** 163–170.

Mantel, N. (1967). Ranking procedures for arbitrarily restricted observation. *Biometrics* **23,** 65–78.

Mykland, P. and Ren, J.-J. (1996). Algorithms for Computing Self-Consistent and Maximum Likelihood Estimators with DoublyCensored Data. *The Annals of Statistics* **24,** 1740–1764.

Neocleous, T., Branden, K. V., and Portnoy, S. (2006). Correction to censored regression quantiles by S. Portnoy. *Journal of the American Statistical Association* **101,** 860–861.

Peng, L. (2011). Self-consistent estimation of censored quantile regression. *Journal of Multivariate Analysis. doi10.1016/j.jmva.2011.10.005* .

Peng, L. and Huang, Y. (2008). Survival analysis with quantile regression models. *Journal of the American Statistical Association* **103,** 637–649.

Pepe, M. Inference for events with dependent risks in multi- ple endpoint studies. *Journal of the American Statistical Association* .

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

Portnoy, S. (2003). Censored regression quantiles. *Journal of the American Statistical Association* **98,** 1001–1012.

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.

Ren, J.-J. (2003). Regression m-estimators with non-i.i.d doubly censored data. *The Annals of Statistics* **31,** 1186–1219.

Ren, J.-J. (2008). Weighted empirical likelihood in some two-sample semiparametric models with various types of censored data. *The Annals of Statistics* **36,** 147–166.

Ren, J.-J. and Gu, M. (1997). Regression m-estimators with doubly censored data. *The Annals of Statistics* **25,** 2638–2664.

Samuleson, S. (1989). Asymptotic theory for nonparametric estimators from doubly censored data. *Scandinavian Journal of Statistics* **16,** 1–21.

Scharfstein, D. and Robins, J. (2002). Estimation of the failure time distribution in the presence of informative censoring. *Biometrika* **89,** 617–634.

Scharfstein, D., Robins, J., Eddings, W., and Rotnitzky, A. (2001). Inference in Randomized Studies with Informative Censoring and Discrete Time-to-Event Endpoints. *Biometrics* **57,** 404–413.

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

Slud, E. and Rubinstein, L. (1983). Dependent competing risks and summary survival curves. *Biometrika* **70,** 643–649.

Thorogood, J., Persijn, G., Schreuder, G., D'amaro, J., Zantvoort, F., Van Houwelingen, J., and Van Rood, J. (1990). The effect of hla matching on kidney graft survival in separate posttransplantation intervals. *Transplantation* **50,** 146–150.

Tsai, W., Jewell, N., and Wang, M. (1987). A note on the product-limit estimator under right censoring and left truncation. *Biometrika* **74,** 883–886.

Tsai, W.-Y. and Crowley, J. (1985). A Large Sample Study of Generalized Maximum Likelihood Estimators from Incomplete Data Via Self-Consistency. *The Annals of Statistics* **13,** 1317–1334.

Tsiatis, A. (1975). A nonidentifiability aspect of the problem of competing risks. *Proceedings of the National Academy of Sciences USA* **72,** 20–22.

Turnbull, B. W. (1974). Nonparametric estimation of a survivorship function with doubly censored data. *Journal of the American Statistical Association* **69,** 169–173.

Verweij, P. and Van Houwelingen, H. (1995). Time-dependent effects of fixed covariates in cox regression. *Biometrics* **51,** 1550–1556.

Wang, H. and Wang, L. (2009). Locally weighted censored quantile regression. *Journal of the American Statistical Association* **104,** 1117–1128.

Yan, J., Cheng, Y., Fine, J. P., and Lai, H. J. (2009). Uncovering symptom progression history from disease registry data with application to young cystic fibrosis patients. *Biometrics* **66,** 594–602.

Yang, S. (1999). Censored median regression using weighted empirical survival and hazard functions. *Journal of the American Statistical Association* **94,** 137–145.

Ying, Z., Jung, S., and Wei, L. (1995). Survival analysis with median regression models. *Journal of the American Statistical Association* **90,** 178–184.

Zhan, Y. and Wellner, J. (1995). Double censoring: characterization and computation of the nonparametric maximum likelihood estimator. Technical Report 292, Dept. Statistics, Univ. Washington, Seattle.

Zhang, C.-H. and Li, X. (1996). Linear regression with doubly censored data. *The Annals of Statistics* **24,** 2720–2743.

Zheng, M. and Klein, J. (1995). Estimates of marginal survival for dependent competing risks based on an assumed copula. *Biometrika* **82,** 127–138.