Regression analysis of survival data with covariates subject to censoring

Yu, Tonghui

2018

Yu, T. (2018). Regression analysis of survival data with covariates subject to censoring.
Doctoral thesis, Nanyang Technological University, Singapore.

<https://hdl.handle.net/10356/89713>

<https://doi.org/10.32657/10220/47144>

Regression Analysis of Survival Data with Covariates Subject to Censoring

Tonghui Yu

DIVISION OF MATHEMATICAL SCIENCES
SCHOOL OF PHYSICAL AND MATHEMATICAL
SCIENCES

2018

Regression Analysis of Survival Data with Covariates Subject to Censoring

Tonghui Yu

Division of Mathematical Sciences
School of Physical and Mathematical Sciences

A thesis submitted to Nanyang Technological University
in partial fulfillment of the requirement for the degree of
Doctor of Philosophy in Mathematical Sciences

2018

Abstract

In this thesis, we concern about some issues in survival data with censored covariates.

In the first part, we explore the conditional modeling of the semi-competing risks data, where individuals are likely to experience two types of events: non-terminal and terminal. It is often of interest in practice to predict the terminal event based on the progression of the non-terminal event. It endeavors to aggregate the censored non-terminal event time with other risk factors for better modeling the terminal event time. We propose a new semiparametric model for the terminal event time based on proportional hazards regression conditioning on the non-terminal event time. The model allows for a covariate subject to dependent and independent censoring on the hazard of the terminal event.

In the second part, we propose a quantile regression model for survival data with covariates subject to limits of detection. A novel multiple imputation approach based on quantile regression for the censored covariates is developed to estimate model parameters. The proposed method extends the existing work based on an AFT model to quantile regression. Thus it possesses more flexibility by relaxing stringent constraints on error distribution used in the AFT model and allowing the effects of covariates varying across different quantile levels of the survival distribution.

We develop two-stage estimation procedures in both parts and establish theoretical properties of the proposed estimators. The finite sample performances are demonstrated via extensive simulation studies. Real data applications illustrate the effectiveness of the proposed methods.

Acknowledgements

This dissertation could have never been completed without the help of these people during my Ph.D. study at Nanyang Technological University.

First, I would like to give my deepest thank to my supervisor, prof. Liming Xiang, for her continuous support and excellent guidance over these four years. When I started my Ph.D. study four years ago, I had no idea how to do research and even I had no background in statistics. She guided me patiently how to think creatively, how to find significant problems and how to write an article clearly and logically. I benefit so much from her constructive suggestions and profound insight into statistics. Without her, this thesis would not be started and completed. I also really appreciate her for inviting me to speak at the conference *CMStatistics2017* in the UK.

I sincerely appreciate NTU for offering me the research scholarship during my graduate studies. I am so grateful to prof. Chan Song Heng and the Division of Mathematical Science for providing me financial support twice to attend the international conferences in the UK and US, respectively. Many thanks go to prof. Guangming Pan, prof. Guohua Wu, prof. Chaoping Xing and other faculty and staff members in the Division of Mathematical Science for providing various courses, seminars as well as all support and assistance. I want to express my gratitude to the three reviewers for their valuable comments and suggestions, which have helped me a lot to improve the thesis.

I want to express my love to my parents who always stand firmly behind me and give me unconditional warmth, love and encouragement. Finally, thank my friends in the statistics group for their selfless help and friendship.

Contents

Abstract	i
Acknowledgements	iii
1 Introduction	1
1.1 Survival Analysis	1
1.1.1 Time-to-event Data	1
1.1.2 Basic Quantities	2
1.2 Censoring	3
1.2.1 Right Censoring and Left Censoring	3
1.2.2 Dependent Censoring	4
1.3 Semi-parametric Models	5
1.3.1 Cox Proportional Hazards Model	6
1.3.2 Accelerated Failure Time Model	6
1.3.3 Censored Quantile Regression	7
1.4 Motivation	9
1.4.1 Breast Cancer Data	9
1.4.2 GenIMS Data	10

1.5	Thesis Outline	11
2	Conditional Modeling of Survival Data with Semi-competing Risks	13
2.1	Introduction	13
2.2	Model	15
2.3	Estimation Procedure and Implementation	18
2.3.1	First-stage of the Estimation for γ	20
2.3.2	Kaplan-Meier-induced Estimator for Q	21
2.3.3	Weighted Breslow-type Estimator for λ_0	23
2.3.4	Second-stage of the Estimation for (α, β)	25
2.3.5	Implementation	26
2.4	Theoretical Justification	27
2.4.1	Asymptotic Properties of the First-stage Estimator	28
2.4.2	Asymptotic Properties of the Second-stage Estimator	32
2.5	Simulation	34
2.6	Real Data Analysis: Breast Cancer Data	43
2.7	Summary	48
2.8	Appendix	50
2.8.1	Proofs of Lemmas 2.1-2.2 and Theorem 2.1	50
2.8.2	Proofs of Lemma 2.3 and Theorem 2.2	58
2.8.3	Proof of Theorem 2.3	63
2.8.4	Proof of Theorem 2.4	69
3	Quantile Regression for Survival Data with Covariates Subject to Detection Limits	79
3.1	Introduction	79
3.2	Methodology	80

3.2.1	Censored Quantile Regression with Fully Observed Covariates	81
3.2.2	Censored Quantile Regression with Univariate Left- censored Covariate	82
3.2.3	Censored Quantile Regression with Univariate Right- censored Covariate	85
3.2.4	Censored Quantile Regression with Multiple Censored Covariates	85
3.3	Theoretical Justification	87
3.4	Simulation Results	89
3.5	Real Data Analysis: GenIMS Data	95
3.6	Summary	103
3.7	Appendix	103
3.7.1	Proof of Theorem 3.1 (Uniform consistency)	105
3.7.2	Proof of Theorem 3.2 (Asymptotic normality)	111
4	Discussion and Future Research	115
	References	119

List of Figures

2.1	Estimation results for true baseline cumulative hazard function $\Lambda_0(t) = t^2$ and survival function $S_0(t) = \exp(-t^2)$ under Scenario I with $n = 100, 200, 400$. The black solid lines are the estimates, the red dot-dashed lines are the true ones, and the black dashed lines are the 95% upper/lower bounds over 200 replicates.	41
2.2	Estimation of the baseline cumulative hazard function (solid curve) together with its 95% upper and lower bounds (dashed curves) for the breast cancer data.	44
2.3	Goodness of fit of the Cox proportional hazards model regarding binary covariates for the breast cancer data.	45
2.4	Goodness of fit of the Cox proportional hazards model regarding continuous covariates for the breast cancer data.	46
3.1	Comparisons of solution paths for estimated regression coefficients (solid curves) along with their 95% upper and lower bounds (dashed curves) using methods “MI-logY” and “CC” for the GenIMS data.	100

- 3.2 Comparisons of solution paths for estimated regression coefficients (solid curves) along with their 95% upper and lower bounds (dashed curves) using methods “MI-logY*” and “CC” for the GenIMS data. 101
- 3.3 Comparisons of solution paths for estimated regression coefficients (solid curves) along with their 95% upper and lower bounds (dashed curves) using methods “MIs” and “CC” for the GenIMS data. 102

List of Tables

2.1	Simulation results are shown for Scenario I. The censoring time is generated from $C \sim U(1, 3)$, gaining the censoring rates of $32 \sim 55\%$ for V and $20 \sim 36\%$ for T . The total observed rate for the pair (V, T) is around 42%	36
2.2	Simulation results are shown for Scenario II. The censoring time is generated from $C \sim U(1, 3)$, gaining the censoring rates of $27 \sim 44\%$ for V and $33 \sim 52\%$ for T . The total observed rate for the pair (V, T) is around 41%	37
2.3	Simulation results are shown for Scenario III. The censoring time is generated from $C \sim U(0, 3)$, gaining the censoring rates of $38 \sim 58\%$ for V and $49 \sim 66\%$ for T . The total observed rate for the pair (V, T) is around 29%	38
2.4	Simulation results are shown for Scenario IV. The censoring time is generated from $C \sim U(1, 3)$, gaining the censoring rates of $34 \sim 58\%$ for V and $42 \sim 60\%$ for T . The total observed rate for the pair (V, T) is around 36%	39

2.5	Simulation results are shown for Scenario V. The censoring time is generated from $C \sim U(1, 3)$, gaining the censoring rates of $25 \sim 42\%$ for V and $25 \sim 42\%$ for T . The total observed rate for the pair (V, T) is around 40%.	40
2.6	Estimation results of conditional modeling for the breast cancer data.	43
3.1	Simulation results based on 500 simulation runs for the proposed method with various choices of $\mathbf{W}_i$ in the imputation model in comparison with the “full”, “CC” and “fill-d” methods, where CR.V1=20%, CR.V2=20% and CR.T=40%. . .	92
3.2	Simulation results based on 500 simulation runs for the proposed method with various choices of $\mathbf{W}_i$ in the imputation model in comparison with the “full”, “CC” and “fill-d” methods, where CR.V1=20%, CR.V2=40% and CR.T=40%. . .	93
3.3	Simulation results based on 500 simulation runs for the proposed method with various choices of $\mathbf{W}_i$ in the imputation model in comparison with the “full”, “CC” and “fill-d” methods, where CR.V1=40%, CR.V2=40% and CR.T=40%. . .	94
3.4	Estimation results obtained from the “CC”, “MI-logY”, “MI-logY*” and “MIs” methods for the GenIMS data.	99

Chapter 1

Introduction

1.1 Survival Analysis

1.1.1 Time-to-event Data

Survival analysis in statistical terminology represents a collection of data analysis approaches for which the response of interest is the time from an origin until a specified event or end-point. In biomedical studies, the time origin is the entry of an individual into a study, and the end-point can be a relapse from remission, development of a disease, distant metastasis of cancer or death. The time to event is also known as survival time. In some biomedical studies, the phrase overall survival (OS) refers to the time to death. A clear fact is that the survival time is non-negative.

Assume that the collected data from an individual has the exact event time, say T . In this thesis, we concentrate on the more common case in which the distribution function of the nonnegative random variable T is continuous.

1.1.2 Basic Quantities

There are two basic functions in the statistical subject, namely, probability density function (PDF) and cumulative distribution function (CDF). The CDF of T , defined by $F(t) = Pr(T \leq t)$, is right continuous. The PDF of T , which is specified as

$$f(t) = \frac{dF(t)}{dt} = \lim_{h \rightarrow 0} \frac{Pr(T \in [t, t+h))}{h},$$

gives the rate of occurrence of event at t .

We now address two functions playing important roles in survival analysis, that is, survival function and hazard function. The survival function of T , denoted by $S(t)$, specifies the probability that an individual survives longer than some specified time t . The interrelationship among the survival function, CDF and PDF are described as follows:

$$S(t) = Pr(T \geq t) = \int_t^{\infty} f(x)dx = 1 - F(t).$$

One can see that $S(t)$ is non-increasing, besides, $S(0) = 1$ and $S(\infty) = 0$. Survival function can be used to compute mean survival time by $E[T] = \int_0^{\infty} S(t)dt$.

The hazard function, denoted by $\lambda(t)$, gives the instantaneous rate at which an individual experiences the specified event, conditional on that he or she has survived to the time t , so that

$$\lambda(t) = \lim_{h \rightarrow 0} \frac{Pr(T \in [t, t+h)|T \geq t)}{h} = -\frac{d}{dt} \log S(t) = \frac{f(t)}{S(t)}.$$

There is a fact on $\lambda(t)$ that it is non-negative. A related quantity is the cumulative hazard function defined as

$$\Lambda(t) = \int_0^t \lambda(x)dx = -\log S(t).$$

With explanatory variable X measured at the entry into the study, we can further specify the hazard function as

$$\lambda(t|X) = \lim_{h \rightarrow 0} \frac{Pr(T \in [t, t+h] | T \geq t, X)}{h}.$$

Correspondingly, we can write $f(t|X)$, $F(t|X)$, $S(t|X)$ and $\Lambda(t|X)$.

1.2 Censoring

1.2.1 Right Censoring and Left Censoring

The main feature of survival time is censoring. Censoring is unavoidable in clinical trials or medical investigations because one cannot wait very long time for the occurrences of the specified events to know the complete survival time interval. The survival time of an individual is censored when the individual has not experienced a given event at the end of the study or is lost to follow-up during the study duration. In these situations, the exact survival time is unknown, and the survival time of the individual is called to be right censored. In many biomedical studies, right-censoring occurs frequently when some individuals are still alive during the study period. We consider a random censorship model where the individual has the time to censoring denoted by C , and C is a random variable. T is observable if $T < C$, otherwise T is censored. In other words, the observation for T is $Y = \min(T, C)$. The censoring indicator denoted by $\delta = I(T < C)$ indicates whether T is observable or not.

Another typical type of censoring is left censoring. It occurs when an individual experiences the event prior to the time t with the unknown actual time to the event. For example, the observation may be $[0, t]$ in a left

censoring case or $[t, \infty)$ in a right censoring setting. Other types of censoring are not under our consideration.

1.2.2 Dependent Censoring

Independence between the survival time and the censoring time is often assumed in conventional statistical methods for survival analysis, and leads to a simple censoring mechanism. A more general case is that the independence occurs conditionally on other explanatory variables.

However, if the time to the event of interest is censored by the time to another event and the correlation between times to both events cannot be explained by other covariates, the independence assumption will be violated. This leads to data settings including the competing risks and semi-competing risks data.

As an example, overall survival is defined as the time from study entry until death, which can be referred to as a terminal event, while the progression-free survival (PFS), or the time to tumor progression, is defined as the time elapsed between study entry and tumor progression (e.g., distant metastasis), which we refer to as the non-terminal event. Individuals in the trials are likely to experience either the terminal or non-terminal event or both. It is well known that both types of events may be dependent and the terminal event time can censor the non-terminal one but not vice versa. Observations arising from such a setting fall into the category of semi-competing risks data (Fine et al., 2001).

If one only cares about the time to the first event, the reduced data is recognized as competing risks data. In the competing risks setting, either of the non-terminal or terminal events happens, then another event will never happen. One of our interests in this thesis focuses the semi-competing risks

data. Conditional analysis of such data will be discussed in Chapter 2.

A variety of methodological approaches have been developed for analysis of survival data in the presence of semi-competing risks. Most of them account for dependency between both types of events using the joint distribution formulated by copula models. For example, an estimation procedure without considering covariates can be found in Fine et al. (2001), which introduced a concordance estimating equation to obtain a consistent estimator of the association parameter. This method was extended to a time-dependent copula model for accommodating covariates by Peng and Fine (2007), in which a nonlinear estimating equations approach was developed. Chen (2012) further studied the joint analysis based on a copula model with two marginal semiparametric transformation models. As an alternative to the copula-based models, an illness-death model was introduced by Xu et al. (2010) with a shared gamma frailty accounting for dependency between two event times. They also developed a restricted model and showed its equivalency to a Clayton copula model. When the primary goal is only to assess the effects of covariates on either non-terminal or terminal event, Lin et al. (1996), Peng and Fine (2006) and Ding et al. (2009) considered two marginal models and introduced artificial censoring techniques. Their methods do not describe explicitly the association between the two events.

1.3 Semi-parametric Models

Three kinds of semi-parametric models will be considered in this thesis.

1.3.1 Cox Proportional Hazards Model

The Cox (1972) proportional hazards (PH) model has been widely used to model the relationship between covariates and a censored outcome. It states that given potential covariates, the conditional hazard functions of T for different individuals i and j are proportional. In other words, the ratio of $\lambda(t|X_i)$ and $\lambda(t|X_j)$ does not vary with time t , where X_i and X_j are covariate vectors for individuals i and j ($i, j = 1, \dots, n$), respectively, and n is the sample size in the study. Cox (1972) also suggested that the relative risk is an exponential function of explanatory covariates. Thus the Cox PH model is specified as

$$\lambda(t|X) = \lambda_0(t) \exp(\beta^T X),$$

where $\lambda_0(t)$ is an unknown baseline function, which only depends on time t , and β is an unknown parameter associated with the covariate vector X . The developments of estimation procedures and inferences for the Cox PH model were provided in Andersen and Gill (1982). We refer to Lin and Wei (1991) and Lin et al. (1993) for more information about the goodness-of-fit for the Cox PH model.

1.3.2 Accelerated Failure Time Model

As an alternative to the Cox PH model, the accelerated failure time (AFT) model has a nice interpretation. It links linearly the explanatory covariates to the logarithm transformation of survival time. Specifically, the semi-parametric AFT model can be expressed as

$$\log T = \gamma^T X + \xi,$$

where γ is an unknown parameter associated with the covariate vector X , and ξ is the error term with an unknown distribution function and is independent of X . The estimation for such a semi-parametric AFT model is typically achieved by solving a rank-based estimating equation (Ying, 1993). Details will be discussed in the next chapter.

1.3.3 Censored Quantile Regression

Quantile regression (QR) is a powerful tool to estimate the conditional distribution of outcomes through modeling the relationship between its conditional quantiles and the explanatory predictors, while the specification of a parametric distribution for the outcomes is not required. Analogous to the comparison between median and mean, quantile regression is a robust analysis against the contamination of outlying observations (Koenker, 2005).

In survival data settings, where outcomes are often censored, quantile regression has received increasing attention in recent years as it directly links a set of risk factors to quantiles of survival time or its log-transformation, providing easier interpretations for heteroscedasticity and structural inhomogeneity in the population. Particularly, the censored quantile regression (CQR) can be specified as

$$Q_T(\tau|X) = \exp[X^T \beta(\tau)],$$

where $Q_T(\tau|X) = \inf\{t : \Pr(T \leq t|X) \geq \tau\}$ for quantile level $\tau \in (0, 1)$, and $\beta(\tau)$ is an unknown parameter corresponding to the covariate vector X and varies with τ . The change of the regression coefficients in the quantile regression reflects that some quantiles of a subject's survival time may be more affected by risk factors than the other quantiles. Unlike the conventional survival models, such as the Cox proportional hazards model or

proportional odds model in which the effect of a risk factor is interpreted concerning hazard ratio or odds ratio, the censored quantile regression explains such an effect more directly and easier to be understood by practitioners.

A variety of approaches has been developed for fitting quantile regression models on survival data. Powell (1986) focused on survival data subject to fixed censoring, which is not practical in many application settings. Ying et al. (1995) proposed an estimating equation with inverse probability of censoring weight for the median regression model with randomly censored survival data. Their work relied on the independence condition between the censoring time and multiple covariates. For more general cases in which the censoring time and the outcomes are independent conditionally on covariates, Portnoy (2003) applied the redistribution-of-mass construction for the Kaplan-Meier estimator to develop weighted quantile regression through a recursive reweighting scheme. Peng and Huang (2008) started with the martingale representation of the Nelson-Aalen estimator to develop an estimating procedure for survival data under conditionally independent censoring.

Adopting the same redistribution-of-mass idea as in Portnoy's work, Wang and Wang (2009) proposed a locally weighted estimator with weight functions estimated nonparametrically via the local Kaplan-Meier estimator, which relaxed the strict global linearity assumption made by Portnoy (2003) and Peng and Huang (2008). Alternatively, Bottai and Zhang (2010) studied a likelihood-based approach to estimate regression quantiles with censored data based on the asymmetric Laplace distribution. More recently, Leng et al. (2013) extended Ying et al. (1995)'s unbiased estimating equations to more general random censoring models by incorporating the inverse

probability of censoring weighting technique. Their probability of censoring weight was obtained by the local Kaplan-Meier estimator following the work of Wang and Wang (2009). It is noted that both Wang and Wang (2009) and Leng et al. (2013) required pre-specifications of the kernel function and smoothing parameter in their estimation procedures.

1.4 Motivation

1.4.1 Breast Cancer Data

We study the breast cancer data collected by Haibe-Kains et al. (2012). Breast cancer is one of the most common cancer and the leading cause of cancer death among women around the world. Despite the overall survival of breast cancer patients increases remarkably in the past decades due to progress in diagnosis and treatment of early breast cancer, 20 – 30% of the patients were found developing a distant recurrence (Lobbezoo et al., 2015). Once distant metastatic occurs, the median survival of breast cancer patients reduces substantially to a range of 2 to 3 years (Cardoso et al., 2012). In this study, two types of events are of interest, that is distant metastasis-free survival (DMFS) and overall survival. A patient may experience distant metastasis and death sequentially, but she will not have distant metastasis once died. Thus the data shall be categorized as the semi-competing risks data.

In applications of cancer studies, instead of assessing association, to study whether improved tumor control translates into prolonged overall survival is attracting greater interests among medical researchers. We refer to Saad and Katz (2009), Halabi et al. (2009), Heng et al. (2011) and Ter-

Minassian et al. (2017) for examples of recent developments on this topic in cancer studies. All of these works found that PFS was a meaningful predictor for OS. Particularly, Halabi et al. (2009) considered PFS as an independent binary predictor (1=occurrence of tumor progression, 0=others) of OS for breast cancer patients. Félix et al. (2013) used a simple censored normal regression to fit the relationship between the intermediate event and terminal event. Empirical investigations in the aforementioned studies have shed some light on the role of the non-terminal event as a covariate that may make gains in prediction accuracy in the regression model on OS. This motivates us to study a new regression model of OS conditional on the aggregation of the non-terminal event information together with other risk factors in Chapter 2 of the thesis, which can provide an insightful understanding of the relationship between the two events.

1.4.2 GenIMS Data

We study the genetic and inflammatory markers of sepsis (GenIMS) data, a research project done by Kellum et al. (2007) with purpose of identifying how biomarkers for inflammatory cytokine responses to pneumonia affect the overall survival of patients with the community acquired pneumonia (CAP), which is recognized as the main cause of severe sepsis. The study collected three biomarkers for inflammatory cytokine responses, including tumor necrosis factor (TNF), interleukin-6 (IL-6) and interleukin-10 (IL-10). Among them, TNF and IL-6 are markers of the proinflammatory cytokine response, while IL-10 is a marker of the anti-inflammatory response. According to the manufacturer's specifications, these cytokine responses are measured having lower limits: 4, 2 or 5, and 5pg/ml for TNF, IL-6 and IL-10, respectively. Our interest in Chapter 3 of this thesis lies in inves-

tigating the effects of these cytokine responses together with demographic characteristics on the distribution of survival time for CAP patients.

1.5 Thesis Outline

The rest of this thesis is organized into three chapters as follows.

Motivated by the breast cancer study, existing joint models cannot explain how an intermediate disease progression (non-terminal event) affects the overall survival (terminal event). In Chapter 2, to emphasize the intermediate effects on overall survival, we propose a novel conditional modeling approach for analyzing survival data in the presence of semi-competing risks. The proposed conditional model can effectively evaluate the effect of the non-terminal event on the prediction of overall survival through a proportional hazards model with a covariate subject to dependent censoring. We develop a two-stage estimation procedure and show the asymptotic properties of the resulting estimator using modern techniques of the empirical process and counting process. Simulation studies and analysis of the motivated data demonstrate good performances of the proposed method.

In Chapter 3, motivated by the GenIMS data in which both covariates and outcomes are subject to censoring, we consider the censored quantile regression and utilize the multiple (quantile) imputation approach without any distribution assumption for censored covariates. We provide the asymptotic properties of the proposed estimator and display estimation results of numerical examples to verify the proposed method.

We will conclude the thesis with discussions and future works in Chapter 4. We attach the detailed technical proofs at the end of each chapter.

Chapter 2

Conditional Modeling of Survival Data with Semi-competing Risks

2.1 Introduction

There are some references on using the information of the non-terminal event time to describe the survival probability of the terminal event time. Without taking other risk factors into account, Parast et al. (2011) and Li and Cheng (2016) considered to predict the terminal event time using information from the non-terminal event. In particular, Parast et al. (2011) utilized the empirical conditional probability along with inverse probability weights in the case having a single categorical covariate only. Li and Cheng (2016) estimated the survival function of the terminal event conditional on the non-terminal event by defining a cross-hazard ratio between two event time variables, while the impacts of risk factors are ignored. When incorporating other risk factors, Yang and Peng (2016) proposed to use a cross-quantile residual ratio function for measuring the association between the two events, and the proposed method is implemented by a working

quantile residual lifetime regression model with the binary indicator of the non-terminal event. From the perspective of the multi-state model, Hu and Tsodikov (2014) considered a Cox PH model for the terminal time using the counting process for the non-terminal event time as a predictor. To the best of our knowledge, there is no study relating the terminal event to the length of the non-terminal event time, rather than its indicator, together with other risk factors in one model.

In this chapter, we aim to extend the methodology and theory of conditional analysis to a regression framework embracing the non-terminal event time and multiple covariate contributions. We specify the conditional model in line with the Cox proportional hazards model but allowing the non-terminal event time to be a covariate subject to dependent and independent censoring. Inspired by the work of Kong and Nan (2016) for generalized linear models with a covariate under a detection limit, we proposed a two-stage pseudolikelihood estimation procedure. In the first stage of the estimation, a semi-parametric accelerated failure time (AFT) model is fitted for relating the nonterminal event time to covariates using rank-based estimating equations, and then the conditional distribution of the nonterminal event time given other covariates can be estimated roughly by the Kaplan-Meier (KM) estimation based on the residuals. A KM induced estimator is then proposed to capture the issue of dependent censoring. In the second stage, given the estimated conditional distribution of the nonterminal event time, parameters can be estimated by maximizing a semiparametric pseudo-likelihood. Note that the implementation of the AFT model in the first stage can be easily done using R package *aftgee*.

The dependent censoring in the semi-competing risks setting poses a great challenge, especially in the theoretical justification of the resulting

estimator, which relies on the estimator obtained from the AFT model in the first stage. The main technical challenge is to show the consistency of the AFT estimator for the nonterminal event time under dependent censoring. We apply the induced-smoothing estimating equations for estimation in the AFT model and show its equivalence to the non-smoothing estimating equations in producing consistent estimators. Empirical process techniques are employed to derive the consistency of the AFT estimator. Moreover, counting process techniques and semi-parametric Z-theorem are applied for proving the asymptotic properties of the resulting estimator.

In the rest of this chapter, first of all, we introduce the conditional model considered and its corresponding likelihood function. A two-stage estimation procedure is proposed. Then we provide the theoretical justification for the proposed estimator. The performances of the proposed method are assessed through simulation studies and illustrated by an analysis of breast cancer data.

2.2 Model

Let V be the non-terminal event time and T be the terminal event time. We consider the semi-competing risk setting under which (V, T) can be viewed as a bivariate failure time, what is more, T censors V but not vice versa. For example, T can be overall survival, and V can be progression-free survival or distant metastasis-free survival. Let $\mathbf{X}$ be $p \times 1$ covariate vector and it is fully observed without censoring. In view of clinical significance, it is of importance to predict overall survival on progression-free survival. Therefore, our proposal is to relate T to $(V, \mathbf{X})$ naturally via the

Cox proportional hazards model:

$$\lim_{h \downarrow 0} \frac{\Pr(t \leq T < t + h | T \geq t, V, X)}{h} = \lambda(t|V, \mathbf{X}) = \lambda_0(t) \exp(\alpha V + \boldsymbol{\beta}^T \mathbf{X}), \quad (2.1)$$

where $\lambda_0(t)$ is an unknown baseline hazard function corresponding to $\mathbf{X} = 0$ and $V = 0$. Note that we consider $V = 0$ here for the technical validity of the model, while V is always greater than zero in practice. $\alpha < 0 (> 0)$ implies that the earlier occurrence of the non-terminal event will lead to a higher (lower) risk of the terminal event; $\alpha = 0$ suggests that there is no association between the non-terminal event and terminal event.

The censoring time C^* is assumed to be independent of (T, V) and $\mathbf{X}$. We assume that the maximum follow-up time in the study exists and is denoted as ς . In model (2.1), the nonterminal event time V is treated as a dependently censored covariate and observed only when $V < \min(T, C)$ with $C = \min(C^*, \varsigma)$. Due to censoring, T and V can be potentially observable as $Y = \min(T, C)$ and $Z = \min(V, T, C)$, respectively. The censoring indicators can also be observed and denoted by $\delta_1 = I(V < \min(T, C))$, $\delta_2 = I(T \leq \min(V, C))$ and $\tilde{\delta}_2 = I(T \leq C)$. Actually, we can easily see that $\delta_2 = (1 - \delta_1)\tilde{\delta}_2$.

Before deriving the likelihood function, we need to analyze the four possible cases in the semi-competing risk setting: (1) the non-terminal and terminal event times are both observed, in this case, we have $\delta_1 = 1$ and $\tilde{\delta}_2 = 1$; (2) the non-terminal event time is observed but the terminal time is censored, then $\delta_1 = 1$ and $\tilde{\delta}_2 = 0$; (3) the terminal event time is observed and it censors the non-terminal one, and we have $\delta_2 = 1$; (4) the non-terminal and terminal event times are both censored, that is, $\delta_1 = 0$ and $\delta_2 = 0$. Based on these four cases, the likelihood for one realization is

derived by Bayes' rule as

$$\begin{aligned}
L(\Omega; Y, \mathbf{X}, Z, \delta_1, \delta_2, \tilde{\delta}_2) &= \Pr(V = Z, T = Y, X)^{\delta_1 \tilde{\delta}_2} \times \Pr(V = Z, T \geq Y, X)^{\delta_1(1-\tilde{\delta}_2)} \\
&\times \Pr(V \geq Z, T = Z, X)^{\delta_2} \times \Pr(V \geq Z, T \geq Z, X)^{1-\delta_1-\delta_2} \\
&= \left\{ f_{T|V,\mathbf{X}}(Y|Z, \mathbf{X}) f_{V|\mathbf{X}}(Z|\mathbf{X}) f_{\mathbf{X}}(\mathbf{X}) \right\}^{\delta_1 \tilde{\delta}_2} \\
&\times \left\{ f_{V|\mathbf{X}}(Z|\mathbf{X}) f_{\mathbf{X}}(\mathbf{X}) \int_Y^\infty f_{T|V,\mathbf{X}}(t|Z, \mathbf{X}) dt \right\}^{\delta_1(1-\tilde{\delta}_2)} \\
&\times \left\{ \int_Z^\infty f_{T|V,\mathbf{X}}(Z|v, \mathbf{X}) f_{V|\mathbf{X}}(v|\mathbf{X}) f_{\mathbf{X}}(\mathbf{X}) dv \right\}^{\delta_2} \\
&\times \left\{ \int_Z^\infty \int_Z^\infty f_{T|V,\mathbf{X}}(t|v, \mathbf{X}) f_{V|\mathbf{X}}(v|\mathbf{X}) f_{\mathbf{X}}(\mathbf{X}) dv dt \right\}^{1-\delta_1-\delta_2}, \tag{2.2}
\end{aligned}$$

where $f_{T|V,\mathbf{X}}$, $f_{V|\mathbf{X}}$ and $f_{\mathbf{X}}$ are densities of T given $(V, \mathbf{X})$, V given $\mathbf{X}$, and $\mathbf{X}$ respectively. Taking logarithm on the likelihood in (2.2) and dropping $f(\mathbf{X})$ as it is irrelevant to the parameters of interest, it gives

$$\begin{aligned}
l(\Omega; Y, \mathbf{X}, Z, \delta_1, \delta_2, \tilde{\delta}_2) &= \delta_1 \tilde{\delta}_2 \log f_{T|V,\mathbf{X}}(Y|Z, \mathbf{X}) + \delta_1 \log f_{V|\mathbf{X}}(Z|\mathbf{X}) \\
&+ \delta_1(1 - \tilde{\delta}_2) \log \left[\int_Y^\infty f_{T|V,\mathbf{X}}(t|Z, \mathbf{X}) dt \right] \\
&+ \delta_2 \log \left\{ \int_Z^\infty f_{T|V,\mathbf{X}}(Z|v, \mathbf{X}) dF_{V|\mathbf{X}}(v|\mathbf{X}) \right\} \\
&+ (1 - \delta_1 - \delta_2) \log \left\{ \int_Z^\infty \int_Z^\infty f_{T|V,\mathbf{X}}(t|v, \mathbf{X}) dF_{V|\mathbf{X}}(v|\mathbf{X}) dt \right\}, \tag{2.3}
\end{aligned}$$

where $\Omega = \{\lambda_0, \alpha, \boldsymbol{\beta}, F_{V|\mathbf{X}}\}$ is a set of unknown parameters and functions to be estimated, and $F_{V|\mathbf{X}}$ is the cumulative density function of V given covariates $\mathbf{X}$.

2.3 Estimation Procedure and Implementation

We propose to estimate unknown parameters (α, β) by maximizing the log-likelihood in (2.3). To this end, we need to handle the conditional density function $f_{T|V, \mathbf{X}}(\cdot)$ and cumulative density function $F_{V|\mathbf{X}}(\cdot)$. Based on model (2.1), the probability density of overall survival T conditional on $(V, \mathbf{X})$ can be written as

$$\begin{aligned} f_{T|V, \mathbf{X}}(t|v, x) \\ = \lambda_0(t) \exp(\alpha v + \beta^T x) \exp \left[- \int_0^t \lambda_0(s) \exp(\alpha v + \beta^T x) ds \right]. \end{aligned} \quad (2.4)$$

Since V can be subject to dependent or independent censoring, it is difficult to specify $F_{V|\mathbf{X}}$ as (2.4). Note that there are observations for V (may be censored) and $\mathbf{X}$. We first propose to model the relationship between V and $\mathbf{X}$ through an AFT model, and then to estimate the conditional distribution function $F_{V|\mathbf{X}}$ using residuals from the model. This idea is inspired by the work of Kong and Nan (2016) who studied the problem of a generalized linear model with a left-censored covariate. In particular, we assume that given $\mathbf{X}$ the semiparametric AFT model for V has the form

$$\log(V_i) = \gamma^T \mathbf{X}_i + \xi_i, \quad i = 1, \dots, n, \quad (2.5)$$

where γ is an unknown p -dimensional coefficient vector and can be regarded as a nuisance parameter in this study, and ξ_i 's are independent of $\mathbf{X}_i$ and have unknown CDF $Q(\xi)$. Here, covariates $\mathbf{X}_i$ in (2.5) may be partial different or overlapped with that in (2.1). To ease presentation, we consider the same covariates in the two models.

With model (2.5), the set of parameters and functions to be estimated

turns to be $\Omega = \{\lambda_0, \alpha, \boldsymbol{\beta}, \boldsymbol{\gamma}, Q\}$. For observed data comprising n independent and identically distributed observations, denoted by $\{(Y_i, \mathbf{X}_i, Z_i, \delta_{1i}, \delta_{2i}, \tilde{\delta}_{2i}) : i = 1, \dots, n\}$, the log-likelihood of Ω becomes

$$\begin{aligned}
& \sum_{i=1}^n l(\Omega; Y_i, \mathbf{X}_i, Z_i, \delta_{1i}, \delta_{2i}, \tilde{\delta}_{2i}) \\
&= \sum_{i=1}^n \delta_{1i} \tilde{\delta}_{2i} \log f_{T|V, \mathbf{X}}(Y_i | Z_i, \mathbf{X}_i) + \delta_{1i} \log \left[\frac{1}{Z_i} Q'(\log Z_i - \boldsymbol{\gamma}^T \mathbf{X}_i) \right] \\
&+ \delta_{1i} (1 - \tilde{\delta}_{2i}) \log \left[\int_{Y_i}^{\infty} f_{T|V, \mathbf{X}}(t | Z_i, \mathbf{X}_i) dt \right] \\
&+ \delta_{2i} \log \left\{ \int_{Z_i}^{\tau} f_{T|V, \mathbf{X}}(Z_i | v, \mathbf{X}_i) dQ(\log v - \boldsymbol{\gamma}^T \mathbf{X}_i) \right\} \\
&+ (1 - \delta_{1i} - \delta_{2i}) \log \left\{ \int_{Z_i}^{\infty} \int_{Z_i}^{\tau} f_{T|V, \mathbf{X}}(t | v, \mathbf{X}_i) dQ(\log v - \boldsymbol{\gamma}^T \mathbf{X}_i) dt \right\},
\end{aligned} \tag{2.6}$$

where Q' represents the first-order derivative of function $Q(\xi)$ with respect to ξ . For the technical purpose, we assume that $\tau < \infty$ is the truncation time for the non-terminal event, while it does not need to be specified in the numerical studies.

The estimation procedure can be implemented through two iterative stages. In the first stage, the nuisance parameter $\boldsymbol{\gamma}$ is estimated through the smooth Gehan rank-based estimating function. In the second stage, a non-parametric estimate of Q can be obtained iteratively using a Kaplan Meier induced estimate of survival (adjusted by dependent censoring) over residuals. We replace $\boldsymbol{\gamma}$ and Q in (2.6) by their corresponding estimators, leading to a semiparametric pseudo-likelihood of the parameters of interest $(\lambda_0, \alpha, \boldsymbol{\beta})$. Details of the two-stage estimation procedure are given below.

2.3.1 First-stage of the Estimation for γ

The regression coefficient vector γ in model (2.5) can be estimated by solving the rank-based weighted estimating equation (Chung et al., 2013)

$$W_G(\gamma) = \frac{1}{n} \sum_{i=1}^n \delta_{1i} \vartheta(\gamma; e_i(\gamma)) \left\{ \mathbf{X}_i - \frac{\sum_{j=1}^n I[e_j(\gamma) \geq e_i(\gamma)] \mathbf{X}_j}{\sum_{j=1}^n I[e_j(\gamma) \geq e_i(\gamma)]} \right\} = 0, \quad (2.7)$$

where $e_i(\gamma) = \log Z_i - \gamma^T \mathbf{X}_i$ and $\vartheta(\gamma; \cdot)$ is a user-specified weight function. Different choices of ϑ include the logrank with $\vartheta = 1$ and Gehan's weight with $\vartheta(\gamma; t) = n^{-1} \sum_{j=1}^n I[e_j(\gamma) \geq t]$. The routine rank-based estimating functions with Gehan's weight for the AFT model (Ying, 1993; Jin et al., 2003) is

$$\mathbf{W}_n(\gamma) = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \delta_{1i} (\mathbf{X}_i - \mathbf{X}_j) I[e_j(\gamma) - e_i(\gamma) \geq 0], \quad (2.8)$$

The asymptotic properties of the estimator obtained by solving equations in (2.8) were justified by Ying (1993) when ignoring the possibility of dependent censoring for V . Note that the logrank estimating function is not monotone and the corresponding estimating equations may produce multiple roots. On the other hand, Gehan's weighted estimating equation method is more popular due to its monotonicity. However, both weighted rank-based estimating equations are non-smoothing, and thus pose large computational challenges.

Jin et al. (2003) proposed to use linear programming for implementation, but it may cause computation complexity for large n or large p cases. Given that smooth approximation becomes more popular in recent ten years owing to its fast and simple computation, see for example Brown and Wang (2007), Johnson and Strawderman (2009), Fu et al. (2010), Wang and Fu (2011)

and Chiou et al. (2014a), one can obtain induced-smoothing estimating equations (ISREE) by substituting $W_G(\gamma)$ with $E_\omega[W_G(\gamma + \Gamma_n^{1/2}\omega)]$, where ω is a $p \times 1$ standard normal random vector and Γ_n is a $p \times p$ user-specified matrix. We use Φ and ϕ to represent the CDF and PDF of the standard normal distribution. The ISREE with Gehan's weight for the AFT model is then given by

$$\widetilde{\mathbf{W}}_n(\gamma) = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \delta_{1i}(\mathbf{X}_i - \mathbf{X}_j) \Phi \left[\frac{e_j(\gamma) - e_i(\gamma)}{\sigma_{ji}} \right] = 0, \quad (2.9)$$

where $\sigma_{ji}^2 = (\mathbf{X}_i - \mathbf{X}_j)^T \Gamma_n (\mathbf{X}_i - \mathbf{X}_j)$ for some non-negative definite and symmetric matrix Γ_n . More details about the ISREE can be found in Brown and Wang (2007) and Johnson and Strawderman (2009). The estimate of γ , denoted by $\hat{\gamma}$ is then obtained by solving equations in (2.9). In practice, this estimation can be easily implemented in R using package *aftgee* created by Chiou et al. (2014b).

2.3.2 Kaplan-Meier-induced Estimator for Q

For the sake of convenience, we take the same empirical process notations as in Van der Vaart (2000). Given a function f , denote the expectation of f under P by Pf and the expectation of f under empirical measure by $\mathbb{P}_n f$. In other words, $Pf = \int f(u) dP(u)$, $\mathbb{P}_n f = n^{-1} \sum_{i=1}^n f(u_i)$. By the law of large numbers, $\mathbb{P}_n f$ converges to Pf almost surely. Next, we estimate the error distribution based on the AFT residuals $e(\gamma) = \log Z - \gamma^T \mathbf{X}$. Note that the Kaplan-Meier (KM) estimator is consistent and asymptotically normal only under independent censoring, and thus cannot be applied to the current case of residuals under dependent censoring. To overcome this problem, we introduce a new estimator for $Q(t)$, which is derived from the KM estimator and thus called Kaplan-Meier induced estimator.

In particular, we first introduce an intermediate quantity defined by

$$Q^*(\gamma, t) = 1 - \exp \left[- \int_{s \leq t} \frac{dh^{(0)}(\gamma, s)}{h^{(1)}(\gamma, s)} \right],$$

which can be estimated nonparametrically as

$$\widehat{Q}_{\widehat{\gamma}}^{km}(t) = 1 - \prod_{i: e_i(\widehat{\gamma}) \leq t} \left\{ 1 - \frac{\delta_{1i}/n}{H_n^{(1)}(\widehat{\gamma}, e_i(\widehat{\gamma}))} \right\}, \quad (2.10)$$

where

$$\begin{aligned} h^{(0)}(\gamma, s) &= P \{ I(e(\gamma) \leq s, \delta_1 = 1) \}, \\ h^{(1)}(\gamma, s) &= P \{ I(e(\gamma) \geq s) \}, \\ H_n^{(1)}(\gamma, e_i(\gamma)) &= \mathbb{P} \{ I(e(\gamma) \geq e_i(\gamma)) \}. \end{aligned} \quad (2.11)$$

$\widehat{Q}_{\widehat{\gamma}}^{km}(t)$ is indeed the Kaplan-Meier estimator based on data $\{e_i(\widehat{\gamma}), \delta_{1i}\}_{i=1, \dots, n}$.

Let Q_0 be the true cumulative distribution function for ξ . Plugging the true values of all parameters, we have

$$Q^*(\gamma_0, t) = 1 - \exp \left\{ - \int_{s \leq t} K_Q(s; \alpha_0, \beta_0, \Lambda_{0,0}, \gamma_0, Q_0, S_c) dQ_0(s) \right\}, \quad (2.12)$$

where

$$\begin{aligned} &K_Q(s; \alpha, \beta, \Lambda_0, \gamma, Q, S_c) \\ &= \frac{E \left[Pr(C \geq e^{s+\gamma^T \mathbf{X}}) Pr(T \geq e^{s+\gamma^T \mathbf{X}} | V = e^{s+\gamma^T \mathbf{X}}, X) \right]}{E \left[Pr(C \geq e^{s+\gamma^T \mathbf{X}}) Pr(T \geq e^{s+\gamma^T \mathbf{X}}, V \geq e^{s+\gamma^T \mathbf{X}}, X) \right]} \\ &= \frac{P \left\{ S_c(e^{s+\gamma^T \mathbf{X}}) \exp \left[-\Lambda_0(e^{s+\gamma^T \mathbf{X}}) \exp(\alpha e^{s+\gamma^T \mathbf{X}} + \beta^T \mathbf{X}) \right] \right\}}{P \left\{ S_c(e^{s+\gamma^T \mathbf{X}}) \int_{u \geq s} \exp \left[-\Lambda_0(e^{s+\gamma^T \mathbf{X}}) \exp(\alpha e^{u+\gamma^T \mathbf{X}} + \beta^T \mathbf{X}) \right] dQ(u) \right\}} \\ &:= \frac{h^{(3)}(s; \alpha, \beta, \Lambda_0, \gamma, Q, S_c)}{h^{(4)}(s; \alpha, \beta, \Lambda_0, \gamma, Q, S_c)}, \end{aligned}$$

and $S_c(\cdot)$ is the survival function of C . Note that $K_Q(s; \alpha, \beta, \Lambda_0, \gamma, Q, S_c) = 1/[1 - Q(s)]$ when T and V are independent. Treating (2.12) as an equation of Q_0 and solving it, we can consequently express Q_0 in relation to other parameters as follows:

$$Q_0(t) = \int_{s \leq t} \frac{dQ^*(\gamma_0, s)}{K_Q(s; \alpha_0, \beta_0, \Lambda_{0,0}, \gamma_0, Q_0, S_c)[1 - Q^*(\gamma_0, s)]} \quad (2.13)$$

when $t \in \{t : Q^*(\gamma_0, t) \neq 1\}$. Then the Kaplan-Meier induced estimator can be constructed in the form of

$$\hat{Q}(t) = \int_{s \leq t} \frac{d\hat{Q}_{\hat{\gamma}}^{km}(s)}{\hat{K}_Q(s; \hat{\alpha}, \hat{\beta}, \hat{\Lambda}_0, \hat{\gamma}, \hat{Q}, \hat{S}_c)[1 - \hat{Q}_{\hat{\gamma}}^{km}(s)]} \quad (2.14)$$

if $t \in \{t : \hat{Q}_{\hat{\gamma}}^{km}(t) \text{ is smaller than } 1\}$ and $\hat{Q}(t) = 1$ otherwise, where

$$\begin{aligned} & \hat{K}_Q(s; \alpha, \beta, \Lambda_0, \gamma, Q, S_c) \\ &= \frac{\mathbb{P}_n \left\{ S_c(e^{s+\gamma^T \mathbf{X}}) \exp \left[-\Lambda_0(e^{s+\gamma^T \mathbf{X}}) \exp(\alpha e^{s+\gamma^T \mathbf{X}} + \beta^T \mathbf{X}) \right] \right\}}{\mathbb{P}_n \left\{ S_c(e^{s+\gamma^T \mathbf{X}}) \int_{u \geq s} \exp \left[-\Lambda_0(e^{s+\gamma^T \mathbf{X}}) \exp(\alpha e^{u+\gamma^T \mathbf{X}} + \beta^T \mathbf{X}) \right] dQ(u) \right\}}, \end{aligned}$$

and $\hat{S}_c$ is the Kaplan-Meier estimator of S_c based on data $\{Y_i, \tilde{\delta}_{2i}\}_{i=1, \dots, n}$.

2.3.3 Weighted Breslow-type Estimator for λ_0

For estimating the baseline hazard function $\lambda_0(t)$, we can use the non-parametric maximum likelihood estimation (NPMLE) method in general line with Andersen and Gill (1982) and Chen (2012). Specifically, one can treat the baseline function $\Lambda_0 = \int_0^t \lambda_0(s) ds$ as a step function, and $\lambda_0(t)$ is viewed as jump size of $\Lambda_0(t)$ at observed terminal time t . Denote $Y_l(t) = I(Y_l \geq t)$ and counting processes by $N_{2i}(t) = \delta_{2i} I(Y_i \leq t)$

and $\tilde{N}_{2i}(t) = \delta_{1i}\tilde{\delta}_{2i}I(Y_i \leq t)$. The NPMLE for λ_0 is obtained via maximizing the log-likelihood in (2.6) with respect to λ_0 . Given $(\alpha, \beta, \gamma, Q)$, the weighted Breslow-type estimators of the baseline hazard $\lambda_0(t)$ and the baseline cumulative hazard $\Lambda_0(t)$ can be expressed respectively as

$$\hat{\lambda}_0(t) = \sum_{i=1}^n \frac{dN_{2i}(t) + d\tilde{N}_{2i}(t)}{\sum_{l=1}^n Y_l(t)\mu_l(\alpha, \beta, \Lambda_0, \gamma, Q)}$$

and

$$\hat{\Lambda}_0(t) = \sum_{i=1}^n \frac{(\delta_{1i}\tilde{\delta}_{2i} + \delta_{2i})(1 - Y_i(t))}{\sum_{l=1}^n Y_l(t)\mu_l(\alpha, \beta, \Lambda_0, \gamma, Q)}, \quad (2.15)$$

where

$$\mu_{1,l}(\alpha, \beta) = \exp(\alpha Z_l + \beta^T \mathbf{X}_l),$$

$$\mu_{2,l}(\alpha, \beta, \Lambda_0, \gamma, Q) = \tilde{\delta}_{2l} \frac{B_l(\alpha, \beta, \Lambda_0, \gamma, Q)}{A_l(\alpha, \beta, \Lambda_0, \gamma, Q)} + (1 - \tilde{\delta}_{2l}) \frac{A_l(\alpha, \beta, \Lambda_0, \gamma, Q)}{C_l(\alpha, \beta, \Lambda_0, \gamma, Q)},$$

$$A_l(\alpha, \beta, \Lambda_0, \gamma, Q) = \int_{v \geq Y_l} \exp(\alpha v + \beta^T \mathbf{X}_l) \exp[-\Lambda_0(Y_l) \exp(\alpha v + \beta^T \mathbf{X}_l)] dQ(\log v - \gamma^T \mathbf{X}_l),$$

$$B_l(\alpha, \beta, \Lambda_0, \gamma, Q) = \int_{v \geq Y_l} \exp(2\alpha v + 2\beta^T \mathbf{X}_l) \exp[-\Lambda_0(Y_l) \exp(\alpha v + \beta^T \mathbf{X}_l)] dQ(\log v - \gamma^T \mathbf{X}_l),$$

$$C_l(\alpha, \beta, \Lambda_0, \gamma, Q) = \int_{v \geq Y_l} \exp[-\Lambda_0(Y_l) \exp(\alpha v + \beta^T \mathbf{X}_l)] dQ(\log v - \gamma^T \mathbf{X}_l),$$

$$\mu_l(\alpha, \beta, \Lambda_0, \gamma, Q) = \delta_{1l}\mu_{1,l}(\alpha, \beta) + (1 - \delta_{1l})\mu_{2,l}(\alpha, \beta, \Lambda_0, \gamma, Q).$$

Note that in the numerical calculation if we encounter the problem of $\frac{0}{0}$, we assign 0 for it.

2.3.4 Second-stage of the Estimation for (α, β)

Replacing γ , Q and Λ_0 by their estimators obtained in the first-stage leads to the following pseudo-likelihood of α and β :

$$\begin{aligned}
 pl(\alpha, \beta) &= \frac{1}{n} \sum_{i=1}^n \left\{ \delta_{1i} \tilde{\delta}_{2i} \log f_{T|V, \mathbf{X}}(Y_i | Z_i, \mathbf{X}_i; \hat{\Lambda}_0(t)) \right. \\
 &\quad + \delta_{1i} (1 - \tilde{\delta}_{2i}) \log \left[\int_{Y_i}^{\infty} f_{T|V, \mathbf{X}}(t | Z_i, \mathbf{X}_i; \hat{\Lambda}_0(t)) dt \right] \\
 &\quad + \delta_{2i} \log \left[\int_{Z_i}^{\tau} f_{T|V, \mathbf{X}}(Z_i | v, \mathbf{X}_i) d\hat{Q}(\log v - \hat{\gamma}^T \mathbf{X}_i) \right] \\
 &\quad \left. + (1 - \delta_{1i} - \delta_{2i}) \log \left[\int_{Z_i}^{\infty} \int_{Z_i}^{\tau} f_{T|V, \mathbf{X}}(t | v, \mathbf{X}_i) d\hat{Q}(\log v - \hat{\gamma}^T \mathbf{X}_i) dt \right] \right\}.
 \end{aligned} \tag{2.16}$$

Note that compared with (2.6), the term $\delta_1 \log [\frac{1}{Z} Q'(\log Z - \gamma^T \mathbf{X})]$ is dropped because it is free of (α, β) . We estimate α , β by maximizing (2.16) through the Newton Raphson algorithm. More specifically, given initial values $(\hat{\alpha}^0, \hat{\beta}^0)$,

$$\begin{pmatrix} \hat{\alpha} \\ \hat{\beta} \end{pmatrix} = \begin{pmatrix} \hat{\alpha}^0 \\ \hat{\beta}^0 \end{pmatrix} - \left(\Delta(\hat{\alpha}^0, \hat{\beta}^0) \right)^{-1} \Sigma(\hat{\alpha}^0, \hat{\beta}^0), \tag{2.17}$$

where

$$\Sigma(\alpha, \beta) = \begin{pmatrix} \frac{\partial pl(\alpha, \beta)}{\partial \alpha} \\ \frac{\partial pl(\alpha, \beta)}{\partial \beta} \end{pmatrix}, \quad \Delta(\alpha, \beta) = \begin{pmatrix} \frac{\partial^2 pl(\alpha, \beta)}{\partial \alpha^2} & \frac{\partial^2 pl(\alpha, \beta)}{\partial \alpha \partial \beta^T} \\ \frac{\partial^2 pl(\alpha, \beta)}{\partial \alpha \partial \beta} & \frac{\partial^2 pl(\alpha, \beta)}{\partial \beta \partial \beta^T} \end{pmatrix},$$

and the calculation of partial derivatives of $pl(\alpha, \beta)$ is straightforward but may be tedious. To deal with the integrations in (2.16), one can make good use of properties of the survival function. Since $\hat{Q}$ is estimated as a step function as given in (2.14). Let $v_{(1)}, \dots, v_{(\kappa)}$ represent the observed jump points for estimator $\hat{Q}(\cdot)$ with $\log Z_i - \hat{\gamma}^T \mathbf{X}_i < v_{(1)} < \dots < v_{(\kappa)} < \tau_2$ and

$\widehat{dQ}_k$ be the jump size at time point $v_{(k)}$. For different individuals, $v_{(1)}$ may be different. We use the same notation for simplify here. Thus the required integrals can be computed approximately, for example,

$$\begin{aligned} & \int_{\log Z_i - \widehat{\gamma}^T \mathbf{X}_i}^{\tau_2} f_{T|V, \mathbf{X}}(Z_i | \exp(v + \widehat{\gamma}^T \mathbf{X}_i), \mathbf{X}_i) d\widehat{Q}(v) \\ &= \sum_{k=1}^{\mathcal{K}} \widehat{\lambda}_0(Z_i) a_{1,i}^{(v_{(k)})} \exp\{a_{2,i}^{(v_{(k)})}\} d\widehat{Q}_k, \end{aligned}$$

where

$$a_{1,i}^{(v)} = \exp\{\alpha \exp(v + \widehat{\gamma}^T \mathbf{X}_i) + \beta^T \mathbf{X}_i\}, \quad a_{2,i}^{(v)} = a_{1,i}^{(v)} \log \widehat{S}_0(Z_i),$$

and $\widehat{S}_0(t) = \exp\{-\widehat{\Lambda}_0(t)\}$ is the estimator for baseline survival function.

2.3.5 Implementation

To summarize, the two-stage estimation procedure is described as follows.

Stage 1 Estimate nuisance parameters (γ, Q) and set initial values for $(\Lambda_0, \alpha, \beta)$.

- 1) Estimate γ via the smoothing rank-based method, and denote the estimate by $\widehat{\gamma}$.
- 2) Given $\widehat{\gamma}$, we compute the residuals from the fitted AFT model, and estimate Q by the Kaplan-Meier estimation based on the residuals. We denote this estimate of Q by $\widehat{Q}_{\widehat{\gamma}}^{km}$.
- 3) For initial values of $(\alpha, \beta, \Lambda_0)$, we use the complete-case analysis to fit the Cox model (2.1) based on the validation sample data with exactly observed V . We denote the estimates as $\widehat{\alpha}^0, \widehat{\beta}^0$ and $\widehat{\Lambda}_0^{(0)}$.

Stage 2 Update α, β, Λ_0 .

- 1) Given initial values $\hat{\alpha}^0$ and $\hat{\beta}^0$, update (α, β) by maximizing the pseudo-likelihood in (2.16) using the Newton-Raphson algorithm (2.17). Denote the resulting estimates by $\hat{\alpha}$ and $\hat{\beta}$.
- 2) With estimates $\hat{\alpha}$ and $\hat{\beta}$, update $\hat{\Lambda}_0(t)$ using (2.15).
- 3) Compute $\hat{Q}(\cdot)$ using (2.14).

Repeat stage 2 until convergence.

In stage 1, R package *aftgee* can be used for conducting step 1); R function *coxph* from package *survival* can be used to implement the Cox PH model for initial value value in step 4).

2.4 Theoretical Justification

Note that the behaviors of the second-stage estimators rely on those estimators for the parameters in the AFT model though they are the nuisances. In this section, we need to justify the performance of $\hat{\gamma}$ first, and then that of $\hat{\alpha}$, $\hat{\beta}$ and $\hat{\Lambda}_0(t)$. The AFT model has an advantage in direct interpretation but suffers the theoretical and computational difficulties. In the presence of covariates under dependent censoring, the corresponding asymptotic properties have not been discussed in the literature.

To associate with the sample size n , we rewrite the estimator $(\hat{\gamma}_n, \hat{Q}_{n, \hat{\gamma}_n}^{km}, \hat{Q}_n, \hat{\alpha}_n, \hat{\beta}_n, \hat{\Lambda}_{n,0})$ with index n . The asymptotic properties of these estimators are derived from the counting process and empirical process.

2.4.1 Asymptotic Properties of the First-stage Estimator

We first introduce some necessary notations. Assume that the parameter space $\mathcal{G}$ for γ is compact. Following the notations used in Ding and Nan (2015) and apart from (2.11), we further define for any $\gamma \in \mathcal{G}$,

$$\begin{aligned} H_n^{(0)}(\gamma, s) &= \mathbb{P} \{I(e(\gamma) \leq s, \delta_1 = 1)\}, \\ H_n^{(2)}(\gamma, s) &= \mathbb{P} \{I(e(\gamma) \geq s)\mathbf{X}\}, \\ h^{(2)}(\gamma, s) &= P \{I(e(\gamma) \geq s)\mathbf{X}\}. \end{aligned} \quad (2.18)$$

Note that using the notations in (2.11) and (2.18), the rank-based estimating equations with Gehan's weight in (2.8) is equivalent to

$$\mathbf{W}_n(\gamma, H_n^{(1)}, H_n^{(2)}) = \mathbb{P}_n \{ [H_n^{(1)}(\gamma, e(\gamma))\mathbf{X} - H_n^{(2)}(\gamma, e(\gamma))] \delta_1 \} = 0. \quad (2.19)$$

Correspondingly, we define

$$\mathbf{W}_0(\gamma, h^{(1)}, h^{(2)}) = P \{ [h^{(1)}(\gamma, e(\gamma))\mathbf{X} - h^{(2)}(\gamma, e(\gamma))] \delta_1 \}, \quad (2.20)$$

which is a non-random function.

In the following, we will establish the consistency of the estimator yielded from the ISREE. Let $\tilde{T}_i = \log T_i - \gamma_0^T \mathbf{X}_i$ and $\tilde{C}_i = \log C_i - \gamma_0^T \mathbf{X}_i$ with true parameter γ_0 . Further let $g_C^{(i)}(\cdot)$ and $G_C^{(i)}(\cdot)$ be the density and cumulative density functions, respectively, of $\tilde{C}_i$ conditional on covariates $\mathbf{X}_i$. Denote $G_{\xi, T}^{(i)}(\cdot, \cdot)$ as the joint cumulative distribution function for $(\xi_i, \tilde{T}_i)$ conditional on covariates $\mathbf{X}_i$. Without loss of generality, we can re-denote $G_{\xi, T}(\cdot, \cdot)$. The following regularity conditions are needed for establishing the asymptotic properties.

Condition 2.1. The parameter space $\mathcal{G}$ for γ is compact with true value γ_0 falling in its interior.

Condition 2.2. The covariates $\mathbf{X}_i \in \mathbb{R}^p$ are i.i.d. random variables with $p = O(1)$ and $\max \|\mathbf{X}_i\|_2 = O(1)$, where $\|\cdot\|_2$ represents the 2- norm.

Condition 2.3. The non-negative definite and symmetric matrix has $\|\Gamma_n\|_2 = O(n^{-1})$.

Condition 2.4. The density function $q(\cdot)$ for error ξ and its first-order derivative are bounded, and $\int_{-\infty}^{\infty} (q'(t)/q(t))^2 q(t) dt$ is bounded.

Condition 2.5. The censoring time C_i is independent of (V_i, T_i) , and $\Pr(C_i^* \geq \varsigma) > 0$. Besides, the density function $g_C(\cdot)$ and its first and second derivatives are bounded.

Condition 2.6. The first, second, and third partial derivatives of $G_{\xi, T}(\cdot, \cdot)$ are all bounded.

Condition 2.1 is a general assumption and considered in most of the literature. Conditions 2.2, 2.4 and the uniform boundedness of $g_C(\cdot)$ in Condition 2.5 are the same as some conditions given in Ying (1993). A weak assumption on the working covariance matrix Γ_n is imposed in Condition 2.3. It was also addressed as Condition C5 by Fu et al. (2010) and Condition C6 by Wang and Fu (2011). The assumptions on uniform boundedness in terms of the derivatives of $g_C(\cdot)$ and $G_{\xi, T}(\cdot, \cdot)$ in Conditions 2.5 and 2.6 respectively are not weakest, but can be satisfied easily in practice. For simplicity, we take some univariate and continuous random variables for instances. There are a large number of probability functions with bounded first and second derivatives. For example, uniform distributions, exponential distributions with finite rate parameter, Weibull distributions with shape parameter larger than 1, normal distributions with standard deviation slightly far from zero, log-normal distribution distributions in which

the standard deviation of the distribution on the log scale is slightly far from zero, and so on.

The following two lemmas are fundamental for deriving the consistency of $\hat{\gamma}$.

Lemma 2.1. Under Conditions 2.1 and 2.2, for any $\gamma \in \mathcal{G}$,

$$\|\mathbf{W}_n(\gamma, H_n^{(1)}, H_n^{(2)}) - \mathbf{W}_0(\gamma, h^{(1)}, h^{(2)})\|_\infty \longrightarrow 0$$

with probability one, where $\|\cdot\|_\infty$ represents the supremum norm.

Lemma 2.1 ensures that the random estimating function $\mathbf{W}_n(\gamma, H_n^{(1)}, H_n^{(2)})$ converge to a non-random function $\mathbf{W}_0(\gamma, h^{(1)}, h^{(2)})$.

Lemma 2.2. Suppose that Conditions 2.1-2.6 hold.

- (i) For any $\gamma \in \mathcal{G}$, $n\|\widetilde{\mathbf{W}}_n(\gamma) - \mathbf{W}_n(\gamma)\|_\infty \longrightarrow 0$ with probability one.
- (ii) Suppose that $\widetilde{\mathbf{W}}_n(\gamma)$ and $E[\mathbf{W}_n(\gamma)]$ are both differentiable with bounded and continuous derivatives in the neighborhood of γ_0 . Denote

$$\widetilde{\mathbf{D}}_n(\gamma) = \frac{\partial}{\partial \gamma} \widetilde{\mathbf{W}}_n(\gamma) \quad \text{and} \quad \mathbf{D}_n(\gamma) = \frac{\partial}{\partial \gamma} E[\mathbf{W}_n(\gamma)].$$

Then with probability one $\widetilde{\mathbf{D}}_n(\gamma_0) \rightarrow \mathbf{D}_n(\gamma_0)$.

Lemma 2.2 highlights the asymptotic equivalence between the non-smoothing estimating equations and induced-smoothing estimating equations. By Lemmas 2.1- 2.2, we can infer the asymptotic properties of $\hat{\gamma}$ as follows.

Theorem 2.1. Assume that γ_0 is the unique root of $\mathbf{W}_0(\gamma) = 0$ and Conditions 2.1- 2.6 hold.

- (i) $\hat{\gamma}_n$ satisfying $\widetilde{\mathbf{W}}_n(\hat{\gamma}_n) = o_p(1)$ converges to γ_0 in probability.

- (ii) Suppose that $\mathbf{D}_0(\gamma)$ exists, which is the derivative of $\mathbf{W}_0(\gamma)$ with respect to γ and bounded continuously in a neighborhood of γ_0 . Meanwhile, $\mathbf{D}_0(\gamma_0)$ is nonsingular. If $\hat{\gamma}_n$ satisfies $\widetilde{\mathbf{W}}_n(\hat{\gamma}_n) = o_p(n^{-1/2})$, then $\|\hat{\gamma}_n - \gamma_0\|_2 < O_p(n^{-1/2})$.
- (iii) Assume that conditions in part (ii) hold and $\hat{\gamma}_n$ satisfies $\widetilde{\mathbf{W}}_n(\hat{\gamma}_n) = o_p(n^{-1/2})$. Then $\sqrt{n}(\hat{\gamma}_n - \gamma_0)$ converges in distribution to a zero-mean Gaussian process.

Remark 2.1. Lemmas 2.1- 2.2 and Theorem 2.1 can be extended for $\max \|\mathbf{X}_i\|_2 = O(n^\varrho)$ with a small positive value ϱ .

Remark 2.2. Theorem 2.1 requires the assumption that the mean function has $\mathbf{W}_0(\gamma_0) = 0$. When $\mathbf{W}_0(\gamma_0)$ is far from zero, the estimator obtained from $\widetilde{\mathbf{W}}_n(\gamma) = 0$ is biased. Artificial censoring technique, introduced by Lin et al. (1996), Peng and Fine (2006) and Ding et al. (2009), is suggested to obtain a consistent estimator of γ but requiring one more marginal model. Inferences are provided in these articles. In numerical examples, we still use $\widetilde{\mathbf{W}}_n(\gamma)$ to compute an estimate of γ because of these reasons: 1) The estimating functions involved in artificial censoring are discontinuous without exact solutions and they may have computational difficulties. 2) By our experience, the bias of the estimator yielded from $\widetilde{\mathbf{W}}_n(\gamma)$ has minor impacts on the estimation of the conditional model. Using the existing package is easy to implement.

Lemma 2.3. Under Conditions 2.1 -2.6 and for any $\varepsilon > 0$ and each $k = 0, 1$ or 2 , we have

$$\sup_{|r-r'|+|s-s'|\leq n^{-\varepsilon}} |h^{(k)}(\gamma, s) - h^{(k)}(\gamma', s')| = O(n^{-\varepsilon}),$$

where $h^{(k)}$'s are defined in (2.11) and (2.18).

By virtue of Lemma 2.3, we can show the convergence of $\widehat{Q}_n^{km}$ toward Q^* . Similar to the preceding lemmas and theorem, we consider dependent censoring as well.

Theorem 2.2. Under Conditions 2.1 -2.6, if $H_n^{(1)}(\gamma, t) \geq n^{-\varepsilon}$ for any $0 < \varepsilon \leq \frac{1}{8}$, then

$$\sup_{t \in \{t: H_n^{(1)}(\gamma, t) \geq n^{-\varepsilon}\}} |\widehat{Q}_{n, \gamma}^{km}(t) - Q^*(\gamma, t)| = O_p(n^{-\varepsilon}),$$

where $\gamma \in \mathcal{G}$. Moreover, $\sqrt{n}(\widehat{Q}_{n, \gamma}^{km}(t) - Q^*(\gamma, t))$ converges to a zero-mean Gaussian process within the set $\{t : H_n^{(1)}(\gamma, t) \geq n^{-\varepsilon}\}$.

From Theorem 2.2, we can easily get the following corollary.

Corollary 2.1. Under Conditions 2.1 -2.6 and let $\widehat{\gamma}_n$ be the consistent estimator of γ with a polynomial convergence rate, that is, $\|\widehat{\gamma}_n - \gamma_0\|_2 < n^{-3\varepsilon}$ for any $0 < \varepsilon \leq \frac{1}{8}$, if $H_n^{(1)}(\widehat{\gamma}_n, t) \geq n^{-\varepsilon}$, then

$$\sup_{t \in \{t: H_n^{(1)}(\widehat{\gamma}_n, t) \geq n^{-\varepsilon}, \|\widehat{\gamma}_n - \gamma_0\|_2 < n^{-3\varepsilon}\}} |\widehat{Q}_{n, \widehat{\gamma}_n}^{km}(t) - Q^*(\gamma_0, t)| = O_p(n^{-\varepsilon}).$$

2.4.2 Asymptotic Properties of the Second-stage Estimator

Given the consistency properties in the proposed AFT model, we consider the asymptotic properties of the two-stage estimators $\widehat{\alpha}$ and $\widehat{\beta}$ in Theorems 2.3 and 2.4 under some additional regularity conditions as below.

Condition 2.7. The collection of baseline cumulative hazard functions $\Lambda_0(\cdot)$, which are strictly increasing and continuously differentiable and satisfy $\Lambda_0(\varsigma) < \infty$, is compact convex with true parameter $\Lambda_{0,0}$ falling in its interior.

Condition 2.8. The parameter spaces $\mathcal{A}$ and $\mathcal{B}$ corresponding to α and β , respectively, are both compact. True coefficients α_0 and β_0 fall in the interiors of $\mathcal{A}$ and $\mathcal{B}$, respectively.

Condition 2.9. For those i with $\delta_{1i} = 0$, $C_i(\alpha, \beta, \Lambda_0, \gamma, Q)$ is bounded away from zero, besides, $A_i(\alpha, \beta, \Lambda_0, \gamma, Q)$ and $B_i(\alpha, \beta, \Lambda_0, \gamma, Q)$ defined in (2.15), are both uniformly bounded in $\{(\alpha, \beta, \Lambda_0, \gamma, Q) : |\alpha - \alpha_0| + \|\beta - \beta_0\|_2 + \sup_t |\Lambda_0(t) - \Lambda_{0,0}(t)| + \|\gamma - \gamma_0\|_2 + \sup_t |Q(t) - Q_0(t)| < d_0\}$.

Condition 2.10. The KM estimator of survival function S_c has $\sup_c |\hat{S}_c(c) - S_c(c)| = o_p(1)$.

Condition 2.11. If there exist vectors c and h such that $c[1, \epsilon, \mathbf{X}_i^T]^T = 0$ and $h[1, \mathbf{X}_i^T]^T = 0$ almost surely for any small positive value ϵ and $i = 1, \dots, n$, then $c = 0$ and $h = 0$.

Condition 2.12. For any s belonging to the support of ξ , $K_Q(s; \alpha, \beta, \Lambda_0, \gamma, Q, S_c)$ is uniformly bounded in the neighbourhood of $(\alpha_0, \beta_0, \Lambda_{0,0}, \gamma_0, Q_0, S_c)$. Besides, if there exist two CDFs Q_1 and Q_2 satisfying Condition 2.4 such that $h^{(4)}(s; \alpha, \beta, \Lambda_0, \gamma, Q_1 - Q_2, S_c)dQ_2(s) = h^{(4)}(s; \alpha, \beta, \Lambda_0, \gamma, Q_2, S_c)d(Q_1 - Q_2)(s)$, then $Q_1 = Q_2$.

Similar arguments to Conditions 2.7, 2.8 and 2.11 can be found in Zeng et al. (2005) and Gamst et al. (2009). Condition 2.9 is required for the NPMLE of the baseline hazard function. The consistency of the KM-induced estimator of Q is based on Conditions 2.10 and 2.12.

Let $\hat{\gamma}_n$ and $\hat{Q}_{n, \hat{\gamma}_n}^{km}$ be consistent estimators of true parameters γ_0 and $Q^*(\gamma_0, t)$, respectively, we have the following two theorems.

Theorem 2.3. (Consistency) Under Conditions 2.2 and 2.7-2.12, suppose that $\|\hat{\gamma}_n - \gamma_0\|_2 + \sup_t |\hat{Q}_{n, \hat{\gamma}_n}^{km}(t) - Q^*(\gamma_0, t)| = o_p^*(1)$, where $o_p^*(\cdot)$ represents

convergence in outer probability, and that $(\alpha_0, \beta_0, \Lambda_{0,0})$ is the maximizer of $pl(\alpha, \beta, \Lambda_0; \gamma_0, Q_0)$. If $(\hat{\alpha}_n, \hat{\beta}_n, \hat{\Lambda}_{n,0})$ is the maximizer of $pl_n(\alpha, \beta, \Lambda_0; \hat{\gamma}_n, \hat{Q}_n)$, then $(\hat{\alpha}_n, \hat{\beta}_n, \hat{\Lambda}_{n,0})$ converges in outer probability to $(\alpha_0, \beta_0, \Lambda_{0,0})$. Moreover, $\hat{Q}_n$ converges in outer probability to Q_0 .

Theorem 2.4. (Asymptotic normality) Under Conditions 2.2 and 2.7-2.12, suppose that $\|\hat{\gamma}_n - \gamma_0\|_2 + \sup_t |\hat{Q}_{n,\hat{\gamma}_n}^{km} - Q^*(\gamma_0, t)| = O_p^*(n^{-1/2})$, $\sqrt{n}(\hat{\gamma}_n - \gamma_0, \hat{Q}_{n,\hat{\gamma}_n}^{km} - Q^*(\gamma_0, \cdot))$ converges weakly to a zero-mean Gaussian process, and that $(\alpha_0, \beta_0, \Lambda_{0,0})$ is the maximizer of $pl(\alpha, \beta, \Lambda_0; \gamma_0, Q_0)$, if $pl_n(\hat{\alpha}_n, \hat{\beta}_n, \hat{\Lambda}_{n,0}; \hat{\gamma}_n, \hat{Q}_n) \geq \sup pl_n(\alpha, \beta, \Lambda_0; \gamma, Q) - o_p^*(n^{-1})$, then $\sqrt{n}(\hat{\alpha}_n - \alpha_0, \hat{\beta}_n - \beta_0, \hat{\Lambda}_{n,0} - \Lambda_{0,0})$ converges weakly to a zero-mean Gaussian process.

Remark 2.3. In Theorem 2.4, we have shown that $\sqrt{n}(\hat{\alpha}_n - \alpha_0, \hat{\beta}_n - \beta_0, \hat{\Lambda}_{n,0} - \Lambda_{0,0})$ follows a Gaussian process. But the explicit form of the asymptotic variance can be very complicated. In the numerical analysis, we use the bootstrap to obtain the standard errors of estimates approximately.

2.5 Simulation

We carry out simulation studies to assess the finite sample performance of our proposed method. Two covariates, denoted by X_1 and X_2 , are considered. For each individual i ($i = 1, \dots, n$), covariates X_{1i} and X_{2i} are independently generated from the standard normal distribution truncated at ± 4 and the Bernoulli distribution with probability 0.5, respectively. The error term ξ_i in the AFT model is assumed to follow a standard extreme value distribution. Then V_i is generated from the exponential distribution with rate $= \exp\{-(\gamma_1 X_{1i} + \gamma_2 X_{2i})\}$.

For the terminal event time T_i and the values of the regression coefficients, we consider the following five different scenarios. In Scenarios I-IV,

we set a Weibull baseline cumulative hazard $\Lambda_0(t) = t^2$, while a standard lognormal baseline is set in Scenario V. The censoring time C_i in Scenario III is generated from the uniform distribution between 0 and study duration $\varsigma = 3$, gaining the overall censoring rate for both V and T around 71%, and $C_i \sim U[1, 3]$ is considered in the remaining scenarios, which achieves the overall censoring rate of 60% for the pair (V, T) . Besides, regarding the choices of the values for true coefficients:

Scenario I $\alpha = -0.5$, $\beta = (\beta_1, \beta_2) = (0.5, -0.5)$ and $\gamma = (\gamma_1, \gamma_2) = (0.5, 0.5)$.

Scenarios II/III are same with Scenario I but $\alpha = -1$.

Scenario IV has the same setting with Scenario II but $\gamma = (1, 1)$.

Scenario V $\alpha = -0.5$, $\beta = (0.5, 0.5)$ and $\gamma = (0.5, -0.5)$.

We set the sample size $n = 100, 200$, and 400 . Based on 200 simulation runs, estimated parameters are computed by the proposed method and summarized in terms of the mean bias, standard deviation (SD), mean square error (MSE), and coverage percentage (CP) with nominal level 95%. Results obtained from Scenario I-V are reported in Tables 2.1-2.5, respectively. We evaluate the performance of the bootstrap SE in Table 2.1-2.4, where 100 Bootstrap samples are used. The values of SE are the average of bootstrap standard error estimates over 200 replicates. It can be seen that the SE and the sample SD are close to each other in all simulation cases.

We also consider two naive methods: cCox and Cox.imp in Tables 2.1. cCox represents the complete case analysis, which fits the Cox model for T with the validation sample only by removing censored observations of V . Cox.imp fits the Cox model for T using all observations of V (though

Table 2.1: *Simulation results are shown for Scenario I. The censoring time is generated from $C \sim U(1, 3)$, gaining the censoring rates of 32 ~ 55% for V and 20 ~ 36% for T . The total observed rate for the pair (V, T) is around 42%.*

n	method		$\alpha(-0.5)$	$\beta_1(0.5)$	$\beta_2(-0.5)$	$\gamma_1(0.5)$	$\gamma_2(0.5)$
100	proposed	bias	-0.053	0.001	0.088	0.079	0.045
		SD	0.293	0.143	0.254	0.155	0.296
		SE	0.282	0.164	0.284	0.178	0.336
		MSE	0.089	0.020	0.072	0.030	0.090
		CP	0.950	0.939	0.944	0.939	0.927
	cCox	bias	-0.927	0.057	-0.004		
		SD	0.441	0.211	0.352		
		MSE	1.054	0.048	0.124		
	Cox.imp	bias	-1.009	0.057	0.060		
		SD	0.266	0.142	0.269		
		MSE	1.089	0.023	0.076		
200	proposed	bias	0.001	-0.018	0.064	0.090	0.008
		SD	0.217	0.096	0.190	0.115	0.237
		SE	0.208	0.103	0.190	0.129	0.236
		MSE	0.047	0.010	0.040	0.021	0.056
		CP	0.956	0.940	0.940	0.918	0.956
	cCox	bias	-0.837	0.049	-0.055		
		SD	0.263	0.129	0.243		
		MSE	0.770	0.019	0.062		
	Cox.imp	bias	-0.964	0.051	0.025		
		SD	0.174	0.094	0.197		
		MSE	0.960	0.011	0.039		
400	proposed	bias	0.043	-0.018	0.068	0.092	0.055
		SD	0.189	0.075	0.133	0.087	0.173
		SE	0.170	0.073	0.130	0.089	0.163
		MSE	0.037	0.006	0.022	0.016	0.033
		CP	0.945	0.960	0.915	0.829	0.915
	cCox	bias	-0.820	0.043	-0.026		
		SD	0.191	0.099	0.182		
		MSE	0.709	0.012	0.034		
	Cox.imp	bias	-0.927	0.049	0.051		
		SD	0.124	0.075	0.136		
		MSE	0.875	0.008	0.021		

Table 2.2: *Simulation results are shown for Scenario II. The censoring time is generated from $C \sim U(1, 3)$, gaining the censoring rates of 27 ~ 44% for V and 33 ~ 52% for T . The total observed rate for the pair (V, T) is around 41%.*

n		$\alpha(-1)$	$\beta_1(0.5)$	$\beta_2(-0.5)$	$\gamma_1(0.5)$	$\gamma_2(0.5)$
100	bias	-0.067	0.025	0.047	0.085	0.027
	SD	0.340	0.164	0.305	0.159	0.279
	SE	0.334	0.177	0.312	0.166	0.316
	MSE	0.120	0.027	0.095	0.032	0.079
	CP	0.957	0.957	0.926	0.926	0.936
200	bias	-0.002	0.000	0.000	0.081	-0.010
	SD	0.245	0.104	0.197	0.111	0.224
	SE	0.228	0.111	0.202	0.117	0.224
	MSE	0.060	0.011	0.039	0.019	0.050
	CP	0.960	0.950	0.950	0.905	0.940
400	bias	-0.012	0.006	0.022	0.083	0.034
	SD	0.168	0.078	0.142	0.083	0.166
	SE	0.172	0.078	0.139	0.082	0.153
	MSE	0.028	0.006	0.021	0.014	0.029
	CP	0.950	0.965	0.955	0.835	0.915

Table 2.3: *Simulation results are shown for Scenario III. The censoring time is generated from $C \sim U(0, 3)$, gaining the censoring rates of 38 ~ 58% for V and 49 ~ 66% for T . The total observed rate for the pair (V, T) is around 29%.*

n		$\alpha(-1)$	$\beta_1(0.5)$	$\beta_2(-0.5)$	$\gamma_1(0.5)$	$\gamma_2(0.5)$
100	bias	-0.074	0.021	0.027	0.062	0.035
	SD	0.430	0.212	0.346	0.178	0.317
	SE	0.433	0.221	0.391	0.180	0.350
	MSE	0.190	0.046	0.120	0.036	0.102
	CP	0.949	0.939	0.944	0.949	0.944
200	bias	-0.035	-0.003	0.011	0.067	-0.003
	SD	0.310	0.129	0.239	0.125	0.250
	SE	0.275	0.134	0.246	0.130	0.244
	MSE	0.097	0.017	0.057	0.020	0.062
	CP	0.959	0.943	0.933	0.923	0.943
400	bias	-0.020	0.010	0.039	0.071	0.039
	SD	0.192	0.088	0.176	0.092	0.175
	SE	0.201	0.092	0.169	0.091	0.168
	MSE	0.037	0.008	0.033	0.013	0.032
	CP	0.958	0.932	0.953	0.890	0.937

Table 2.4: *Simulation results are shown for Scenario IV. The censoring time is generated from $C \sim U(1, 3)$, gaining the censoring rates of $34 \sim 58\%$ for V and $42 \sim 60\%$ for T . The total observed rate for the pair (V, T) is around 36% .*

n		$\alpha(-1)$	$\beta_1(0.5)$	$\beta_2(-0.5)$	$\gamma_1(1)$	$\gamma_2(1)$
100	bias	0.016	-0.003	0.020	0.097	0.056
	SD	0.334	0.198	0.330	0.185	0.312
	SE	0.353	0.214	0.351	0.188	0.344
	MSE	0.111	0.039	0.109	0.044	0.101
	CP	0.964	0.943	0.927	0.907	0.933
200	bias	0.042	-0.012	-0.015	0.114	0.014
	SD	0.237	0.126	0.233	0.127	0.240
	SE	0.236	0.137	0.231	0.136	0.244
	MSE	0.058	0.016	0.054	0.029	0.058
	CP	0.960	0.955	0.944	0.879	0.949
400	bias	0.033	-0.009	0.015	0.110	0.052
	SD	0.159	0.092	0.159	0.095	0.177
	SE	0.170	0.093	0.158	0.097	0.170
	MSE	0.026	0.009	0.025	0.021	0.034
	CP	0.935	0.955	0.945	0.789	0.935

Table 2.5: *Simulation results are shown for Scenario V. The censoring time is generated from $C \sim U(1, 3)$, gaining the censoring rates of 25 ~ 42% for V and 25 ~ 42% for T . The total observed rate for the pair (V, T) is around 40%.*

n		$\alpha(-0.5)$	$\beta_1(0.5)$	$\beta_2(0.5)$	$\gamma_1(0.5)$	$\gamma_2(-0.5)$
100	bias	0.001	-0.007	0.089	0.070	-0.057
	SD	0.299	0.153	0.280	0.163	0.321
	MSE	0.089	0.023	0.087	0.032	0.107
	CP	0.975	0.954	0.939	0.924	0.939
200	bias	0.001	0.011	0.043	0.057	-0.028
	SD	0.205	0.102	0.179	0.106	0.217
	MSE	0.042	0.010	0.034	0.015	0.048
	CP	0.954	0.949	0.939	0.914	0.949
400	bias	0.025	-0.002	0.050	0.079	-0.018
	SD	0.166	0.074	0.145	0.080	0.160
	MSE	0.028	0.006	0.024	0.013	0.026
	CP	0.960	0.950	0.950	0.865	0.955

Figure 2.1: Estimation results for true baseline cumulative hazard function $\Lambda_0(t) = t^2$ and survival function $S_0(t) = \exp(-t^2)$ under Scenario I with $n = 100, 200, 400$. The black solid lines are the estimates, the red dot-dashed lines are the true ones, and the black dashed lines are the 95% upper/lower bounds over 200 replicates.

some may be censored) as if they are all exactly observed. Both the naive methods yield larger biases and standard deviations for the coefficient α of V in comparison with the proposed method.

From Tables 2.1- 2.5, to sum up, it can be seen that under different scenarios, biases of the estimates in the conditional model are all very small with CPs nearly the nominal level 0.95. As the sample size increases, SD and MSE both decrease. These results agree with the theoretical justification. Although the biases and CPs of the estimates of γ become worse as n increases, which is in line with observations in Lin et al. (1996), it slightly affects the estimations of α and β of our interest. To be specific, Tables 2.1 and 2.2 compare the performance of the proposed method under different impacts of V on T . Estimation results of scenarios with different censoring rates are shown in Tables 2.2 and 2.3. When expanding the effects of X_1 and X_2 on V as Table 2.4 shows, it yields worse estimates of γ , but the performances of estimators of α and β , are still well compared with Table 2.2. Under the setting in scenario V, the Cox proportional hazards model is no longer equivalent to the AFT model. Table 2.5 shows that the estimators of α and β still perform well regardless of the estimates of γ .

We further illustrate the versatility of the proposed method in estimating the baseline hazard. The mean of the estimated baseline cumulative hazard and survival curves together with 2.5% and 97.5% quantiles are presented in Figure 2.1 for Scenarios I respectively. The figure shows that the estimated baseline cumulative hazard curves and survival curves are all close to the truth and their biases decrease as sample size n increases. The estimates of baseline cumulative hazard and survival functions exhibit similar trends in Scenarios II-V. Details are omitted to save space.

2.6 Real Data Analysis: Breast Cancer Data

We present an application example to assess the contribution of the non-terminal event to the overall survival of patients in a study of breast cancer (Haibe-Kains et al., 2012). The distant metastatic is a known strong prognostic factor for survival and thus it is important to assess its impact on the risk of death for breast cancer patients.

Table 2.6: *Estimation results of conditional modeling for the breast cancer data.*

	proposed		cCox		Cox.imp	
	Estimate	SE	Estimate	SE	Estimate	SE
DMFS(α)	-0.305**	0.052	-1.049**	0.162	-0.856**	0.078
age(β_1)	0.451**	0.127	-0.068	0.200	0.310**	0.113
er(β_2)	-0.442*	0.242	-0.864**	0.353	-0.734**	0.237
node(β_3)	0.209	0.256	0.135	0.344	0.383*	0.224
size(β_4)	0.133	0.124	-0.219	0.270	0.132	0.116

significance level: * indicates 0.1 and ** indicates 0.05.

cCox: Cox model estimation based on complete data.

Cox.imp: Cox model estimation with ignoring the censoring of the non-terminal event time.

A total of 1238 breast cancer patients were collected from 6 databases including database CAL (118 patients), NKI (337 patients), STNO2 (118 patients), TRANSBIG (198 patients), UCSF (162 patients) and UNC4 (306 patients). In the study, the distant metastasis-free survival is treated as a non-terminal event, while overall survival is the terminal event. We aim to explore how changes in distant metastasis-free survival (years) affect

Figure 2.2: *Estimation of the baseline cumulative hazard function (solid curve) together with its 95% upper and lower bounds (dashed curves) for the breast cancer data.*

Figure 2.3: *Goodness of fit of the Cox proportional hazards model regarding binary covariates for the breast cancer data.*

Figure 2.4: *Goodness of fit of the Cox proportional hazards model regarding continuous covariates for the breast cancer data.*

overall survival (years). Our analysis focuses on a subset of 250 patients who have valid times for DMFS, OS and potential factors, who were 50 years or older at diagnose, and whose tumor sizes are bigger than 2 cm at baseline. Among data from these patients, censoring rates are 69.2% for DMFS and 60.8% for OS. As a result, there are overall 81.2% observations censored for either DMFS or OS or both. Besides DFMS, other available Risk factors include: patients' age, estrogen receptor status (er for short), nodal status (node for short), tumor size (per cm). Age of all patients ranges from 28 to 89, while tumor size is observed within 2.1 cm~7 cm. In our estimation, we standardize the variable age by subtracting its mean and dividing with standard deviation. The standardization is also performed for variable tumor size.

We fit the Cox model to predict the hazard ratio of death based on the time for distant metastasis subject to censoring and potential risk factors. Figure 2.2 presents the estimated baseline cumulative hazard function and its 95% upper and lower bounds. In Table 2.6, we compare the results produced by the proposed method with those from two naive methods. SE is obtained based on 400 Bootstrap samples. Coefficient α associated with the DMFS is significant, which indicates the important impact of DMFS on the risk of death. The estimated value of α shows that 1-year increase in DMFS results in around 26% decrease in the relative risk of dying. When considering the complete-case Cox model (cCox) and the Cox model with ignoring censoring for DMFS (Cox.imp) in Table 2.6, both models having loss of information produce inflated estimates for some significant regression coefficients in comparison with the proposed one. These results are consistent with the empirical evidences obtained from the original data set that there are 47 individuals died after distant metastasis. Further inves-

tigations on this subgroup show that the earlier the patient experiences distant metastasis, the sooner she dies. Specifically, the median survival time among those patients with DMFS below median level is 1.71 years, while the median survival time for women with DMFS above median level is 5.51 years. In all different types of models considered, factors size appears insignificant.

The adequacy of the proportional hazards model for the breast cancer data is assessed in Figures 2.3 and 2.4. For binary covariates *er* and *node*, we draw the logarithm transformation of the baseline cumulative hazard function under each stratum in Figure 2.3. No gross departure from the hypothesis of parallel curves for covariate *er* suggests that the proportional hazards model for covariate *er* is reasonable. We see that there is a close distance between curves for covariate *node*, which also implies that covariate *node* is insignificant. For continuous covariates *age* and *size*, we consider score processes to verify the reasonability of the proportional hazards model as Figure 2.4 shows. The empirical score processes are indicated with solid black lines which randomly fluctuate around zero. 20 bootstrapped processes with dashed red lines show similar patterns with their corresponding observed score processes. These suggest that the proportional hazards model is reasonably well.

2.7 Summary

In this chapter, we propose a new method for analysis of semi-competing risks data by handling the nonterminal event time as a covariate subject to dependent or independent censoring along with other risk factors in a regression model for the terminal event time. Compared to the existing

joint modeling method for the semi-competing risks data, the impact of the nonterminal event time on the terminal event time or overall survival can be interpreted explicitly via the estimated coefficient corresponding to the dependently censored covariate. We propose a semiparametric AFT model for the intermediate event first to estimate its distribution through residuals so that a pseudo-likelihood method can be built-up in the second stage. Such a two-stage estimation procedure has been developed for estimating the model parameters and unknown functionals along with an easy-to-implement algorithm. We have shown the consistency and asymptotic normality of the resulting estimator.

2.8 Appendix

The proofs of Theorem 2.1 and Theorem 2.4 are mainly based on the techniques of empirical process theory, including the Donsker class and Glivenko-Cantelli class (Van Der Vaart and Wellner, 1996). As such, we give their definitions first. Let the empirical process $\mathbb{G}_n(f) = \sqrt{n}(\mathbb{P}_n f - Pf)$. A class $\mathcal{F}$ is a Donsker if $\mathbb{G}_n$ weakly converges to $\mathbb{G}$ in $L^\infty(\mathcal{F})$ for a tight Borel measurable element $\mathbb{G}$ in $L^\infty(\mathcal{F})$. A class $\mathcal{F}$ is a Glivenko-Cantelli if $\sup\{|\mathbb{P}_n f - Pf| : f \in \mathcal{F}\} \rightarrow 0$ almost surely. It can be shown that a Donsker class implies a Glivenko-Cantelli class almost surely.

2.8.1 Proofs of Lemmas 2.1-2.2 and Theorem 2.1

Proof of Lemma 2.1. The proof of this lemma is quite similar to that of Proposition 2.3.4 (1) in Ding (2010). Nan et al. (2009) also provided some similar results in the different model framework.

The class of functions $\{(\boldsymbol{\gamma}, s) : e(\boldsymbol{\gamma}) \geq s\}$ is a collection of finite-dimensional half spaces, so it is a VC-class which is named after Vapnik and Červonenkis. It follows that the collection of indicators $\{I(e(\boldsymbol{\gamma}) \geq s)\}$ is also a VC-class of the same index with $\{(\boldsymbol{\gamma}, s) : e(\boldsymbol{\gamma}) \geq s\}$. Besides, the envelope function f for $\{I(e(\boldsymbol{\gamma}) \geq s)\}$ is actually an identity function, thus measurable square integrable i.e. $Pf^2 < \infty$. By Theorem 2.5.2, 2.6.7 and 2.6.8 in Van Der Vaart and Wellner (1996), $\{I(e(\boldsymbol{\gamma}) \geq s)\}$ is a Donsker class. Since $\mathbf{X}$ is uniformly bounded, (see Theorem 2.6.8, Theorem 2.10.6 and Example 2.10.10 in Van Der Vaart and Wellner (1996)), the class of functions $\{I(e(\boldsymbol{\gamma}) \geq s)\mathbf{X}\}$ is also a Donsker class. By the definition, $H_n^{(1)}$ and $h^{(1)}$ are convex hulls of the closure of $\{I(e(\boldsymbol{\gamma}) \geq s)\}$, and $H_n^{(2)}$ and $h^{(2)}$ are convex hulls of the closure of $\{I(e(\boldsymbol{\gamma}) \geq s)\mathbf{X}\}$. Therefore, $H_n^{(k)}$ and $h^{(k)}$

($k = 1, 2$) are all Donsker classes and thus Glivenko-Cantelli classes (by Theorems 2.10.2 and 2.10.3 in (Van Der Vaart and Wellner, 1996)). Then with probability one

$$\begin{aligned} \|H_n^{(1)} - h^{(1)}\|_\infty &= \|(\mathbb{P}_n - P)I(e(\gamma) \geq s)\|_\infty \longrightarrow 0, \\ \|H_n^{(2)} - h^{(2)}\|_\infty &= \|(\mathbb{P}_n - P)I(e(\gamma) \geq s)\mathbf{X}\|_\infty \longrightarrow 0. \end{aligned}$$

Due to $\|P(\delta_1 \mathbf{X})\|_\infty < \infty$ and $\|P(\delta_1)\|_\infty \leq 1$, we get

$$\begin{aligned} \|P\{[H_n^{(1)} - h^{(1)}]\delta_1 \mathbf{X}\}\|_\infty &\leq \|H_n^{(1)} - h^{(1)}\|_\infty \|P(\delta_1 \mathbf{X})\|_\infty \longrightarrow 0, \\ \|P\{[H_n^{(2)} - h^{(2)}]\delta_1\}\|_\infty &\leq \|H_n^{(2)} - h^{(2)}\|_\infty \|P(\delta_1)\|_\infty \longrightarrow 0. \end{aligned}$$

On the other hand, since δ_1 can be viewed as a random variable following a Bernoulli distribution, δ_1 has finite second moment. So $\{(H_n^{(1)} - H_n^{(2)})\delta_1\}$ is a Donsker class and thus a Glivenko-Cantelli class. This means that $\|(\mathbb{P}_n - P)\{(H_n^{(1)} - H_n^{(2)})\delta_1\}\|_\infty \longrightarrow 0$ with probability one. Consequently, the convergence resulting in Lemma 2.1 follows from the fact

$$\begin{aligned} &\|\mathbf{W}_n(\gamma, H_n^{(1)}, H_n^{(2)}) - \mathbf{W}_0(\gamma, h^{(1)}, h^{(2)})\|_\infty \\ &\leq \|(\mathbb{P}_n - P)\{(H_n^{(1)} - H_n^{(2)})\delta_1\}\|_\infty \\ &+ \|P\{[H_n^{(1)} - h^{(1)}]\delta_1 \mathbf{X}\}\|_\infty + \|P\{[H_n^{(2)} - h^{(2)}]\delta_1\}\|_\infty. \end{aligned}$$

□

Proof of Lemma 2.2. (i) Note that

$$\begin{aligned} E\left[\widetilde{\mathbf{W}}_n(\gamma) - \mathbf{W}_n(\gamma)\right] &= E\left\{E\left[\widetilde{\mathbf{W}}_n(\gamma) - \mathbf{W}_n(\gamma)|\mathbf{X}\right]\right\} \\ &= \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \int_{\mathcal{X}^2} (\mathbf{X}_i - \mathbf{X}_j) \\ &\quad \times E\left\{\delta_{1i} \Phi\left[\frac{e_j(\gamma) - e_i(\gamma)}{\sigma_{ji}}\right] - \delta_{1i} I[e_j(\gamma) - e_i(\gamma) \geq 0]\right\} f_X(\mathbf{X}_i) f_X(\mathbf{X}_j) d\mathbf{X}_i d\mathbf{X}_j, \end{aligned}$$

where $\mathcal{X}$ is the parameter space for variable $\mathbf{X}$ and

$$\begin{aligned}\delta_{1i} &= I[V_i \leq \min(T_i, C_i)] = I[V_i \leq T_i] \times I[V_i \leq C_i] \\ &= I[\log V_i - \gamma_0^T \mathbf{X}_i \leq \log T_i - \gamma_0^T \mathbf{X}_i] \times I[\log V_i - \gamma_0^T \mathbf{X}_i \leq \log C_i - \gamma_0^T \mathbf{X}_i] \\ &:= I[\xi_i \leq \tilde{T}_i] \times I[\xi_i \leq \tilde{C}_i],\end{aligned}$$

$\xi_i = \log V_i - \gamma_0^T \mathbf{X}_i$, $\tilde{T}_i = \log T_i - \gamma_0^T \mathbf{X}_i$ and $\tilde{C}_i = \log C_i - \gamma_0^T \mathbf{X}_i$. We use $g_{T|\xi}^{(i)}(\cdot|\cdot)$ and $G_{T|\xi}^{(i)}(\cdot|\cdot)$ to represent the conditional distribution functions of $\tilde{T}_i$ given $(\xi_i, \mathbf{X}_i)$.

Let $a(\gamma) = (\gamma_0 - \gamma)^T(\mathbf{X}_i - \mathbf{X}_j)$, we have

$$\begin{aligned}& I[e_i(\gamma) - e_j(\gamma) \leq 0] \\ &= I[\log[\min(V_i, T_i, C_i)] - \gamma^T \mathbf{X}_i - \log[\min(V_j, T_j, C_j)] + \gamma^T \mathbf{X}_j \leq 0] \\ &= I[\log[\min(V_i, T_i, C_i)] - \gamma_0^T \mathbf{X}_i - \log[\min(V_j, T_j, C_j)] + \gamma_0^T \mathbf{X}_j + a(\gamma) \leq 0] \\ &= I[\log[\min(V_i, T_i, C_i)] - \gamma_0^T \mathbf{X}_i + a(\gamma) \leq \xi_j] \\ &\times I[\log[\min(V_i, T_i, C_i)] - \gamma_0^T \mathbf{X}_i + a(\gamma) \leq \tilde{T}_j] \\ &\times I[\log[\min(V_i, T_i, C_i)] - \gamma_0^T \mathbf{X}_i + a(\gamma) \leq \tilde{C}_j].\end{aligned}$$

Consequently,

$$\begin{aligned}& E\{\delta_{1i} I[e_j(\gamma) - e_i(\gamma) \geq 0]\} \\ &= E\left\{I[\xi_i \leq \tilde{T}_i] I[\xi_i \leq \tilde{C}_i] I[\xi_i + a(\gamma) \leq \xi_j] I[\xi_i + a(\gamma) \leq \tilde{C}_j] I[\xi_i + a(\gamma) \leq \tilde{T}_j]\right\} \\ &= \int_{-\infty}^{\infty} \bar{G}_C^{(i)}(u) \bar{G}_{T|\xi}^{(i)}(u|\xi = u) \bar{G}_C^{(j)}(u + a(\gamma)) \bar{G}_{\xi, T}^{(j)}(u + a(\gamma), u + a(\gamma)) dQ(u),\end{aligned}$$

where $\bar{G}_C^{(i)}(u) = 1 - G_C^{(i)}(u)$, $\bar{G}_{T|\xi}^{(i)}(u) = 1 - G_{T|\xi}^{(i)}(u)$ and $\bar{G}_{\xi, T}^{(i)}(u) = 1 - G_{\xi, T}^{(i)}(u)$, $G_C^{(i)}(\cdot)$ and $Q(\cdot)$ are the CDFs of $\tilde{C}_i$ and ξ_i , respectively. $G_{T|\xi}^{(i)}(\cdot)$ and $G_{\xi, T}^{(i)}(\cdot)$ are the joint and conditional CDF of ξ_i and $\tilde{T}_i$. Define

$$m_j(s) = g_C^{(j)}(s) \bar{G}_{\xi, T}^{(j)}(s, s) + G_{\xi, T}^{(j)(1)}(s, s) \bar{G}_C^{(j)}(s) + G_{\xi, T}^{(j)(2)}(s, s) \bar{G}_C^{(j)}(s),$$

where $G_{\xi,T}^{(j)(1)}(s, s)$ and $G_{\xi,T}^{(j)(2)}(s, s)$ are the first-order partial derivatives of $G_{\xi,T}^{(j)}(s, s)$ with respect to ξ and $\tilde{T}$, respectively. Then

$$\begin{aligned} & E \left\{ \delta_{1i} \Phi \left[\frac{e_j(\gamma) - e_i(\gamma)}{\sigma_{ji}} \right] \right\} \\ &= \iint_{\mathbb{R}^2} m_j(s) \Phi \left(\frac{s - u - a(\gamma)}{\sigma_{ji}} \right) \overline{G}_C^{(i)}(u) \overline{G}_{T|\xi}^{(i)}(u|\xi = u) dQ(u) ds \\ &= \iint_{\mathbb{R}^2} m_j(s(u, t)) \Phi(t) \overline{G}_C^{(i)}(u) \overline{G}_{T|\xi}^{(i)}(u|\xi = u) \sigma_{ji} dt dQ(u), \end{aligned}$$

where $s(u, t) = u + a(\gamma) + \sigma_{ji}t$. By multiple applications of integration by parts, we have

$$\begin{aligned} & \int_{\mathbb{R}} m_j(s(u, t)) \Phi(t) \sigma_{ji} dt \\ &= - \int_{\mathbb{R}} \overline{G}_{\xi,T}^{(j)}(s(u, t), s(u, t)) \Phi(t) d\overline{G}_C^{(j)}(s(u, t)) \\ & \quad - \int_{\mathbb{R}} \overline{G}_C^{(j)}(s(u, t)) \Phi(t) d\overline{G}_{\xi,T}^{(j)}(s(u, t), s(u, t)) \\ &= \int_{\mathbb{R}} \overline{G}_C^{(j)}(s(u, t)) \left[\overline{G}_{\xi,T}^{(j)}(s(u, t), s(u, t)) \Phi(t) \right]' dt \\ & \quad + \int_{\mathbb{R}} \overline{G}_{\xi,T}^{(j)}(s(u, t), s(u, t)) \left[\overline{G}_C^{(j)}(s(u, t)) \Phi(t) \right]' dt \\ &= 2 \int_{\mathbb{R}} \overline{G}_C^{(j)}(s(u, t)) \overline{G}_{\xi,T}^{(j)}(s(u, t), s(u, t)) \phi(t) dt \tag{2.21} \\ & \quad - \int_{\mathbb{R}} g_C^{(j)}(s(u, t)) G_{\xi,T}^{(j)}(s(u, t), s(u, t)) \Phi(t) \sigma_{ji} dt \\ & \quad - \int_{\mathbb{R}} \overline{G}_C^{(j)}(s(u, t)) G_{\xi,T}^{(j)(1)}(s(u, t), s(u, t)) \Phi(t) \sigma_{ji} dt \\ & \quad - \int_{\mathbb{R}} \overline{G}_C^{(j)}(s(u, t)) G_{\xi,T}^{(j)(2)}(s(u, t), s(u, t)) \Phi(t) \sigma_{ji} dt \\ &= 2 \int_{\mathbb{R}} \overline{G}_C^{(j)}(s(u, t)) \overline{G}_{\xi,T}^{(j)}(s(u, t), s(u, t)) \phi(t) dt \\ & \quad - \int_{\mathbb{R}} m_j(s(u, t)) \Phi(t) \sigma_{ji} dt. \end{aligned}$$

By changing the order of equation (2.21), we obtain

$$\int_{\mathbb{R}} m_j(s(u, t)) \Phi(t) \sigma_{ji} dt = \int_{\mathbb{R}} \overline{G}_C^{(j)}(s(u, t)) \overline{G}_{\xi,T}^{(j)}(s(u, t), s(u, t)) \phi(t) dt,$$

which leads to

$$\begin{aligned} & E \left\{ \delta_{1i} \Phi \left[\frac{e_j(\gamma) - e_i(\gamma)}{\sigma_{ji}} \right] \right\} \\ &= \iint_{\mathbb{R}^2} \overline{G}_C^{(j)}(s(u, t)) \overline{G}_{\xi, T}^{(j)}(s(u, t), s(u, t)) \overline{G}_C^{(i)}(u) \overline{G}_{T|\xi}^{(i)}(u|u) \phi(t) d\mathbf{d}Q(u) \end{aligned}$$

Applying the second-order Taylor expansion around $t = 0$,

$$\begin{aligned} & \overline{G}_C^{(j)}(s(u, t)) \overline{G}_{\xi, T}^{(j)}(s(u, t), s(u, t)) \\ &= \overline{G}_C^{(j)}(u + a(\gamma)) \overline{G}_{\xi, T}^{(j)}(u + a(\gamma), u + a(\gamma)) - m_j(u + a(\gamma)) \sigma_{ji} t \\ & \quad - \frac{1}{2} m_j'(u + a(\gamma)) \sigma_{ji}^2 t^2 - \frac{1}{6} m_j''(u + a(\gamma) + \sigma_{ji} t^*) \sigma_{ji}^3 t^3, \end{aligned}$$

where t^* is between 0 and t . Define

$$\begin{aligned} J_j(s) &= g_C^{(j)}(s) \overline{G}_{\xi, T}^{(j)}(s, s) - 2g_C^{(j)}(s) \left(G_{\xi, T}^{(j)(1)}(s, s) + G_{\xi, T}^{(j)(2)}(s, s) \right) \\ & \quad + \overline{G}_C^{(j)}(s) \left(G_{\xi, T}^{(j)(11)}(s, s) + 2G_{\xi, T}^{(j)(12)}(s, s) + G_{\xi, T}^{(j)(22)}(s, s) \right) \end{aligned}$$

with the second-order partial derivatives $G_{\xi, T}^{(j)(11)}(s, s)$, $G_{\xi, T}^{(j)(12)}(s, s)$ and $G_{\xi, T}^{(j)(22)}(s, s)$ of $G_{\xi, T}^{(j)}(s, s)$ with respect to (ξ, ξ) , $(\xi, \tilde{T})$ and $(\tilde{T}, \tilde{T})$, respectively. Then,

$$\begin{aligned} & E \left\{ \delta_{1i} \Phi \left[\frac{e_j(\gamma) - e_i(\gamma)}{\sigma_{ji}} \right] \right\} \\ &= \iint_{\mathbb{R}^2} \overline{G}_C^{(j)}(u + a(\gamma)) \overline{G}_{\xi, T}^{(j)}(u + a(\gamma), u + a(\gamma)) \overline{G}_C^{(i)}(u) \overline{G}_{T|\xi}^{(i)}(u|\xi = u) \phi(t) d\mathbf{d}Q(u) \\ & \quad - \sigma_{ji} \iint_{\mathbb{R}^2} m_j(u + a(\gamma)) \overline{G}_C^{(i)}(u) \overline{G}_{T|\xi}^{(i)}(u|\xi = u) \phi(t) t d\mathbf{d}Q(u) \\ & \quad - \frac{\sigma_{ji}^2}{2} \iint_{\mathbb{R}^2} J_j(u + a(\gamma)) \overline{G}_C^{(i)}(u) \overline{G}_{T|\xi}^{(i)}(u|\xi = u) \phi(t) t^2 d\mathbf{d}Q(u) \\ & \quad - \frac{\sigma_{ji}^3}{6} \iint_{\mathbb{R}^2} J_j'(u + a(\gamma) + \sigma_{ji} t^*) \overline{G}_C^{(i)}(u) \overline{G}_{T|\xi}^{(i)}(u|\xi = u) \phi(t) t^3 d\mathbf{d}Q(u). \end{aligned} \tag{2.22}$$

Note that the second and fourth terms in (2.22) are zeros since $\int t \phi(t) dt = \int t^3 \phi(t) dt = 0$. Let

$$\eta_{ij}^{(1)} = -\frac{\sigma_{ji}^2}{2} \int_{\mathbb{R}} J_j(u + a(\gamma) + \sigma_{ji} t^*) \overline{G}_C^{(i)}(u) \overline{G}_{T|\xi}^{(i)}(u|\xi = u) dQ(u),$$

$$E \left\{ \delta_{1i} \Phi \left[\frac{e_j(\boldsymbol{\gamma}) - e_i(\boldsymbol{\gamma})}{\sigma_{ji}} \right] - \delta_{1i} I[e_j(\boldsymbol{\gamma}) - e_i(\boldsymbol{\gamma}) \geq 0] \right\} = \eta_{ij}^{(1)}.$$

Since distribution functions and their first-order and second-order derivatives involved in $\eta_{ij}^{(1)}$ are all bounded, there exists a constant $M_\eta^{(1)} > 0$ such that $|\eta_{ij}^{(1)}| < M_\eta^{(1)} \sigma_{ji}^2$. It follows that $\left\| E \left[\widetilde{\mathbf{W}}_n(\boldsymbol{\gamma}) - \mathbf{W}_n(\boldsymbol{\gamma}) \right] \right\|_2$ is bounded by

$$\begin{aligned} & n^{-2} M_\eta^{(1)} \sum_{i=1}^n \sum_{j=1}^n \int_{\mathcal{X}^2} \|\mathbf{X}_i - \mathbf{X}_j\|_2 \sigma_{ji}^2 f_X(\mathbf{X}_i) f_X(\mathbf{X}_j) d\mathbf{X}_i d\mathbf{X}_j \\ & < n^{-3} M_\eta^{(1)} \sum_{i=1}^n \sum_{j=1}^n \|\mathbf{X}_i - \mathbf{X}_j\|_2^3 \leq o(n^{-1}). \end{aligned}$$

By Chebyshev's inequality, $n \|\widetilde{\mathbf{W}}_n(\boldsymbol{\gamma}) - \mathbf{W}_n(\boldsymbol{\gamma})\|_\infty \rightarrow 0$ with probability one. The proof of part (i) is complete.

(ii) To show result in part (ii), we write

$$\begin{aligned} E[\mathbf{W}_n(\boldsymbol{\gamma})] &= \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \int_{\mathcal{X}^2} (\mathbf{X}_i - \mathbf{X}_j) \\ &\quad \times \int_{-\infty}^{\infty} \overline{G}_C^{(i)}(u) \overline{G}_{T|\xi}^{(i)}(u|\xi = u) \overline{G}_C^{(j)}(u + a(\boldsymbol{\gamma})) \overline{G}_{\xi,T}^{(j)}(u + a(\boldsymbol{\gamma}), u + a(\boldsymbol{\gamma})) dQ(u) \\ &\quad \times f_X(\mathbf{X}_i) f_X(\mathbf{X}_j) d\mathbf{X}_i d\mathbf{X}_j. \end{aligned} \tag{2.23}$$

Differentiating the right hand side of equation (2.23) with respect to $\boldsymbol{\gamma}$, we can get

$$\begin{aligned} \mathbf{D}_n(\boldsymbol{\gamma}) &= \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \int_{\mathcal{X}^2} (\mathbf{X}_i - \mathbf{X}_j)(\mathbf{X}_i - \mathbf{X}_j)^T \\ &\quad \times \int_{-\infty}^{\infty} \overline{G}_C^{(i)}(u) \overline{G}_{T|\xi}^{(i)}(u|u) m_j(u + a(\boldsymbol{\gamma})) dQ(u) f_X(\mathbf{X}_i) f_X(\mathbf{X}_j) d\mathbf{X}_i d\mathbf{X}_j, \end{aligned}$$

$$\begin{aligned} \mathbf{D}_n(\gamma_0) &= \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \int_{\mathcal{X}^2} (\mathbf{X}_i - \mathbf{X}_j)(\mathbf{X}_i - \mathbf{X}_j)^T \\ &\quad \times \int_{-\infty}^{\infty} \bar{G}_C^{(i)}(u) \bar{G}_{T|\xi}^{(i)}(u|u) m_j(u) dQ(u) f_X(\mathbf{X}_i) f_X(\mathbf{X}_j) d\mathbf{X}_i d\mathbf{X}_j. \end{aligned} \quad (2.24)$$

Note that

$$\begin{aligned} \widetilde{\mathbf{D}}_n(\gamma_0) &= \left. \frac{\partial \widetilde{\mathbf{W}}_n(\gamma)}{\partial \gamma} \right|_{\gamma=\gamma_0} \\ &= \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \int_{\mathcal{X}^2} \delta_{1i}(\mathbf{X}_i - \mathbf{X}_j)(\mathbf{X}_i - \mathbf{X}_j)^T \sigma_{ji}^{-1} \phi \left[\frac{e_j(\gamma_0) - e_i(\gamma_0)}{\sigma_{ji}} \right] f_X(\mathbf{X}_i) f_X(\mathbf{X}_j) d\mathbf{X}_i d\mathbf{X}_j. \end{aligned} \quad (2.25)$$

To calculate the expectation of $\widetilde{\mathbf{D}}_n(\gamma_0)$ in (2.25), we have

$$\begin{aligned} E \left[\widetilde{\mathbf{D}}_n(\gamma_0) \right] &= \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \int_{\mathcal{X}^2} (\mathbf{X}_i - \mathbf{X}_j)(\mathbf{X}_i - \mathbf{X}_j)^T \sigma_{ji}^{-1} \\ &\quad \times E \left\{ \delta_{1i} \phi \left[\frac{e_j(\gamma_0) - e_i(\gamma_0)}{\sigma_{ji}} \right] \right\} f_X(\mathbf{X}_i) f_X(\mathbf{X}_j) d\mathbf{X}_i d\mathbf{X}_j, \end{aligned} \quad (2.26)$$

where

$$\begin{aligned} &E \left\{ \delta_{1i} \phi \left[\frac{e_j(\gamma_0) - e_i(\gamma_0)}{\sigma_{ji}} \right] \right\} \\ &= \iint_{\mathbb{R}^2} \bar{G}_C^{(j)}(u + \sigma_{ji}t) \bar{G}_{\xi,T}^{(j)}(u + \sigma_{ji}t, u + \sigma_{ji}t) \bar{G}_C^{(i)}(u) \bar{G}_{T|\xi}^{(i)}(u|u) \phi'(t) dt dQ(u) \\ &= \iint_{\mathbb{R}^2} -\bar{G}_C^{(j)}(u + \sigma_{ji}t) \bar{G}_{\xi,T}^{(j)}(u + \sigma_{ji}t, u + \sigma_{ji}t) \bar{G}_C^{(i)}(u) \bar{G}_{T|\xi}^{(i)}(u|u) t \phi(t) dt dQ(u) \\ &= \iint_{\mathbb{R}^2} -\bar{G}_C^{(j)}(u) \bar{G}_{\xi,T}^{(j)}(u, u) \bar{G}_C^{(i)}(u) \bar{G}_{T|\xi}^{(i)}(u|u) t \phi(t) dt dQ(u) \\ &\quad + \sigma_{ji} \iint_{\mathbb{R}^2} m_j(u) \bar{G}_C^{(i)}(u) \bar{G}_{T|\xi}^{(i)}(u|u) t^2 \phi(t) dt dQ(u) \\ &\quad + \frac{\sigma_{ji}^2}{2} \iint_{\mathbb{R}^2} J_j(u) \bar{G}_C^{(i)}(u) \bar{G}_{T|\xi}^{(i)}(u|u) t^3 \phi(t) dt dQ(u) \\ &\quad + \frac{\sigma_{ji}^3}{6} \iint_{\mathbb{R}^2} J_j'(u + \sigma_{ji}t^*) \bar{G}_C^{(i)}(u) \bar{G}_{T|\xi}^{(i)}(u|u) t^4 \phi(t) dt dQ(u) \end{aligned} \quad (2.27)$$

The first and third terms in the last equation of (2.27) are both zeros. Thus combining (2.24), (2.26) and (2.27) we obtain

$$E \left[\widetilde{\mathbf{D}}_n(\gamma_0) \right] - \mathbf{D}_n(\gamma_0) = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \int_{\mathcal{X}^2} (\mathbf{X}_i - \mathbf{X}_j)(\mathbf{X}_i - \mathbf{X}_j)^T \eta_{ij}^{(2)} f_X(\mathbf{X}_i) f_X(\mathbf{X}_j) d\mathbf{X}_i d\mathbf{X}_j,$$

where

$$\eta_{ij}^{(2)} = \frac{\sigma_{ji}^2}{2} \int_{\mathbb{R}} J_j'(u + \sigma_{ji} t^*) \overline{G}_C^{(i)}(u) \overline{G}_{T|\xi}^{(i)}(u|u) dQ(u).$$

Similar to the argument for $\eta_{ij}^{(1)}$, given that the distribution functions and their first/second/third-order derivatives involved in $\eta_{ij}^{(2)}$ are all bounded, there exists a large enough constant $M_\eta^{(2)} > 0$ such that $|\eta_{ij}^{(2)}| < M_\eta^{(2)} \sigma_{ji}^2$. Given Conditions 2.2 and 2.3 that all X_i are bounded and $\Gamma_n = O(n^{-1})$, it follows from $\sigma_{ji}^2 = (\mathbf{X}_i - \mathbf{X}_j)^T \Gamma_n (\mathbf{X}_i - \mathbf{X}_j) = O(n^{-1})$ that $E \left[\widetilde{\mathbf{D}}_n(\gamma_0) \right] - \mathbf{D}_n(\gamma_0) < O(n^{-1})$. However, by the strong law of large numbers, $\widetilde{\mathbf{D}}_n(\gamma_0) - E \left[\widetilde{\mathbf{D}}_n(\gamma_0) \right] \rightarrow \mathbf{0}$ with probability one. Therefore, $\widetilde{\mathbf{D}}_n(\gamma_0) - \mathbf{D}_n(\gamma_0) \rightarrow \mathbf{0}$ with probability one. $\square$

Proof of Theorem 2.1. (i) By Lemmas 2.1, 2.2(i) and the triangle inequality, for any $\gamma \in \mathcal{G}$, we have

$$\begin{aligned} & \| \widetilde{\mathbf{W}}_n(\gamma) - \mathbf{W}_0(\gamma) \|_\infty \\ & \leq \| \widetilde{\mathbf{W}}_n(\gamma) - \mathbf{W}_n(\gamma) \|_\infty + \| \mathbf{W}_n(\gamma) - \mathbf{W}_0(\gamma) \|_\infty \\ & = o_p(n^{-1}) + o_p(1) = o_p(1). \end{aligned}$$

Thereby, $\sup_{\gamma \in \mathcal{G}} \| \widetilde{\mathbf{W}}_n(\gamma) - \mathbf{W}_0(\gamma) \|_\infty \rightarrow 0$ with probability one. Since γ_0 is the unique root of $\mathbf{W}_0(\gamma) = 0$, according to Theorem 5.9 in Van der Vaart (2000), $\hat{\gamma}_n$ satisfying $\widetilde{\mathbf{W}}_n(\hat{\gamma}_n) = o_p(1)$ converges to γ_0 in probability.

(ii) By Lemma 2.2 (i), $o_p(n^{-1/2}) = \widetilde{\mathbf{W}}_n(\hat{\gamma}_n) = \mathbf{W}_n(\hat{\gamma}_n) + o_p(n^{-1})$. Consequently, $\mathbf{W}_n(\hat{\gamma}_n) = o_p(n^{-1/2})$. From Proposition 2.2.4(2) in Ding (2010), we obtain that $\| \hat{\gamma}_n - \gamma_0 \|_\infty < O_p(n^{-1/2})$ if $\hat{\gamma}_n$ satisfies $\mathbf{W}_n(\hat{\gamma}_n) = o_p(n^{-1/2})$.

(iii) Note that

$$o_p(n^{-1/2}) = \widetilde{\mathbf{W}}_n(\hat{\gamma}_n) = \widetilde{\mathbf{W}}_n(\gamma_0) + \widetilde{\mathbf{D}}_n(\gamma_0)(\hat{\gamma}_n - \gamma_0) + o(\|\hat{\gamma}_n - \gamma_0\|_\infty),$$

and $\mathbf{D}_0(\gamma) = \frac{\partial \mathbf{W}_0}{\partial \gamma}$ which leads to $\mathbf{D}_n(\gamma_0) = \mathbf{D}_0(\gamma_0) + o_p(1)$. By Lemma 2.2 and Theorem 2.1(ii),

$$n^{1/2}(\hat{\gamma}_n - \gamma_0) = -\mathbf{D}_0(\gamma_0)^{-1}n^{1/2}\mathbf{W}_n(\gamma_0) + o_p(1).$$

From the proof of Proposition 2.3.4(3) in Ding (2010), we obtain

$$\begin{aligned} n^{1/2}\mathbf{W}_n(\gamma_0) &= \int \mathbb{G}_n\{I(e(\gamma_0) \geq t)\}x \, dP_{e(\gamma_0), \delta_1, \mathbf{X}}(t, 1, x) \\ &\quad - \int \mathbb{G}_n\{I(e(\gamma_0) \geq t)\mathbf{X}\} \, dP_{e(\gamma_0), \delta_1}(t, 1) + \mathbb{G}_n\{[h^{(1)}\mathbf{X} - h^{(2)}]\delta_1\} + o_p(1). \end{aligned}$$

Thus, the asymptotic representation can be shown as $\sqrt{n}(\hat{\gamma}_n - \gamma_0) = \mathbb{G}_n\{m_3(\gamma_0, e(\gamma_0); \delta_1, \mathbf{X})\} + o_p(1)$, where

$$\begin{aligned} &m_3(\gamma_0, e(\gamma_0); \delta_1, \mathbf{X}) \\ &= [-\mathbf{D}_0(\gamma_0)]^{-1} \left\{ [h^{(1)}\mathbf{X} - h^{(2)}]\delta_1 + \int I(e(\gamma_0) \geq t)x \, dP_{e(\gamma_0), \delta_1, \mathbf{X}}(t, 1, x) \right. \\ &\quad \left. - \int \{I(e(\gamma_0) \geq t)\mathbf{X}\} \, dP_{e(\gamma_0), \delta_1}(t, 1) \right\}. \end{aligned}$$

The proof of Theorem 2.1 is complete. □

2.8.2 Proofs of Lemma 2.3 and Theorem 2.2

Our proof of Theorem 2.2 is similar to that of Theorem 2.1 in Ding and Nan (2015). We need the following three technical Lemmas for preparation, where Lemmas 2.4-2.5 are extracted from Ding and Nan (2015). They can apply to both independent and dependent censoring cases. We list them

here for the reader's convenience. For Lemmas 2.4-2.5 and Theorem 2.2, we can follow the same proofs of Lemmas 1.1, 1.3 and Theorem 2.1 in the supplementary material of Ding and Nan (2015). Details are thus omitted here.

Lemma 2.4. For any $\varepsilon > 0$ and $k = 0$ or 1 , with probability one, we have

$$\sup_{\gamma \in \mathcal{G}, -\infty < s < \infty} \sqrt{n} |H_n^{(k)}(\gamma, s) - h^{(k)}(\gamma, s)| = o(n^\varepsilon).$$

Lemma 2.5. Let $U_n(\gamma; s)$ be random variables for which there exist non-random Borel functions $u_n(\gamma; s)$ such that for every $\varepsilon > 0$,

$$(i) \quad \sup_{\gamma \in \mathcal{G}, -\infty < s < \infty} \sqrt{n} |U_n(\gamma, s) - u_n(\gamma, s)| = o(n^{-1/2+\varepsilon}).$$

$$(ii) \quad \sup_{\gamma \in \mathcal{G}} \int_{\mathbb{R}} |dU_n(\gamma, s)| = O(1) \text{ almost surely.}$$

$$(iii) \quad \sup_{\gamma \in \mathcal{G}, -\infty < s < \infty} |u_n(\gamma, s)| = O(1).$$

Then under Conditions 2.2, 2.4, 2.5, for every $0 < \varepsilon \leq 1/2$, with probability one we have

$$\sup_{\gamma \in \mathcal{G}, -\infty < y < \infty} \left| \int_{s=-\infty}^y U_n(\gamma, s) dH_n^{(0)}(\gamma, s) - \int_{s=-\infty}^y u_n(\gamma, s) dh_n^{(0)}(\gamma, s) \right| = o(n^{-1/2+\varepsilon})$$

Proof of Lemma 2.3. We prove the result in this lemma for $k = 0, 1, 2$, respectively. First, when $k = 0$, we need to show $\sup_{|r-r'|+|s-s'| \leq n^{-\varepsilon}} |h^{(0)}(\gamma, s) - h^{(0)}(\gamma', s')| = O(n^{-\varepsilon})$. Note that

$$I(e(\gamma) \leq s, \delta_1 = 1) = I(\xi \leq \tilde{T})I(\xi \leq \tilde{C})I(\xi \leq s + (\gamma - \gamma_0)^T \mathbf{X}).$$

So,

$$\begin{aligned}
h^{(0)}(\boldsymbol{\gamma}, s) &= P\{I(e(\boldsymbol{\gamma}) \leq s, \delta_1 = 1)\} \\
&= \int_{\mathcal{X}} \int_{-\infty}^s q(u + (\boldsymbol{\gamma} - \boldsymbol{\gamma}_0)^T \mathbf{X}) \bar{G}_C(u + (\boldsymbol{\gamma} - \boldsymbol{\gamma}_0)^T \mathbf{X}) \\
&\quad \times \bar{G}_{T|\xi}(u + (\boldsymbol{\gamma} - \boldsymbol{\gamma}_0)^T \mathbf{X} | u + (\boldsymbol{\gamma} - \boldsymbol{\gamma}_0)^T \mathbf{X}) du f_{\mathbf{X}}(x) dx,
\end{aligned}$$

where $\mathcal{X}$ is the parameter space for variable $\mathbf{X}$. Then,

$$\begin{aligned}
&|h^{(0)}(\boldsymbol{\gamma}, s) - h^{(0)}(\boldsymbol{\gamma}', s)| \\
&\leq \int_{\mathcal{X}} \int_{-\infty}^{\infty} \left| q'(u) \bar{G}_C(u) \bar{G}_{T|\xi}(u|u) - q(u) g_C(u) \bar{G}_{T|\xi}(u|u) - q(u) \bar{G}_C(u) g_{T|\xi}(u|u) \right| \\
&\quad \times |r - r'| |\mathbf{X}| du f_{\mathbf{X}}(x) dx \\
&\leq O(1) |r - r'| \int_{\mathcal{X}} \int_{-\infty}^{\infty} \{|q'(u)| + 2q(u)\} du |\mathbf{X}| f_{\mathbf{X}}(x) dx \\
&\leq O(1) |r - r'|,
\end{aligned} \tag{2.28}$$

where the second last inequality is based on Conditions 2.4-2.6, while the last inequality is based on Conditions 2.2, 2.4 and the Cauchy-Schwartz inequality. Similar calculations to that in (2.28) are conducted with respect

to s . It follows that

$$\begin{aligned}
& |h^{(0)}(\boldsymbol{\gamma}, s) - h^{(0)}(\boldsymbol{\gamma}, s')| \\
&= \left| \int_{\mathcal{X}} \int_s^{s'} q(u + (\boldsymbol{\gamma} - \boldsymbol{\gamma}_0)^T \mathbf{X}) \overline{G}_C(u + (\boldsymbol{\gamma} - \boldsymbol{\gamma}_0)^T \mathbf{X}) \right. \\
&\quad \left. \times \overline{G}_{T|\xi}(u + (\boldsymbol{\gamma} - \boldsymbol{\gamma}_0)^T \mathbf{X}) |u + (\boldsymbol{\gamma} - \boldsymbol{\gamma}_0)^T \mathbf{X}| du f_{\mathbf{X}}(x) dx \right| \\
&\leq O(1) \int_{\mathcal{X}} \int_s^{s'} du |\mathbf{X}| f_{\mathbf{X}}(x) dx \\
&\leq O(1) |s - s'|,
\end{aligned}$$

where the last inequality follows from Condition 2.2. Thus,

$$\sup_{|r-r'|+|s-s'|\leq n^{-\varepsilon}} |h^{(0)}(\boldsymbol{\gamma}, s) - h^{(0)}(\boldsymbol{\gamma}', s')| = O(1) \sup(|r-r'|+|s-s'|) = O(n^{-\varepsilon}).$$

Second, when $k = 1$, we decompose

$$\begin{aligned}
I(e(\boldsymbol{\gamma}) \geq s) &= I(\log \min(V, T, C) - \boldsymbol{\gamma}_0^T \mathbf{X} - (\boldsymbol{\gamma} - \boldsymbol{\gamma}_0)^T \mathbf{X} \geq s) \\
&= I(\xi \geq s + (\boldsymbol{\gamma} - \boldsymbol{\gamma}_0)^T \mathbf{X}) I(\tilde{T} \geq s + (\boldsymbol{\gamma} - \boldsymbol{\gamma}_0)^T \mathbf{X}) I(\tilde{C} \geq s + (\boldsymbol{\gamma} - \boldsymbol{\gamma}_0)^T \mathbf{X}),
\end{aligned}$$

which leads to

$$\begin{aligned}
h^{(1)}(\boldsymbol{\gamma}, s) &= P\{I(e(\boldsymbol{\gamma}) \geq s)\} \\
&= \int_{\mathcal{X}} \overline{G}_C(s + (\boldsymbol{\gamma} - \boldsymbol{\gamma}_0)^T \mathbf{X}) \overline{G}_{\xi, T}(s + (\boldsymbol{\gamma} - \boldsymbol{\gamma}_0)^T \mathbf{X}, s + (\boldsymbol{\gamma} - \boldsymbol{\gamma}_0)^T \mathbf{X}) f_{\mathbf{X}}(x) dx.
\end{aligned}$$

Then

$$\begin{aligned}
& |h^{(1)}(\boldsymbol{\gamma}, s) - h^{(1)}(\boldsymbol{\gamma}', s)| \\
&\leq |r - r'| \int_{\mathcal{X}} \left| m(s + \tilde{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0)^T \mathbf{X} \right| |\mathbf{X}| f_{\mathbf{X}}(x) dx \quad (2.29) \\
&\leq O(1) |r - r'|,
\end{aligned}$$

where $\tilde{\gamma}$ is chosen between γ and γ' and

$$m(s) = g_C(s)\overline{G}_{\xi,T}(s, s) + G_{\xi,T}^{(1)}(s, s)\overline{G}_C(s) + G_{\xi,T}^{(2)}(s, s)\overline{G}_C(s).$$

By similar calculation with respect to variable s , we have

$$\begin{aligned} & |h^{(1)}(\gamma, s) - h^{(1)}(\gamma, s')| \\ & \leq |s - s'| \int_{\mathcal{X}} \left| m(\tilde{s} + (\gamma - \gamma_0)^T \mathbf{X}) \right| f_{\mathbf{X}}(x) dx \\ & \leq O(1)|s - s'|, \end{aligned} \tag{2.30}$$

where $\tilde{s}$ is chosen between s and s' . Thus $\sup_{|r-r'|+|s-s'|\leq n^{-\varepsilon}} |h^{(1)}(\gamma, s) - h^{(1)}(\gamma', s')| = O(n^{-\varepsilon})$.

At last, the proof of the result for $k = 2$ is similar. We apply the same calculations as in (2.29) and (2.30) to $h^{(2)}(\cdot, \cdot)$. It follows that

$$\begin{aligned} & |h^{(2)}(\gamma, s) - h^{(2)}(\gamma', s)| \\ & \leq |r - r'| \int_{\mathcal{X}} \left| m(s + (\tilde{\gamma} - \gamma_0)^T \mathbf{X}) \right| |\mathbf{X}|^2 f_{\mathbf{X}}(x) dx \\ & \leq O(1)|r - r'| \end{aligned}$$

and

$$\begin{aligned} & |h^{(2)}(\gamma, s) - h^{(2)}(\gamma, s')| \\ & \leq |s - s'| \int_{\mathcal{X}} \left| m(\tilde{s} + (\gamma - \gamma_0)^T \mathbf{X}) \right| |\mathbf{X}| f_{\mathbf{X}}(x) dx \\ & \leq O(1)|s - s'|. \end{aligned}$$

Thus $\sup_{|r-r'|+|s-s'|\leq n^{-\varepsilon}} |h^{(2)}(\gamma, s) - h^{(2)}(\gamma', s')| = O(n^{-\varepsilon})$. This completes the proof of the lemma. $\square$

Proof of Theorem 2.2. Using Lemmas 2.3-2.5 and similar arguments as in Theorem 2.1 of Ding and Nan (2015), we can accomplish the proof of the consistency. Using the similar arguments in the proof of Theorem 2.3.4 in Ding (2010) and changing some of their calculations as we did in Lemma 2.3, we see that $\sqrt{n}(\widehat{Q}_{n,\gamma}^{km}(t) - Q^*(\gamma, t))$ converges to a zero-mean Gaussian process. $\square$

2.8.3 Proof of Theorem 2.3

Define

$$\begin{aligned}\bar{\Lambda}_{n,0}(t) &= \frac{\sum_{i=1}^n (\delta_{1i}\tilde{\delta}_{2i} + \delta_{2i})(1 - Y_i(t))}{\sum_{l=1}^n Y_l(t)\mu_l(\alpha_0, \beta_0, \Lambda_{0,0}, \gamma_0, Q_0)}, \\ \bar{\bar{\Lambda}}_{n,0}(t) &= \frac{\sum_{i=1}^n (\delta_{1i}\tilde{\delta}_{2i} + \delta_{2i})(1 - Y_i(t))}{\sum_{l=1}^n Y_l(t)\mu_l(\alpha_0, \beta_0, \Lambda_{0,0}, \hat{\gamma}_n, \bar{Q}_{n,\hat{\gamma}_n})}, \\ \hat{\Lambda}_{n,0}(t) &= \frac{\sum_{i=1}^n (\delta_{1i}\tilde{\delta}_{2i} + \delta_{2i})(1 - Y_i(t))}{\sum_{l=1}^n Y_l(t)\mu_l(\hat{\alpha}_n, \hat{\beta}_n, \hat{\Lambda}_{n,0}, \hat{\gamma}_n, \hat{Q}_n)},\end{aligned}$$

where

$$\bar{Q}_{n,\hat{\gamma}_n}(t) = \int_{s \leq t} \frac{d\hat{Q}_{n,\hat{\gamma}_n}^{km}(s)}{\hat{K}_Q(s; \alpha_0, \beta_0, \Lambda_{0,0}, \hat{\gamma}_n, Q_0, S_c)[1 - \hat{Q}_{n,\hat{\gamma}_n}^{km}(s)]}.$$

First of all, we need to show the consistency of $\bar{\Lambda}_{n,0}$. It can be proved using the techniques of counting process and martingale residuals. Conditional on X_i , the local martingales associated with $dN_{2i}(t) + d\tilde{N}_{2i}(t)$ given by

$$M_i(t) = \int_0^t dN_{2i}(u) + d\tilde{N}_{2i}(u) - Y_i(u)\mu_i(\alpha_0, \beta_0, \Lambda_{0,0}, \gamma_0, Q_0)d\Lambda_{0,0}(u)$$

have mean zero ($\forall t \in [0, \varsigma]$). By the strong law of large numbers, we have $\frac{1}{n} \sum_{i=1}^n M_i(t) \xrightarrow{a.s.} 0$. Similarly, we can also define martingales

$$\overline{M}_i(t) = \int_0^t dN_{2i}(u) + d\tilde{N}_{2i}(u) - Y_i(u)\mu_i(\alpha_0, \beta_0, \Lambda_{0,0}, \gamma_0, Q_0)d\overline{\Lambda}_{n,0}(u).$$

By the definition of $\overline{\Lambda}_{n,0}(t)$, which is a piecewise constant function, the sum of all $\overline{M}_i(t)$ can be calculated by

$$\begin{aligned} & \sum_{i=1}^n \overline{M}_i(t) \\ &= \sum_{i=1}^n \sum_{k: T_k \leq t} \left(dN_{2i}(T_k) + d\tilde{N}_{2i}(T_k) \right) \\ & \quad - \sum_{i=1}^n \sum_{k: T_k \leq t} \frac{Y_i(T_k)\mu_i(\alpha_0, \beta_0, \Lambda_{0,0}, \gamma_0, Q_0) \left(\sum_{j: T_j = T_k} dN_{2j}(T_k) + d\tilde{N}_{2j}(T_k) \right)}{\sum_{l=1}^n Y_l(T_k)\mu_l(\alpha_0, \beta_0, \Lambda_{0,0}, \gamma_0, Q_0)} \\ &= \sum_{k: T_k \leq t} \sum_{i=1}^n (\delta_{2i} + \delta_{1i}\tilde{\delta}_{2i})I(T_i = T_k) \\ & \quad - \sum_{k: T_k \leq t} \frac{\sum_{i=1}^n Y_i(T_k)\mu_i(\alpha_0, \beta_0, \Lambda_{0,0}, \gamma_0, Q_0)}{\sum_{l=1}^n Y_l(T_k)\mu_l(\alpha_0, \beta_0, \Lambda_{0,0}, \gamma_0, Q_0)} \left(\sum_{j: T_j = T_k} dN_{2j}(T_k) + d\tilde{N}_{2j}(T_k) \right) \\ &= \sum_{k: T_k \leq t} \sum_{i=1}^n (\delta_{2i} + \delta_{1i}\tilde{\delta}_{2i})I(T_i = T_k) - \sum_{k: T_k \leq t} \sum_{j: T_j = T_k} (\delta_{2j} + \delta_{1j}\tilde{\delta}_{2j})I(T_j = T_k) = 0. \end{aligned} \tag{2.31}$$

It follows that $\frac{1}{n} \sum_{i=1}^n M_i(t) - \overline{M}_i(t) \xrightarrow{a.s.} 0$, that is,

$$\int_0^t \left(\frac{1}{n} \sum_{i=1}^n Y_i(u)\mu_i(\alpha_0, \beta_0, \Lambda_{0,0}, \gamma_0, Q_0) \right) d(\overline{\Lambda}_{n,0}(u) - \Lambda_{0,0}(u)) \xrightarrow{a.s.} 0.$$

By the fact that Y_i and Z_i fall in $[0, \varsigma]$ and Conditions 2.2, 2.7 and 2.8, we can see that $\mu_{1,i}(\alpha_0, \beta_0)$, $A_i(\alpha_0, \beta_0, \Lambda_{0,0}, \gamma_0, Q_0)$, $B_i(\alpha_0, \beta_0, \Lambda_{0,0}, \gamma_0, Q_0)$ and

$C_i(\alpha_0, \beta_0, \Lambda_{0,0}, \gamma_0, Q_0)$, defined in (2.15), are all bounded away from zero. Thus from Condition 2.9,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n Y_i(u) \mu_i(\alpha_0, \beta_0, \Lambda_{0,0}, \gamma_0, Q_0) = E[Y_i(u) \mu_i(\alpha_0, \beta_0, \Lambda_{0,0}, \gamma_0, Q_0)]$$

is also bounded away from zero. For a large enough n , there exists a small positive constant c_0 such that $\frac{1}{n} \sum_{i=1}^n Y_i(u) \mu_i(\alpha_0, \beta_0, \Lambda_{0,0}, \gamma_0, Q_0) \geq c_0$. Therefore,

$$\begin{aligned} 0 &\leq \left| c_0 \int_0^t d(\bar{\Lambda}_{n,0}(u) - \Lambda_{0,0}(u)) \right| \\ &\leq \left| \int_0^t \left(\frac{1}{n} \sum_{i=1}^n Y_i(u) \mu_i(\alpha_0, \beta_0, \Lambda_{0,0}, \gamma_0, Q_0) \right) d(\bar{\Lambda}_{n,0}(u) - \Lambda_{0,0}(u)) \right| \xrightarrow{a.s.} 0. \end{aligned}$$

By Squeeze Theorem, we have

$$\bar{\Lambda}_{n,0}(t) - \Lambda_{0,0}(t) \xrightarrow{a.s.} 0 \quad \forall t \in [0, \varsigma]. \quad (2.32)$$

In fact, if a sequence of non-decreasing functions converges pointwise to a continuous function, it also converges uniformly. Given that uniform convergence of this sequence to a continuous limit in a compact space implies local uniform continuity, $\bar{\Lambda}_{n,0}(t)$ is locally and uniformly continuous on $[0, \varsigma]$.

Second, we prove the consistency of $\bar{\bar{\Lambda}}_{n,0}$. Let

$$\bar{\bar{M}}_i(t) = \int_0^t dN_{2i}(u) + d\tilde{N}_{2i}(u) - Y_i(u) \mu_i(\alpha_0, \beta_0, \Lambda_{0,0}, \hat{\gamma}_n, \bar{Q}_{n, \hat{\gamma}_n}) d\bar{\bar{\Lambda}}_{n,0}(u).$$

Performing calculations similar to those in (2.31), we can obtain $\sum_{i=1}^n \bar{\bar{M}}_i(t) =$

0. Combining with $\sum_{i=1}^n \overline{M}_i(t) = 0$, we have

$$\begin{aligned}
 0 &= \frac{1}{n} \sum_{i=1}^n \int_0^t Y_i(u) \mu_i(\alpha_0, \beta_0, \Lambda_{0,0}, \hat{\gamma}_n, \overline{Q}_{n,\hat{\gamma}_n}) d\overline{\overline{\Lambda}}_{n,0}(u) \\
 &\quad - \int_0^t Y_i(u) \mu_i(\alpha_0, \beta_0, \Lambda_{0,0}, \gamma_0, Q_0) d\overline{\Lambda}_{n,0}(u) \\
 &= \frac{1}{n} \sum_{i=1}^n \int_0^t Y_i(u) [\mu_i(\alpha_0, \beta_0, \Lambda_{0,0}, \hat{\gamma}_n, \overline{Q}_{n,\hat{\gamma}_n}) - \mu_i(\alpha_0, \beta_0, \Lambda_{0,0}, \gamma_0, Q_0)] d\overline{\overline{\Lambda}}_{n,0}(u) \\
 &\quad + \int_0^t Y_i(u) \mu_i(\alpha_0, \beta_0, \Lambda_{0,0}, \gamma_0, Q_0) d(\overline{\overline{\Lambda}}_{n,0}(u) - \overline{\Lambda}_{n,0}(u)).
 \end{aligned} \tag{2.33}$$

Consider the condition that

$$|\hat{\gamma}_n - \gamma_0| + \sup_t |\hat{Q}_{n,\hat{\gamma}_n}^{km}(t) - Q^*(\gamma_0, t)| = o_p^*(1),$$

where $o_p^*(1)$ is defined in terms of convergence in outer probability, that is, $Pr^*(|U| > \epsilon) \rightarrow 0$ if a random variable U converges in outer probability to 0, and $Pr^*(A) = \inf\{Pr(B) : A \subset B, B \in \mathcal{B}\}$ for any A of Ω in the probability space $\{\Omega, \mathcal{B}, Pr\}$. We see that from Condition 2.2 and 2.4, $Q_0(\log u - \hat{\gamma}_n^T \mathbf{X}_l) - Q_0(\log u - \gamma_0^T \mathbf{X}_l) = Q'_0(\log u - \tilde{\gamma}^T \mathbf{X}_l)(\hat{\gamma}_n - \gamma_0)^T \mathbf{X}_l = o_p^*(1)$, where $\tilde{\gamma}$ is chosen between $\hat{\gamma}_n$ and γ_0 , moreover, $K_Q(s; \alpha_0, \beta_0, \Lambda_{0,0}, \gamma_0, Q_0, S_c) / \hat{K}_Q(s; \alpha_0, \beta_0, \Lambda_{0,0}, \hat{\gamma}_n, Q_0, S_c) = 1 + o_p^*(1)$. So, $\overline{Q}_{n,\hat{\gamma}_n}(\log u - \hat{\gamma}_n^T \mathbf{X}_l) - Q_0(\log u - \gamma_0^T \mathbf{X}_l) = o_p^*(1)$. It follows from Condition 2.9 that $\mu_i(\alpha_0, \beta_0, \Lambda_{0,0}, \hat{\gamma}_n, \overline{Q}_{n,\hat{\gamma}_n}) - \mu_i(\alpha_0, \beta_0, \Lambda_{0,0}, \gamma_0, Q_0) = o_p^*(1)$. Then the first term in the last equation of (2.33) is $o_p^*(1)$. Thus the second term in the last equation of (2.33) converges in outer probability to zero. Consequently, $\overline{\overline{\Lambda}}_{n,0}(t) - \overline{\Lambda}_{n,0}(t) \xrightarrow{p^*} 0$. By (2.32), $\overline{\overline{\Lambda}}_{n,0}(t) - \Lambda_{0,0}(t) \xrightarrow{p^*} 0, \forall t \in [0, \varsigma]$.

Condition 2.7 implies the boundedness of $\hat{\Lambda}_{n,0}(t)$, thus there is a limit $\Lambda_0^*(t)$ for sequence $\{\hat{\Lambda}_{n,0}(t) : n \in \mathbb{N}\}$. Due to the compact parameter

spaces $\mathcal{A}$ and $\mathcal{B}$, we can consider the existence of a convergent subsequences $\{(\hat{\alpha}_n, \hat{\beta}_n)\}$ with limits (α^*, β^*) .

Now it remains to show $(\alpha^*, \beta^*, \Lambda_0^*) = (\alpha_0, \beta_0, \Lambda_{0,0})$. Note that $\bar{\bar{\Lambda}}_{n,0}$ is right continuous with jumps on Y_i . Let $\bar{\bar{\lambda}}_{n,0}(Y_i)$ be the size of jump of $\bar{\bar{\Lambda}}_{n,0}$ at the point Y_i . Given the uniform convergence of $\bar{\bar{\Lambda}}_{n,0}$ to a continuous limit on $[0, \varsigma]$, we have $\bar{\bar{\lambda}}_{n,0}(t) \xrightarrow{p^*} \lambda_{0,0}(t)$ for any $t \in [0, \varsigma]$. Since $(\hat{\alpha}_n, \hat{\beta}_n, \hat{\Lambda}_{n,0})$ is the maximizer of $pl_n(\alpha, \beta, \Lambda_0; \hat{\gamma}_n, \hat{Q}_{n,\hat{\gamma}_n})$, we can see that

$$\begin{aligned}
& 0 \leq pl_n(\hat{\alpha}_n, \hat{\beta}_n, \hat{\Lambda}_{n,0}; \hat{\gamma}_n, \hat{Q}_n) - pl_n(\alpha_0, \beta_0, \bar{\bar{\Lambda}}_{n,0}; \hat{\gamma}_n, \bar{Q}_{n,\hat{\gamma}_n}) \\
&= \frac{1}{n} \sum_{i=1}^n \left\{ \log \left\{ \left[\frac{\hat{\lambda}_{n,0}(Y_i) \exp(\hat{\alpha}_n Z_i + \hat{\beta}_n^T \mathbf{X}_i)}{\bar{\bar{\lambda}}_{n,0}(Y_i) \exp(\alpha_0 Z_i + \beta_0^T \mathbf{X}_i)} \right]^{\delta_{1i} \tilde{\delta}_{2i}} \right. \right. \\
&\quad \times \left[\frac{\exp[-\hat{\Lambda}_{n,0}(Y_i) \exp(\hat{\alpha}_n Z_i + \hat{\beta}_n^T \mathbf{X}_i)]}{\exp[-\bar{\bar{\Lambda}}_{n,0}(Y_i) \exp(\alpha_0 Z_i + \beta_0^T \mathbf{X}_i)]} \right]^{\delta_{1i}} \\
&\quad \times \left[\frac{\int_{Z_i}^{\tau} \hat{\lambda}_{n,0}(Z_i) \exp[-\hat{\Lambda}_{n,0}(Z_i) \exp(\hat{\alpha}_n v + \hat{\beta}_n^T \mathbf{X}_i)] \exp(\hat{\alpha}_n v + \hat{\beta}_n^T \mathbf{X}_i) d\hat{Q}_n(\log v - \hat{\gamma}_n^T \mathbf{X}_i)}{\int_{Z_i}^{\tau} \bar{\bar{\lambda}}_{n,0}(Z_i) \exp[-\bar{\bar{\Lambda}}_{n,0}(Z_i) \exp(\alpha_0 v + \beta_0^T \mathbf{X}_i)] \exp(\alpha_0 v + \beta_0^T \mathbf{X}_i) d\bar{Q}_{n,\hat{\gamma}_n}(\log v - \hat{\gamma}_n^T \mathbf{X}_i)} \right]^{\delta_{2i}} \\
&\quad \times \left. \left[\frac{\int_{Z_i}^{\tau} \exp[-\hat{\Lambda}_{n,0}(Z_i) \exp(\hat{\alpha}_n v + \hat{\beta}_n^T \mathbf{X}_i)] d\hat{Q}_n(\log v - \hat{\gamma}_n^T \mathbf{X}_i)}{\int_{Z_i}^{\tau} \exp[-\bar{\bar{\Lambda}}_{n,0}(Z_i) \exp(\alpha_0 v + \beta_0^T \mathbf{X}_i)] d\bar{Q}_{n,\hat{\gamma}_n}(\log v - \hat{\gamma}_n^T \mathbf{X}_i)} \right]^{1-\delta_{1i}-\delta_{2i}} \right\} \Bigg\} \\
&\xrightarrow{p^*} E \left\{ \log \left\{ \left[\frac{\lambda_0^*(Y_i) \exp(\alpha^* Z_i + \beta^{*T} \mathbf{X}_i)}{\lambda_{0,0}(Y_i) \exp(\alpha_0 Z_i + \beta_0^T \mathbf{X}_i)} \right]^{\delta_{1i} \tilde{\delta}_{2i}} \times \left[\frac{\exp[-\Lambda_0^*(Y_i) \exp(\alpha^* Z_i + \beta^{*T} \mathbf{X}_i)]}{\exp[-\Lambda_{0,0}(Y_i) \exp(\alpha_0 Z_i + \beta_0^T \mathbf{X}_i)]} \right]^{\delta_{1i}} \right. \right. \\
&\quad \times \left[\frac{\int_{Z_i}^{\tau} \lambda_0^*(Z_i) \exp[-\Lambda_0^*(Z_i) \exp(\alpha^* v + \beta^{*T} \mathbf{X}_i)] \exp(\alpha^* v + \beta^{*T} \mathbf{X}_i) d\hat{Q}^*(\log v - \gamma_0^T \mathbf{X}_i)}{\int_{Z_i}^{\tau} \lambda_{0,0}(Z_i) \exp[-\Lambda_{0,0}(Z_i) \exp(\alpha_0 v + \beta_0^T \mathbf{X}_i)] \exp(\alpha_0 v + \beta_0^T \mathbf{X}_i) dQ_0(\log v - \gamma_0^T \mathbf{X}_i)} \right]^{\delta_{2i}} \\
&\quad \times \left. \left[\frac{\int_{Z_i}^{\tau} \exp[-\Lambda_0^*(Z_i) \exp(\alpha^* v + \beta^{*T} \mathbf{X}_i)] d\hat{Q}^*(\log v - \gamma_0^T \mathbf{X}_i)}{\int_{Z_i}^{\tau} \exp[-\Lambda_{0,0}(Z_i) \exp(\alpha_0 v + \beta_0^T \mathbf{X}_i)] dQ_0(\log v - \gamma_0^T \mathbf{X}_i)} \right]^{1-\delta_{1i}-\delta_{2i}} \right\} \Bigg\} \\
&= pl(\alpha^*, \beta^*, \Lambda_0^*; \gamma_0, \hat{Q}^*) - pl(\alpha_0, \beta_0, \Lambda_{0,0}; \gamma_0, Q_0),
\end{aligned}$$

(2.34)

where $\widehat{Q}^*$ is the limit of $\widehat{Q}_n$. $(\alpha_0, \beta_0, \Lambda_{0,0})$ is the maximizer of $pl(\alpha, \beta, \Lambda_0; \gamma_0, Q_0)$, then the right hand side of (2.34) is not more than zero. Therefore, we get

$$pl(\alpha^*, \beta^*, \Lambda_0^*; \gamma_0, \widehat{Q}^*) = pl(\alpha_0, \beta_0, \Lambda_{0,0}; \gamma_0, Q_0). \quad (2.35)$$

Without loss of generality, we set $\delta_{1i} = 1$, $\widetilde{\delta}_{2i} = 0$, $Z_i = \epsilon$ for some small positive value ϵ and $Y_i = \varsigma$. From (2.34) and (2.35),

$$\Lambda_0^*(\varsigma) \exp(\alpha^* \epsilon + \beta^{*T} \mathbf{X}_i) = \Lambda_{0,0}(\varsigma) \exp(\alpha_0 \epsilon + \beta_0^T \mathbf{X}_i). \quad (2.36)$$

Taking logarithm on both sides of equation (2.36), we have

$$c[1, \epsilon, \mathbf{X}_i^T]^T = 0, \quad i = 1, \dots, n,$$

where $c = [\log \Lambda_0^*(\varsigma) - \log \Lambda_{0,0}(\varsigma), \alpha^* - \alpha_0, \beta^{*T} - \beta_0^T]$. It follows from Condition 2.11 that $c = 0$, that is, $\alpha^* = \alpha_0$, $\beta^* = \beta_0$, and $\Lambda_0^*(\varsigma) = \Lambda_{0,0}(\varsigma)$. Next, we need to prove $\Lambda_0^*(t) = \Lambda_{0,0}(t)$ for any $t \in [0, \varsigma]$. Let $\delta_{1i} = 1$, $\widetilde{\delta}_{2i} = 1$, $Z_i = 0$ and integrate (2.35) with respect to Y_i from 0 to t , we have

$$\begin{aligned} & \int_0^t \lambda_0^*(y) \exp(\beta_0^T \mathbf{X}_i) \exp[-\Lambda_0^*(y) \exp(\beta_0^T \mathbf{X}_i)] dy \\ &= \int_0^t \lambda_{0,0}(y) \exp(\beta_0^T \mathbf{X}_i) \exp[\Lambda_{0,0}(y) \exp(\beta_0^T \mathbf{X}_i)] dy \end{aligned}$$

or

$$\exp[-\Lambda_0^*(t) \exp(\beta_0^T \mathbf{X}_i)] = \exp[\Lambda_{0,0}(t) \exp(\beta_0^T \mathbf{X}_i)],$$

which implies that $\Lambda_0^*(t) = \Lambda_{0,0}(t)$. Furthermore, we see that

$$\widehat{Q}^*(t) = \int_{s \leq t} \frac{dQ^*(\gamma_0, s)}{K_Q(s; \alpha_0, \beta_0, \Lambda_{0,0}, \gamma_0, \widehat{Q}^*, \widehat{S}_c)[1 - Q^*(\gamma_0, s)]}.$$

From (2.13) and Condition 2.10, we have

$$K_Q(t; \alpha_0, \beta_0, \Lambda_{0,0}, \gamma_0, \widehat{Q}^*, S_c) d\widehat{Q}^*(t) = K_Q(t; \alpha_0, \beta_0, \Lambda_{0,0}, \gamma_0, Q_0, S_c) dQ_0(t)$$

and

$$\begin{aligned}
& P \left\{ d_{Q_0}(\log t - \gamma_0^T \mathbf{X}) \int_{v \geq t} \exp[-\Lambda_{0,0}(t) \exp(\alpha_0 v + \beta_0^T \mathbf{X})] d(\hat{Q}^* - Q_0)(\log v - \gamma_0^T \mathbf{X}) \right\} \\
&= P \left\{ d(\hat{Q}^* - Q_0)(\log t - \gamma_0^T \mathbf{X}) \int_{v \geq t} \exp[-\Lambda_{0,0}(t) \exp(\alpha_0 v + \beta_0^T \mathbf{X})] dQ_0(\log v - \gamma_0^T \mathbf{X}) \right\}
\end{aligned}$$

It follows from Condition 2.12 that $\hat{Q}^* = Q_0$. The proof of this theorem is complete.

2.8.4 Proof of Theorem 2.4

From the consistency property, we can define a neighborhood of the true parameter pair $(\alpha_0, \beta_0, \Lambda_{0,0})$ as

$$B_{\epsilon_0}(\alpha_0, \beta_0, \Lambda_{0,0}) = \left\{ (\alpha, \beta, \Lambda_0) : |\alpha - \alpha_0| + \|\beta - \beta_0\| + \sup_t |\Lambda_0(t) - \Lambda_{0,0}(t)| < \epsilon_0 \right\}$$

for a small positive constant ϵ_0 . We also define a set

$$\begin{aligned}
\mathcal{H} = & \left\{ (h_1, h_2, h_3) : h_1 \in \mathbb{R}, h_2 \in \mathbb{R}^p \text{ and } h_3(\cdot) \text{ is a function of bounded variation on } [0, \varsigma] \right. \\
& \left. \text{with } |h_1| \leq 1, \|h_2\| \leq 1, \text{ and } |h_3|_{BV} \leq 1 \right\},
\end{aligned}$$

where $|h_3|_{BV}$ is the total variation of h_3 in $[0, \varsigma]$. Note that $\mathcal{H}$ is Donsker due to the reason that the collection of half spaces is a Donsker class. Let l_α and l_β be scores for α and β , respectively, and $l_{\Lambda_0}[h_3]$ be the score for Λ_0 along the submodel $\Lambda_0 + \epsilon \int_0^t h_3 d\Lambda_0$. In detail, the explicit forms of l_α ,

l_{β} and $l_{\Lambda_0}[h_3]$ are specified as follows:

$$\begin{aligned}
 l_{\alpha}(\alpha, \beta, \Lambda_0; \gamma, Q) &= \delta_{1i} \tilde{\delta}_{2i} Z_i - \delta_{1i} Z_i \Lambda_0(Y_i) \exp(\alpha Z_i + \beta^T \mathbf{X}_i) \\
 &\quad + \delta_{2i} R_{1i}^{-1} \int_{Z_i}^{\tau} v [1 - \Lambda_0(Z_i) \exp(\alpha v + \beta^T \mathbf{X}_i)] \\
 &\quad \times \exp [\alpha v + \beta^T \mathbf{X}_i - \Lambda_0(Z_i) \exp(\alpha v + \beta^T \mathbf{X}_i)] dQ(\log v - \gamma^T \mathbf{X}_i) \\
 &\quad + (1 - \delta_{1i} - \delta_{2i}) R_{2i}^{-1} \int_{Z_i}^{\tau} -v \Lambda_0(Z_i) \exp(\alpha v + \beta^T \mathbf{X}_i) \\
 &\quad \times \exp [-\Lambda_0(Z_i) \exp(\alpha v + \beta^T \mathbf{X}_i)] dQ(\log v - \gamma^T \mathbf{X}_i),
 \end{aligned}$$

$$\begin{aligned}
 l_{\beta}(\alpha, \beta, \Lambda_0; \gamma, Q) &= \delta_{1i} \tilde{\delta}_{2i} \mathbf{X}_i - \delta_{1i} \mathbf{X}_i \Lambda_0(Y_i) \exp(\alpha Z_i + \beta^T \mathbf{X}_i) \\
 &\quad + \delta_{2i} R_{1i}^{-1} \int_{Z_i}^{\tau} \mathbf{X}_i [1 - \Lambda_0(Z_i) \exp(\alpha v + \beta^T \mathbf{X}_i)] \\
 &\quad \times \exp [\alpha v + \beta^T \mathbf{X}_i - \Lambda_0(Z_i) \exp(\alpha v + \beta^T \mathbf{X}_i)] dQ(\log v - \gamma^T \mathbf{X}_i) \\
 &\quad + (1 - \delta_{1i} - \delta_{2i}) R_{2i}^{-1} \int_{Z_i}^{\tau} -\mathbf{X}_i \Lambda_0(Z_i) \exp(\alpha v + \beta^T \mathbf{X}_i) \\
 &\quad \times \exp [-\Lambda_0(Z_i) \exp(\alpha v + \beta^T \mathbf{X}_i)] dQ(\log v - \gamma^T \mathbf{X}_i),
 \end{aligned}$$

$$\begin{aligned}
l_{\Lambda_0}(\alpha, \beta, \Lambda_0; \gamma, Q)[h_3] &= \delta_{1i} \tilde{\delta}_{2i} h_3(Y_i) + \delta_{2i} h_3(Z_i) - \delta_{1i} \exp(\alpha Z_i + \beta^T \mathbf{X}_i) \int_0^{Y_i} h_3 d\Lambda_0 \\
&\quad + \delta_{2i} R_{1i}^{-1} \int_0^{Z_i} h_3 d\Lambda_0 \int_{Z_i}^{\tau} -\exp(2\alpha v + 2\beta^T \mathbf{X}_i) \\
&\quad \times \exp[-\Lambda_0(Z_i) \exp(\alpha v + \beta^T \mathbf{X}_i)] dQ(\log v - \gamma^T \mathbf{X}_i) \\
&\quad + (1 - \delta_{1i} - \delta_{2i}) R_{2i}^{-1} \int_0^{Z_i} h_3 d\Lambda_0 \int_{Z_i}^{\tau} -\exp(\alpha v + \beta^T \mathbf{X}_i) \\
&\quad \times \exp[-\Lambda_0(Z_i) \exp(\alpha v + \beta^T \mathbf{X}_i)] dQ(\log v - \gamma^T \mathbf{X}_i),
\end{aligned}$$

where

$$R_{1i}(\alpha, \beta, \Lambda_0; \gamma, Q) = \int_{Z_i}^{\tau} \exp(\alpha v + \beta^T \mathbf{X}_i) \exp[-\Lambda_0(Z_i) \exp(\alpha v + \beta^T \mathbf{X}_i)] dQ(\log v - \gamma^T \mathbf{X}_i),$$

and

$$R_{2i}(\alpha, \beta, \Lambda_0; \gamma, Q) = \int_{Z_i}^{\tau} \exp[-\Lambda_0(Z_i) \exp(\alpha v + \beta^T \mathbf{X}_i)] dQ(\log v - \gamma^T \mathbf{X}_i).$$

Consider the random mapping from $\Psi_n : B_\epsilon(\alpha_0, \beta_0, \Lambda_{0,0}) \rightarrow l^\infty(\mathcal{H})$ and corresponding asymptotic mapping $\Psi : B_\epsilon(\alpha_0, \beta_0, \Lambda_{0,0}) \rightarrow l^\infty(\mathcal{H})$, where $l^\infty(\mathcal{H})$ includes all the bounded functions on $\mathcal{H}$. Here,

$$\begin{aligned}
&\Psi_n(\alpha, \beta, \Lambda_0; \gamma, Q)[h_1, h_2, h_3] \\
&= \mathbb{P}_n \left\{ h_1 l_\alpha(\alpha, \beta, \Lambda_0; \gamma, Q) + h_2^T l_\beta(\alpha, \beta, \Lambda_0; \gamma, Q) + l_{\Lambda_0}(\alpha, \beta, \Lambda_0; \gamma, Q)[h_3] \right\} \\
&:= A_{n1}(\alpha, \beta, \Lambda_0; \gamma, Q)[h_1] + A_{n2}(\alpha, \beta, \Lambda_0; \gamma, Q)[h_2] + A_{n3}(\alpha, \beta, \Lambda_0; \gamma, Q)[h_3]
\end{aligned}$$

and

$$\begin{aligned}
&\Psi(\alpha, \beta, \Lambda_0; \gamma, Q)[h_1, h_2, h_3] \\
&= P \left\{ h_1 l_\alpha(\alpha, \beta, \Lambda_0; \gamma, Q) + h_2^T l_\beta(\alpha, \beta, \Lambda_0; \gamma, Q) + l_{\Lambda_0}(\alpha, \beta, \Lambda_0; \gamma, Q)[h_3] \right\} \\
&:= A_1(\alpha, \beta, \Lambda_0; \gamma, Q)[h_1] + A_2(\alpha, \beta, \Lambda_0; \gamma, Q)[h_2] + A_3(\alpha, \beta, \Lambda_0; \gamma, Q)[h_3].
\end{aligned}$$

By definition, $\Psi(\alpha_0, \beta_0, \Lambda_{0,0}; \gamma_0, Q_0) = 0$. $(\hat{\alpha}_n, \hat{\beta}_n, \hat{\Lambda}_{n,0})$ satisfying $\Psi_n(\hat{\alpha}_n, \hat{\beta}_n, \hat{\Lambda}_{n,0}; \hat{\gamma}_n, \hat{Q}_n) = o_{p^*}(n^{-1/2})$ is a consistent estimator of $(\alpha_0, \beta_0, \Lambda_{0,0})$. The asymptotic representation of $\sqrt{n}(\hat{\alpha}_n - \alpha_0, \hat{\beta}_n - \beta_0, \hat{\Lambda}_{n,0} - \Lambda_{0,0})$ can be derived directly from the semiparametric Z-estimation theory in Nan and Wellner (2013) when the following Conditions B1)-B4) hold.

B1) (stochastic equicontinuity)

$$\frac{\sqrt{n}(\Psi_n - \Psi)(\hat{\alpha}_n, \hat{\beta}_n, \hat{\Lambda}_{n,0}; \hat{\gamma}_n, \hat{Q}_n) - \sqrt{n}(\Psi_n - \Psi)(\alpha_0, \beta_0, \Lambda_{0,0}; \gamma_0, Q_0)}{1 + \sqrt{n}|\Psi_n(\hat{\alpha}_n, \hat{\beta}_n, \hat{\Lambda}_{n,0}; \hat{\gamma}_n, \hat{Q}_n)| + \sqrt{n}|\Psi(\hat{\alpha}_n, \hat{\beta}_n, \hat{\Lambda}_{n,0}; \hat{\gamma}_n, \hat{Q}_n)|} = o_p^*(1).$$

B2) $\sqrt{n}\Psi_n(\alpha_0, \beta_0, \Lambda_{0,0}; \gamma_0, Q_0) = O_{p^*}(1)$.

B3) (smoothness) There exists a continuous linear operator $\Psi' = (\Psi'_1, \Psi'_2, \Psi'_3)^T$, a continuous matrix Ψ'_4 and a continuous linear functional Ψ'_5 such that

$$\begin{aligned} & \left| \Psi(\hat{\alpha}_n, \hat{\beta}_n, \hat{\Lambda}_{n,0}; \hat{\gamma}_n, \hat{Q}_n) - \Psi(\alpha_0, \beta_0, \Lambda_{0,0}; \gamma_0, Q_0) \right. \\ & - \Psi'_1(\alpha_0, \beta_0, \Lambda_{0,0}; \gamma_0, Q_0)(\hat{\alpha}_n - \alpha_0) - \Psi'_2(\alpha_0, \beta_0, \Lambda_{0,0}; \gamma_0, Q_0)(\hat{\beta}_n - \beta_0) \\ & - \Psi'_3(\alpha_0, \beta_0, \Lambda_{0,0}; \gamma_0, Q_0)(\hat{\Lambda}_{n,0} - \Lambda_{0,0}) - \Psi'_4(\alpha_0, \beta_0, \Lambda_{0,0}; \gamma_0, Q_0)(\hat{\gamma}_n - \gamma_0) \\ & \left. - \Psi'_5(\alpha_0, \beta_0, \Lambda_{0,0}; \gamma_0, Q_0)(\hat{Q}_n - Q_0) \right| \\ & = o\left(|\hat{\alpha}_n - \alpha_0| + \|\hat{\beta}_n - \beta_0\|_2 + \sup |\hat{\Lambda}_{n,0} - \Lambda_{0,0}|\right) \\ & + o\left(\|\hat{\gamma}_n - \gamma_0\| + \sup |\hat{Q}_n - Q_0|\right). \end{aligned} \tag{2.37}$$

Besides, Ψ' is nonsingular.

B4) $\sqrt{n}\Psi'_4(\alpha_0, \beta_0, \Lambda_{0,0}; \gamma_0, Q_0)(\hat{\gamma}_n - \gamma_0) = O_{p^*}(1)$ and $\sqrt{n}\Psi'_5(\alpha_0, \beta_0, \Lambda_{0,0}; \gamma_0, Q_0)(\hat{Q}_n - Q_0) = O_{p^*}(1)$.

Among these Conditions, B1) and B2) are satisfied from Donsker property. Specifically, using Conditions 2.2 and similar arguments in Lemma 5 of Kong and Nan (2016)'s supplementary material, we have the Donsker property of the collection $\{Q(\log v - \gamma^T x) : v \text{ is in the support of } V, \gamma \in \mathcal{G}, x \in \mathcal{X} \text{ and } Q \text{ is from the collection of distribution functions satisfying Condition 4}\}$. By Theorem 2.10.6 in Van Der Vaart and Wellner (1996) (page 192), the Lipschitz function over a Donsker class is actually Donsker as well, so $h_1 l_\alpha(\alpha, \beta, \Lambda_0; \gamma, Q)$, $h_2^T l_\beta(\alpha, \beta, \Lambda_0; \gamma, Q)$ and $l_{\Lambda_0}(\alpha, \beta, \Lambda_0; \gamma, Q)[h_3]$ are Lipschitz for $(\alpha, \beta, \Lambda_0, \gamma, Q)$ and $(h_1, h_2, h_3) \in \mathcal{H}$ and thus all belong to Donsker classes. By permanence of the Donsker property in section 2.10 of Van Der Vaart and Wellner (1996), $h_1 l_\alpha(\alpha, \beta, \Lambda_0; \gamma, Q) + h_2^T l_\beta(\alpha, \beta, \Lambda_0; \gamma, Q) + l_{\Lambda_0}(\alpha, \beta, \Lambda_0; \gamma, Q)[h_3]$ is Donsker. Assumption B1) is then verified. We see that $A_{nk}[h_k]$ ($k = 1, 2, 3$) are all bounded Lipschitz functions with respect to $(h_1, h_2, h_3) \in \mathcal{H}$ at the point $(\alpha_0, \beta_0, \Lambda_{0,0}, \gamma_0, Q_0)$, so all belong to Donsker classes. By Theorem 2.11.22 and Example 2.11.24 in Van Der Vaart and Wellner (1996) (page 220-221), $\sqrt{n}(\Psi_n - \Psi)(\alpha_0, \beta_0, \Lambda_{0,0})$ converges in distribution to a tight zero-mean Gaussian process denoted by $\mathcal{W}$. B2) is then obtained from classical central limit theorem.

It remains to show that Conditions B3)-B4) hold. Define

$$\begin{aligned} D_1(v) = & v \left[1 - \Lambda_{0,0}(Z_i) \exp(\alpha_0 v + \beta_0^T \mathbf{X}_i) \right] \exp \left[\alpha_0 v + \beta_0^T \mathbf{X}_i - \Lambda_{0,0}(Z_i) \exp(\alpha_0 v + \beta_0^T \mathbf{X}_i) \right] \\ & + \mathbf{X}_i \left[1 - \Lambda_{0,0}(Z_i) \exp(\alpha_0 v + \beta_0^T \mathbf{X}_i) \right] \exp \left[\alpha_0 v + \beta_0^T \mathbf{X}_i - \Lambda_{0,0}(Z_i) \exp(\alpha_0 v + \beta_0^T \mathbf{X}_i) \right] \\ & - \exp(2\alpha_0 v + 2\beta_0^T \mathbf{X}_i) \exp \left[-\Lambda_{0,0}(Z_i) \exp(\alpha_0 v + \beta_0^T \mathbf{X}_i) \right] \int_0^{Z_i} h_3 d\Lambda_{0,0}, \end{aligned}$$

$$\begin{aligned} D_2(v) = & -v \Lambda_{0,0}(Z_i) \exp(\alpha_0 v + \beta_0^T \mathbf{X}_i) \exp \left[-\Lambda_{0,0}(Z_i) \exp(\alpha_0 v + \beta_0^T \mathbf{X}_i) \right] \\ & - \mathbf{X}_i \Lambda_{0,0}(Z_i) \exp(\alpha_0 v + \beta_0^T \mathbf{X}_i) \exp \left[-\Lambda_{0,0}(Z_i) \exp(\alpha_0 v + \beta_0^T \mathbf{X}_i) \right] \\ & - \exp \left[\alpha_0 v + \beta_0^T \mathbf{X}_i - \Lambda_{0,0}(Z_i) \exp(\alpha_0 v + \beta_0^T \mathbf{X}_i) \right] \int_0^{Z_i} h_3 d\Lambda_{0,0}, \end{aligned}$$

$$D_3(v) = \exp(\alpha_0 v + \beta_0^T \mathbf{X}_i) \exp[-\Lambda_{0,0}(Z_i) \exp(\alpha_0 v + \beta_0^T \mathbf{X}_i)],$$

$$D_4(v) = \exp[-\Lambda_{0,0}(Z_i) \exp(\alpha_0 v + \beta_0^T \mathbf{X}_i)],$$

and

$$\begin{aligned} & \Psi'_5(\alpha_0, \beta_0, \Lambda_{0,0}; \gamma_0, Q_0)(\hat{Q}_n - Q_0) \\ &= P \left\{ \delta_2 \frac{\int_Z^\tau D_1 d(\hat{Q}_n - Q_0) \{\log v - \gamma_0^T X\}}{\int_Z^\tau D_3 dQ_0(\log v - \gamma_0^T X)} \right. \\ & \quad - \delta_2 \frac{\int_Z^\tau D_1 dQ_0(\log v - \gamma_0^T X) \int_Z^\tau D_3 d(\hat{Q}_n - Q_0) \{\log v - \gamma_0^T X\}}{[\int_Z^\tau D_3 dQ_0(\log v - \gamma_0^T X)]^2} \\ & \quad + (1 - \delta_1 - \delta_2) \frac{\int_Z^\tau D_2 d(\hat{Q}_n - Q_0) \{\log v - \gamma_0^T X\}}{\int_Z^\tau D_4 dQ_0(\log v - \gamma_0^T X)} \\ & \quad \left. - (1 - \delta_1 - \delta_2) \frac{\int_Z^\tau D_2 dQ_0(\log v - \gamma_0^T X) \int_Z^\tau D_4 d(\hat{Q}_n - Q_0) \{\log v - \gamma_0^T X\}}{[\int_Z^\tau D_4 dQ_0(\log v - \gamma_0^T X)]^2} \right\}. \end{aligned} \quad (2.38)$$

Performing similar calculations in the proof of asymptotic normality in Kong and Nan (2016), we obtain equation (2.37) via taking the first-order Taylor expansion. Next, we show that Ψ' is nonsingular. Let $\mathcal{T}$ denote the linear operator $\Psi'(\alpha_0, \beta_0, \Lambda_{0,0})$ from $\{(\alpha - \alpha_0, \beta - \beta_0, \Lambda_0 - \Lambda_{0,0}) : (\alpha, \beta, \Lambda_0) \in B_{\epsilon_0}(\alpha_0, \beta_0, \Lambda_{0,0})\}$ to $L^\infty(\mathcal{H})$, then we have

$$\begin{aligned} \mathcal{T}(\alpha - \alpha_0, \beta - \beta_0, \Lambda_0 - \Lambda_{0,0})[h_1, h_2, h_3] &= (\alpha - \alpha_0)\psi_1(h_1, h_2, h_3) \\ &+ (\beta - \beta_0)^T \psi_2(h_1, h_2, h_3) + \int_0^s \psi_3(h_1, h_2, h_3) d(\Lambda_0 - \Lambda_{0,0}), \end{aligned} \quad (2.39)$$

where $\psi_1(h_1, h_2, h_3) = \frac{\partial \Psi}{\partial \alpha}(\alpha, \beta, \Lambda_0)$, $\psi_2(h_1, h_2, h_3) = \frac{\partial \Psi}{\partial \beta}(\alpha, \beta, \Lambda_0)$ and $\psi_3(h_1, h_2, h_3) = \frac{\partial \Psi}{\partial \epsilon}(\alpha, \beta, \Lambda_\epsilon) \Big|_{\epsilon=0}$. To verify the nonsingularity of Ψ' , it is equivalent to

show that the mapping $\mathcal{T}$ is continuously invertible on its range. It in turn becomes to show the invertibility of the linear operator $\psi = (\psi_1, \psi_2, \psi_3)^T(h_1, h_2, h_3)$. This is because we have $\mathcal{T} = 0$ for any $(\alpha, \beta, \Lambda_0) \in B_{\epsilon_0}(\alpha_0, \beta_0, \Lambda_{0,0})$ if $\psi(h_1, h_2, h_3) = 0$. To this end, it requires to show that $h_1 = 0$, $h_2 = 0$, and $h_3 = 0$ when $\psi(h_1, h_2, h_3) = 0$. Let $\alpha = \alpha_0 + \epsilon h_1$, $\beta = \beta_0 + \epsilon h_2$ and $\Lambda_0(t) = \Lambda_{0,0}(t) + \epsilon \int_0^t h_3 d\Lambda_{0,0}$ for a small constant ϵ . By (2.39), we have

$$0 = \mathcal{T}(\alpha - \alpha_0, \beta - \beta_0, \Lambda_0 - \Lambda_{0,0})[h_1, h_2, h_3] = \epsilon E \left[(l_{\alpha_0}[h_1] + l_{\beta_0}[h_2] + l_{\Lambda_{0,0}}[h_3])^2 \right], \quad (2.40)$$

where l_{α_0} , l_{β_0} and $l_{\Lambda_{0,0}}$ represent the score functions with respect to α , β and Λ , respectively and evaluated at the point $(\alpha_0, \beta_0, \Lambda_{0,0})$. Equation (2.40) implies that

$$l_{\alpha_0}[h_1] + l_{\beta_0}[h_2] + l_{\Lambda_{0,0}}[h_3] = 0 \quad (2.41)$$

almost surely. Note that

$$\begin{aligned} l_{\alpha_0}[h_1] = & h_1 \left[\delta_{1i} \tilde{\delta}_{2i} Z_i - \delta_{1i} Z_i \Lambda_{0,0}(Y_i) \exp(\alpha_0 Z_i + \beta_0^T \mathbf{X}_i) \right. \\ & + \delta_{2i} R_{1i}^{(0)-1} \int_{Z_i}^{\tau} v [1 - \Lambda_{0,0}(Z_i) \exp(\alpha_0 v + \beta_0^T \mathbf{X}_i)] \\ & \times \exp [\alpha_0 v + \beta_0^T \mathbf{X}_i - \Lambda_{0,0}(Z_i) \exp(\alpha_0 v + \beta_0^T \mathbf{X}_i)] dQ_0(\log v - \gamma_0^T \mathbf{X}_i) \\ & + (1 - \delta_{1i} - \delta_{2i}) R_{2i}^{(0)-1} \int_{Z_i}^{\tau} -v \Lambda_{0,0}(Z_i) \exp(\alpha_0 v + \beta_0^T \mathbf{X}_i) \\ & \times \exp [-\Lambda_{0,0}(Z_i) \exp(\alpha_0 v + \beta_0^T \mathbf{X}_i)] dQ_0(\log v - \gamma_0^T \mathbf{X}_i) \left. \right], \end{aligned}$$

$$\begin{aligned}
l_{\beta_0}[h_2] = & h_2^T \left[\delta_{1i} \tilde{\delta}_{2i} \mathbf{X}_i - \delta_{1i} \mathbf{X}_i \Lambda_{0,0}(Y_i) \exp(\alpha_0 Z_i + \beta_0^T \mathbf{X}_i) \right. \\
& + \delta_{2i} R_{1i}^{(0)-1} \int_{Z_i}^{\tau} \mathbf{X}_i [1 - \Lambda_{0,0}(Z_i) \exp(\alpha_0 v + \beta_0^T \mathbf{X}_i)] \\
& \times \exp [\alpha_0 v + \beta_0^T \mathbf{X}_i - \Lambda_{0,0}(Z_i) \exp(\alpha_0 v + \beta_0^T \mathbf{X}_i)] dQ_0(\log v - \gamma_0^T \mathbf{X}_i) \\
& + (1 - \delta_{1i} - \delta_{2i}) R_{2i}^{(0)-1} \int_{Z_i}^{\tau} -\mathbf{X}_i \Lambda_{0,0}(Z_i) \exp(\alpha_0 v + \beta_0^T \mathbf{X}_i) \\
& \times \exp [-\Lambda_{0,0}(Z_i) \exp(\alpha_0 v + \beta_0^T \mathbf{X}_i)] dQ_0(\log v - \gamma_0^T \mathbf{X}_i) \left. \right],
\end{aligned}$$

and

$$\begin{aligned}
l_{\Lambda_{0,0}}[h_3] = & \delta_{1i} \tilde{\delta}_{2i} h_3(Y_i) + \delta_{2i} h_3(Z_i) - \delta_{1i} \exp(\alpha_0 Z_i + \beta_0^T \mathbf{X}_i) \int_0^{Y_i} h_3 d\Lambda_{0,0} \\
& + \delta_{2i} R_{1i}^{(0)-1} \int_0^{Z_i} h_3 d\Lambda_{0,0} \int_{Z_i}^{\tau} -\exp(2\alpha_0 v + 2\beta_0^T \mathbf{X}_i) \\
& \times \exp [-\Lambda_{0,0}(Z_i) \exp(\alpha_0 v + \beta_0^T \mathbf{X}_i)] dQ_0(\log v - \gamma_0^T \mathbf{X}_i) \\
& + (1 - \delta_{1i} - \delta_{2i}) R_{2i}^{(0)-1} \int_0^{Z_i} h_3 d\Lambda_{0,0} \\
& \times \int_{Z_i}^{\tau} -\exp [\alpha_0 v + \beta_0^T \mathbf{X}_i - \Lambda_{0,0}(Z_i) \exp(\alpha_0 v + \beta_0^T \mathbf{X}_i)] dQ_0(\log v - \gamma_0^T \mathbf{X}_i),
\end{aligned}$$

where

$$R_{1i}^{(0)} = \int_{Z_i}^{\tau} \exp(\alpha_0 v + \beta_0^T \mathbf{X}_i) \exp [-\Lambda_{0,0}(Z_i) \exp(\alpha_0 v + \beta_0^T \mathbf{X}_i)] dQ_0(\log v - \gamma_0^T \mathbf{X}_i),$$

and

$$R_{2i}^{(0)} = \int_{Z_i}^{\tau} \exp [-\Lambda_{0,0}(Z_i) \exp(\alpha_0 v + \beta_0^T \mathbf{X}_i)] dQ_0(\log v - \gamma_0^T \mathbf{X}_i).$$

Without loss of generality, we let $\mathbf{X}_i$ be fixed, and $\delta_{1i} = 1$, $\delta_{2i} = 0$ for all i . Assume that $\Lambda_{0,0}(0) = 0$ and we set $\tilde{\delta}_{2i} = 1$, $Z_i = Y_i = 0$. Equation (2.41)

becomes $0 = h_2^T \mathbf{X}_i + h_3(0)$, which implies $h_2 = h_3(0) = 0$ from Condition 2.11. If we let $Y_i = y$ for any $y \in (0, \varsigma]$, then the equation (2.41) turns to be $H'_3(y)/\lambda(y) = \exp(\beta_0^T \mathbf{X}_i)H_3(y)$ with boundary condition $H(0) = 0$, where $H(y) = \int_0^y h_3(s)d\Lambda_{0,0}(s)$. Thus, $h_3(y) = 0$. On the other hand, if $\tilde{\delta}_{2i} = 0$, we set $Z_i = \epsilon$ and $Y_i = \varsigma$. Combined with $h_2 = h_3 = 0$, equation (2.41) reduces to $0 = h_1 Z_i \Lambda_{0,0}(\epsilon)$, which implies $h_1 = 0$. Hence, Ψ' is nonsingular. Therefore, Condition B3) is verified. And B4) is automatically from B1)-B3).

Applying the semiparametric Z-estimation theory in Nan and Wellner (2013), we obtain

$$\begin{aligned}
& \sqrt{n}(\alpha - \alpha_0, \beta - \beta_0, \Lambda_0 - \Lambda_{0,0}) \\
&= -\Psi'(\alpha_0, \beta_0, \Lambda_{0,0})\sqrt{n} \left[(\Psi_n - \Psi)(\alpha_0, \beta_0, \Lambda_{0,0}) \right. \\
& \quad \left. + \Psi'_4(\alpha_0, \beta_0, \Lambda_{0,0}; \gamma_0, Q_0)(\hat{\gamma}_n - \gamma_0) + \Psi'_5(\alpha_0, \beta_0, \Lambda_{0,0}; \gamma_0, Q_0)(\hat{Q}_n - Q_0) \right].
\end{aligned} \tag{2.42}$$

Note that the first two terms in the right-hand side of equation (2.42) converge in distribution to zero-mean Gaussian processes, and the last term can be rewritten as

$$\begin{aligned}
& \Psi'_5(\alpha_0, \beta_0, \Lambda_{0,0}; \gamma_0, Q_0)(\hat{Q}_n - Q_0) \\
&= O_{p^*}(1)P \left\{ \delta_2 \frac{\int_Z^{\tau-} \frac{D_1}{K_Q(\log v - \gamma_0^T \mathbf{X})[1 - Q^*(\gamma_0, \log v - \gamma_0^T \mathbf{X})]} d(\hat{Q}_{n,\hat{\gamma}}^{km} - Q_0^*)\{\log v - \gamma_0^T X\}}{\int_Z^{\tau} D_3 dQ_0(\log v - \gamma_0^T X)} \right. \\
&\quad - \delta_2 \frac{\int_Z^{\tau} D_1 dQ_0(\log v - \gamma_0^T X) \int_Z^{\tau-} \frac{D_3}{K_Q(\log v - \gamma_0^T \mathbf{X})[1 - Q^*(\gamma_0, \log v - \gamma_0^T \mathbf{X})]} d(\hat{Q}_{n,\hat{\gamma}}^{km} - Q_0^*)\{\log v - \gamma_0^T X\}}{[\int_Z^{\tau} D_3 dQ_0(\log v - \gamma_0^T X)]^2} \\
&\quad + (1 - \delta_1 - \delta_2) \frac{\int_Z^{\tau-} \frac{D_2}{K_Q(\log v - \gamma_0^T \mathbf{X})[1 - Q^*(\gamma_0, \log v - \gamma_0^T \mathbf{X})]} d(\hat{Q}_{n,\hat{\gamma}}^{km} - Q_0^*)\{\log v - \gamma_0^T X\}}{\int_Z^{\tau} D_4 dQ_0(\log v - \gamma_0^T X)} \\
&\quad - (1 - \delta_1 - \delta_2) \frac{\int_Z^{\tau} D_2 dQ_0(\log v - \gamma_0^T X)}{[\int_Z^{\tau} D_4 dQ_0(\log v - \gamma_0^T X)]^2} \\
&\quad \left. \times \int_Z^{\tau-} \frac{D_4}{K_Q(\log v - \gamma_0^T \mathbf{X})[1 - Q^*(\gamma_0, \log v - \gamma_0^T \mathbf{X})]} d(\hat{Q}_{n,\hat{\gamma}}^{km} - Q_0^*)\{\log v - \gamma_0^T X\} \right\}, \tag{2.43}
\end{aligned}$$

thus converges to zero-mean Gaussian processes as well. This completes the proof.

Chapter 3

Quantile Regression for Survival Data with Covariates Subject to Detection Limits

3.1 Introduction

All the aforementioned works in subsection 1.3.3 concentrated on the censored quantile regression with fully observed covariates. However, in some medical and epidemiological settings, survival times are associated with risk factors, which are censored due to limits of detection. This chapter is motivated by a study of the genetic and inflammatory markers of sepsis (GenIMS) in Subsection 1.4.2.

The existing analysis often focuses on mean-based regression for such data with both outcomes and risk factors subject to censoring. Based on the normal distribution assumption for the censored covariates, D'Angelo and Weissfeld (2008) proposed a two-stage estimation procedure with index approach for the Cox model with censored covariates. Kim and Kong

(2015) studied the Cox model with an interval-censored covariate, where their likelihood-based method allowed a single covariate only. Bernhardt et al. (2014) utilized a classic multiple imputation method for an accelerated failure time model with biomarkers as censored covariates and the error term following a seminonparametric distribution. However, their method required a parametric distribution assumption for the censored covariates.

To overcome limitations in the existing works above, we propose a more flexible model using quantile regression, which is free of any parametric distribution assumptions for the outcome variable as well as the censored covariates. We develop an estimation procedure using a multiple imputation approach along similar lines of the work by Wang and Feng (2012) to handle censored covariates.

In the rest of this chapter, first we present models of interest, including the censored quantile regression with/without censored covariates, and develop their corresponding estimation procedures. Then we provide theoretical justifications for the proposed estimator. Next, We report simulation results and illustrate the utility of our method with analysis of the motivating dataset. Finally, we conclude the paper with some remarks. Detailed proofs of theoretical properties are deferred to the Appendix.

3.2 Methodology

Our overall goal is to understand how the distribution of a survival time T depends on a set of covariates $\mathbf{A}$ in which some covariates could be subject to censoring. We first revisit quantile regression analysis for survival data when all covariates are exactly observed, and then introduce the proposed models and estimation procedures for different cases with censored covariates.

3.2.1 Censored Quantile Regression with Fully Observed Covariates

Suppose a sample for analysis consists of n individuals. For each individual i , $i = 1, \dots, n$, we denote the right-censored survival time by T_i and the p -dimensional covariate vector by $\mathbf{A}_i$. We assume that the right censoring time, denoted by C_i , is independent of T_i conditional on $\mathbf{A}_i$. Due to right random censoring, T_i is observed only up to (Y_i, δ_i) , where $Y_i = \min(T_i, C_i)$ and the censoring indicator $\delta_i = I(T_i < C_i)$. For any quantile level $\tau \in (0, 1)$, the τ -th conditional quantile of T_i given covariate vector $\mathbf{A}_i$ is defined as $Q_{T_i}(\tau|\mathbf{A}_i) = \inf\{t : \Pr(T_i \leq t|\mathbf{A}_i) \geq \tau\}$. To build the relationship between T_i and $\mathbf{A}_i$, we assume the following censored quantile regression (CQR) model,

$$Q_{T_i}(\tau|\mathbf{A}_i) = \exp[\mathbf{A}_i^T \boldsymbol{\beta}(\tau)], \quad (3.1)$$

where $\boldsymbol{\beta}(\tau)$ is an unknown p -dimensional vector of regression coefficients and may vary across different values of quantile level τ . When $\mathbf{A}_i$ are observed for all subjects and no censoring in covariates, according to Peng and Huang (2008), $\boldsymbol{\beta}(\tau)$ can be estimated through solving the following martingale-based estimating equation

$$n^{-1/2} \sum_{i=1}^n \mathbf{s}_i(\boldsymbol{\beta}, \tau|\mathbf{A}_i, Y_i, \delta_i) = 0,$$

where

$$\mathbf{s}_i(\boldsymbol{\beta}, \tau|\mathbf{A}_i, Y_i, \delta_i) = \mathbf{A}_i \left\{ N_i[\exp(\mathbf{A}_i^T \boldsymbol{\beta}(\tau))] - \int_0^\tau I[Y_i \geq \exp(\mathbf{A}_i^T \boldsymbol{\beta}(u))] dH(u) \right\}, \quad (3.2)$$

$N_i(t) = \delta_i I(Y_i \leq t)$ and $H(u) = -\log(1 - u)$ for $0 \leq u < 1$.

3.2.2 Censored Quantile Regression with Univariate Left-censored Covariate

If a subset of covariates in $\mathbf{A}_i$ are censored, the direct use of the observed values of censored covariates may cause bias in estimation. We propose multiple imputation methods to handle those censored covariates hereafter. First, we consider that there is a univariate covariate V_i subjected to left-censoring. Without loss of generality, we assume that the censoring occurs in the first covariate in $\mathbf{A}_i$. Then $\mathbf{A}_i$ can be written in a partitioned vector form, that is, $\mathbf{A}_i = (V_i, \mathbf{X}_i^T)^T$, where $\mathbf{X}_i \in \mathbb{R}^{p_2}$ is the vector of uncensored covariates and $p_2 = p - 1$. The case of multivariate censored covariates will be discussed later in subsection 3.2.4. Suppose that d is a given constant known as a “detection limit” in the biomedical research. Let $Z_i = \max(V_i, d)$ be the observation of V_i with corresponding censoring indicator $\delta_i^v = I(V_i > d)$. Given both completely observed covariates $\mathbf{X}_i$ and censored covariate V_i , model (3.1) becomes

$$Q_{T_i}(\tau|V_i, \mathbf{X}_i) = \exp[\alpha(\tau)V_i + \mathbf{X}_i^T \boldsymbol{\beta}_1(\tau)], \quad 0 < \tau < 1, \quad (3.3)$$

where $\alpha(\tau)$ and $\boldsymbol{\beta}_1(\tau)$ are coefficients associated with V_i and $\mathbf{X}_i$, respectively. Since V_i is censored, a common way to deal with such a censored covariate is based on multiple imputation, which requires a distribution assumption on V_i . For censored observations of V_i , Wang and Feng (2012) suggested to impute them by conditional quantiles of their distribution given the remaining uncensored covariates. Their method not only inherits the spirit of traditional multiple imputation methods, but also possesses robustness against both outlying covariate data and derivation from the parametric model assumption for censored covariates. This inspires us to consider a flexible imputation model in the current work. We model the

relationship between quantiles of V_i and covariate information through the censored quantile regression for V_i in the form of

$$Q_{V_i}(u|\mathbf{W}_i) = \mathbf{W}_i^T \boldsymbol{\eta}(u), \quad 0 < u < 1, \quad (3.4)$$

where $\boldsymbol{\eta}(u)$ is an unknown vector of regression coefficients at quantile level u , and $\mathbf{W}_i$ represents a set of user-specified covariates associated with V_i . For example, $\mathbf{W}_i$ can be specified as a set of all uncensored covariates $\mathbf{X}_i$, or a triplet consisting of $\mathbf{X}_i$, the log survival time and corresponding censoring indicator, i.e., $\mathbf{W}_i = (\mathbf{X}_i, \log Y_i, \delta_i)$. The performance of different choices of $\mathbf{W}_i$ will be compared in our numerical studies in the following section. The imputation model assumes that the conditional distribution function of V given $(T, C, \mathbf{X})$ is equivalent to the conditional distribution function of V given $\mathbf{W}$.

Denote the censoring probability by $\pi(\mathbf{W}_i) = Pr(V_i < d|\mathbf{W}_i) = Pr(\delta_i^v = 0|\mathbf{W}_i)$. It can be estimated via logistic regression

$$\log \frac{\pi(\mathbf{W}_i)}{1 - \pi(\mathbf{W}_i)} = \mathbf{W}_i^T \boldsymbol{\lambda}_0, \quad (3.5)$$

where $\boldsymbol{\lambda}_0$ is an unknown vector of coefficients to be estimated.

Due to left censoring for V , there is an lower identifiability limit such that the parameters in model (3.4) cannot be estimated under some lower quantile levels. Denote the smallest identifiable quantile level in model (3.4) by ξ_1 . In fact, $\xi_1 \leq \inf\{u : \mathbf{W}_i^T \boldsymbol{\eta}(u) = Z_i, \delta_i^v = 0, \text{ and } i = 1, \dots, n.\}$. There is also an identifiability issue in model (3.3) due to right-censoring for T . Let τ_u be the largest identifiable quantile level in model (3.3). τ_u shall satisfy $Pr\{C \geq \exp[\mathbf{A}^T \boldsymbol{\beta}(\tau_u)]\} > 0$ and be unknown in practice. Applying the multiple (quantile) imputation for censored covariate V_i , we propose the following estimation procedure to estimate parameters of interest $\boldsymbol{\beta}(\tau) = (\alpha(\tau), \boldsymbol{\beta}_1^T(\tau))^T$ in model (3.3).

Step 1. Employ R function `glm()` to fit model (3.5) and denote the obtained estimate of λ_0 by $\hat{\lambda}$. Plugging $\hat{\lambda}$ in (3.5) gives the estimate $\hat{\pi}(\mathbf{W}_i)$ for $\pi(\mathbf{W}_i)$.

Step 2. Find $\hat{\eta}(u) = \arg \min \sum_{i=1}^n \rho_u [Z_i - \max(\mathbf{W}_i^T \eta, d)]$, where $\rho_u(y) = y[u - I(y < 0)]$.

Step 3. For individual i with $\delta_i^v = 0$ and $j = 1, \dots, m$, generate $u_{i(j)}$ from $\text{Uniform}[0, \hat{\pi}(\mathbf{W}_i)]$ and let $V_i^{*(j)} = \mathbf{W}_i^T \hat{\eta}\{u_{i(j)}\}$ if $u_{i(j)} \geq \xi_1$.

Step 4. Find the solution, denoted by $\hat{\beta}^{(j)}(\tau)$, of the estimating equation

$$n^{-1/2} \sum_{i=1}^n \left\{ \delta_i^v \mathbf{s}_i(\beta, \tau | \mathbf{A}_i^{(j)}, Y_i, \delta_i) + (1 - \delta_i^v) \mathbf{s}_i(\beta, \tau | \mathbf{A}_i^{*(j)}, Y_i, \delta_i) I[V_i^{*(j)} < d, u_{i(j)} \geq \xi_1] \right\} = 0,$$

where $\mathbf{A}_i^{*(j)} = (V_i^{*(j)}, \mathbf{X}_i^T)^T$.

Step 5. Repeat steps 3 and 4 for m times, the multiple imputation estimate is then computed by $\hat{\beta}_{MI}(\tau) = \frac{1}{m} \sum_{j=1}^m \hat{\beta}^{(j)}(\tau)$.

To implement the optimization in Step 2, we can make use of the existing R codes provided by Portnoy (2003) or Peng and Huang (2008). Based on our experience, the use of both of their codes in Step 2 produces only minor differences in the final result of the estimation procedure. Therefore, we apply Portnoy (2003)'s the code only for Step 2 in the following numerical studies.

It is worth to note that one may encounter the quantile crossing problem in the implementation of the estimation procedure above. To fix this problem, we suggest to further use the quantile rearrangement technique

proposed by Chernozhukov et al. (2010) via function `rearrange()` available in R package `quantreg` in the beginning of Step 3.

3.2.3 Censored Quantile Regression with Univariate Right-censored Covariate

In the case where one univariate covariate V_i is subject to right censoring and observed as $Z_i = \min(V_i, d)$ together with the censoring indicator $\delta_i^v = I(V_i < d)$, the censoring probability is $\pi(\mathbf{W}_i) = Pr(\delta_i^v = 0 | \mathbf{W}_i)$. Let ξ_2 be the largest identifiable quantile level in model (3.4). To incorporate the multiple imputations for the right-censored covariate V_i , we propose to apply the estimation procedure available in subsection 3.2.2 with Steps 2-4 being updated as follows:

Step 2'. Compute $\hat{\boldsymbol{\eta}}(u) = \arg \min \sum_{i=1}^n \rho_u [Z_i - \min(\mathbf{W}_i^T \boldsymbol{\eta}, d)]$.

Step 3'. For i with $\delta_i^v = 0$ and $j = 1, \dots, m$, generate $u_{i(j)} \sim \text{Uniform}[1 - \hat{\pi}(\mathbf{W}_i), 1]$, and then let $V_i^{*(j)} = \mathbf{W}_i^T \hat{\boldsymbol{\eta}}\{u_{i(j)}\}$ if $u_{i(j)} \leq \xi_2$.

Step 4'. Obtain $\hat{\boldsymbol{\beta}}^{(j)}(\tau)$ from the estimating equation

$$n^{-1/2} \sum_{i=1}^n \left\{ \delta_i^v \mathbf{s}_i(\boldsymbol{\beta}, \tau | \mathbf{A}_i^{(j)}, Y_i, \delta_i) + (1 - \delta_i^v) \mathbf{s}_i(\boldsymbol{\beta}, \tau | \mathbf{A}_i^{*(j)}, Y_i, \delta_i) I[V_i^{*(j)} > d, u_{i(j)} \leq \xi_2] \right\}.$$

3.2.4 Censored Quantile Regression with Multiple Censored Covariates

Let $V_i = (V_{i1}, \dots, V_{ip_1})$ be a p_1 -dimensional vector of latent censored covariates with the corresponding vector of censoring indicators $\delta_i^v = (\delta_{i1}^v, \dots, \delta_{ip_1}^v)$,

where $p_1 = p - p_2$. The multiple imputation given in Subsection 3.2.2 or 3.2.3 can be similarly implemented for each censored observations in V_i . We start the imputation with the first censored covariate V_{i1} , and implement Steps 1-3 of the estimation procedure in Subsection 3.2.2 or Steps 1-2 and 3' in Subsection 3.2.3 for left- or right-censored V_{i1} . Then $V_{i1}^{*(j)}$ can be formed. For the second censored covariate V_{i2} , we replace V_{i2} by $V_{i2}^{*(j)} = \hat{Q}_{V_{i2}}(u_{2,i(j)} | \mathbf{W}_i, V_{i1}^{*(j)})$, where $u_{k,i(j)}$ is from Uniform $[0, \pi_k(\mathbf{W}_i)]$ or Uniform $[1 - \pi_k(\mathbf{W}_i), 1]$ for left- or right-censored V_{i2} , and $\pi_k(\mathbf{W}_i) = Pr(\delta_{ik}^v = 0 | \mathbf{W}_i)$ for $k = 1, \dots, p_1$. Similarly, we replace V_{ik} by $V_{ik}^{*(j)} = \hat{Q}_{V_{ik}}(u_{k,i(j)} | \mathbf{W}_i, V_{i1}^{*(j)}, \dots, V_{i(k-1)}^{*(j)})$ for $k = 2, \dots, p_1$. The multiple imputation estimate thereafter can be obtained using the steps analogous to Steps 4-5 in subsection 2.2 for right-censored covariates, while Step 4 has to be replaced by Step 4' if the underlying covariate is left-censored. In numerical examples, it is suggested that a censored covariate with a lower censoring rate has higher preference as a prior target which requires imputation.

If we delete some individuals whose covariates are not exactly observed, the analysis for the remaining data is regarded as complete case (CC) analysis. It is known that the complete case (CC) analysis leads to a consistent estimator but with large standard error due to not making full use of all observed information. Like other classical multiple imputation methods, the proposed method uses the observed data from censored cases and thus improves the efficiency of the estimator, but may produce bigger bias than the CC analysis. To trade off between bias and variance, a shrinkage approach was suggested by Chen et al. (2009), and utilized by Wang and Feng (2012) and Wei et al. (2012) in their multiple imputation procedures for quantile regression. Specifically, let $\hat{\beta}_{CC}(\tau)$ be the CC estimator of $\beta(\tau)$. Let $(v_1, \dots, v_p)^T$ be the diagonal elements of $\text{var}(\hat{\beta}_{MI}(\tau) - \hat{\beta}_{CC}(\tau))$. At a

quantile level τ , we define the shrinkage estimator $\hat{\beta}_S(\tau)$ following the work of Chen et al. (2009) as

$$\hat{\beta}_S(\tau) = \hat{\beta}_{CC}(\tau) + \mathbf{K}(\tau)[\hat{\beta}_{MI}(\tau) - \hat{\beta}_{CC}(\tau)],$$

where $\hat{\beta}_{kMI}(\tau)$ and $\hat{\beta}_{kCC}(\tau)$ are the k -th elements of $\hat{\beta}_{MI}(\tau)$ and $\hat{\beta}_{CC}(\tau)$, respectively, and $\mathbf{K}(\tau)$ is a $p \times p$ diagonal matrix with the k -th diagonal element given by $v_k / \{v_k + [\hat{\beta}_{kMI}(\tau) - \hat{\beta}_{kCC}(\tau)]^2\}$ for $k = 1, \dots, p$. Denote by $\mathbf{I}_p$ a $p \times p$ identity matrix. One can obtain the variance of $\hat{\beta}_S(\tau)$ by

$$(\mathbf{K}(\tau), \mathbf{I}_p - \mathbf{K}(\tau)) \text{cov} \left(\hat{\beta}_{MI}(\tau), \hat{\beta}_{CC}(\tau) \right) (\mathbf{K}(\tau), \mathbf{I}_p - \mathbf{K}(\tau)). \quad (3.6)$$

3.3 Theoretical Justification

To establish the asymptotic properties, we consider the case with a single left-censored covariate only for simplicity in our presentation, while results for the cases with a single right-censored covariate or multiple covariates subject to left- or right-censoring can be shown in a similar manner. It is noted that the estimators $\hat{\eta}(u)$ and $\hat{\beta}_{MI}(\tau)$ for true parameters η_0 and β_0 are right continuous piece-wise functions over the grid points $0 < \xi_1 \leq u_1 < \dots < u_{k_{1n}} < 1$ and $0 < \tau_1 < \dots < \tau_{k_{2n}} \leq \tau_u < 1$. To shown theoretical properties, we need the following mild technical conditions:

Condition 3.1. For those whose latent covariates cannot be exactly observed, there is a positive constant π_0 such that the true censoring probability $\pi(\mathbf{W}_i) > \pi_0 > \xi_1$.

Condition 3.2. a) $\sup \|\mathbf{A}_i\|_2 = O(1)$ where $\|\cdot\|_2$ represents the 2-norm, b) $E(\mathbf{A}\mathbf{A}^T) > 0$, c) $\sup \|\mathbf{W}_i\|_2 = o(n^{1/2-\epsilon_0})$ for a small positive constant ϵ_0 .

Condition 3.3. $\hat{\lambda} - \lambda_0 = \frac{1}{n} \sum_{i=1}^n D(V_i, \mathbf{W}_i, \delta_i^v) + o_p(n^{-\frac{1}{2}})$ where $D(V_i, \mathbf{W}_i, \delta_i^v)$ is Donsker class and $\frac{1}{n} \sum_{i=1}^n D(V_i, \mathbf{W}_i, \delta_i^v)$ converges to a tight Gaussian process.

Condition 3.4. a) $\eta_0(u)$ belonging to a compact space Ω_0 is smooth and its first-order derivative $\eta'_0(u)$ is Lipschitz for $0 < u < 1$.

b) For those whose latent covariates cannot be exactly observed, the density $\{||\mathbf{W}_i||_2^{-1} \mathbf{W}_i^T \eta'_0(u)\}^{-1}$ of V_i is almost surely bounded uniformly on $u \in (0, 1)$.

Condition 3.5. $\hat{\eta}(u) - \eta_0(u) = \frac{1}{n} \sum_{i=1}^n e(V_i, \mathbf{W}_i, \delta_i^v, u) + R_n(u)$ where $e(V_i, \mathbf{W}_i, \delta_i^v, u)$ is Donsker class, $\frac{1}{n} \sum_{i=1}^n e(V_i, \mathbf{W}_i, \delta_i^v, u)$ is a tight Gaussian process for a given u , and $|R_n(u)| = o_p(n^{-\frac{1}{2}})$ uniformly on $u \in [\xi_1, 1)$.

Remark 3.1. We give the definition of Donsker class as follows. Given a function f , denote the expectation of f under P by Pf and the expectation of f under empirical measure by $\mathbb{P}_n f$. Let the empirical process $\mathbb{G}_n(f) = \sqrt{n}(\mathbb{P}_n f - Pf)$. A class $\mathcal{F}$ is a Donsker if $\mathbb{G}_n$ weakly converges to $\mathbb{G}$ in $L^\infty(\mathcal{F})$ for a tight Borel measurable element $\mathbb{G}$ in $L^\infty(\mathcal{F})$.

Theorem 3.1. Under Conditions 3.1-3.5 and Condition 3.6 in the Appendix, as $k_{1n} \rightarrow \infty$, $n^{1/2}k_{1n}^{-1} \rightarrow 0$ and $k_{2n} \rightarrow \infty$, the multiple imputed estimator is a uniformly consistent estimator of $\beta(\tau)$, that is, $\sup_{\tau \in (0, \tau_u]} ||\hat{\beta}_{MI}(\tau) - \beta_0(\tau)||_2 \rightarrow 0$.

Theorem 3.2. Under Conditions 3.1-3.5 and Conditions 3.6-3.8 in the Appendix, as $k_{1n} \rightarrow \infty$, $n^{1/2}k_{1n}^{-1} \rightarrow 0$, $k_{2n} \rightarrow \infty$ and $n^{1/2}k_{2n}^{-1} \rightarrow 0$, $\sqrt{n}(\hat{\beta}_{MI}(\tau) - \beta_0(\tau))$ converges weakly to a Gaussian process for $\tau \in (0, \tau_u]$.

Remark 3.2. Conditions 3.1-3.5 pose some restrictions for the imputation models, the censoring probability function for V and their estimators. In Conditions 3.3 and 3.5, the asymptotic properties of the estimators for $\boldsymbol{\lambda}$ and $\boldsymbol{\eta}$ are verified by existing works. Detail can be found in Portnoy (2003) for the estimator $\hat{\boldsymbol{\eta}}$, and Chapter 5 of McCulloch and Searle (2005) for the estimator $\hat{\lambda}$. Conditions 3.2 a)-b) are required in the working models (3.4) and (3.5) as well as the main model (3.1).

Remark 3.3. The expression of the asymptotic covariance matrix of $\hat{\boldsymbol{\beta}}_{MI}(\tau)$ is very complicated. Thereby we use a simple bootstrap approach by re-sampling the quintuple $(Y_i, Z_i, \mathbf{X}_i, \delta_i, \delta_i^v)$ with replacement.

Remark 3.4. The extensions of Theorems 3.1 and 3.2 to right-censored or multiple censored covariates can be developed through some tedious but straightforward calculations, which are akin to those used in the technical appendices of Wang and Feng (2012).

3.4 Simulation Results

In this section, we conduct simulation studies to evaluate the performance of our proposed method and compare it across four different multiple imputation models with different choices of $\mathbf{W}_i$, referred to as “MI-logY”, “MI-logY*”, “MI-noY” and “MI-noY*”. We take two types of the user-specified matrix $\mathbf{W}$ in the imputation model into account. For the method “MI-logY”, we impute the conditional quantile $Q_{V_{i1}}(u | \log Y_i, \delta_i, \mathbf{X}_i)$ for those censored V_{i1} , while the imputation model $Q_{V_{i1}}(u | \mathbf{X}_i)$ is used in “MI-noY”. When the shrinkage approach is considered in the final estimation step for both “MI-logY” and “MI-noY”, the corresponding methods are named as

“MI-logY*” and “MI-noY*”, respectively. We also compare the proposed method with the naive methods “CC” and “fill-d”, where “CC” refers to complete case analysis based on censored quantile regression ignoring individuals having censored covariates, and “fill-d” uses the detection limits as if they were the latent true values of censored covariates. In addition, we define the “full” method in which all exact observations of V_i are assumed to be known even if $\delta_i^v = 0$. The performance of the “full” method is expected to be the best as it uses the most information. Therefore it serves as a benchmark over which the performances of the other methods are compared.

We consider that both censored and uncensored covariates are bivariate in the simulation. We set the sample size $n = 200$. For each subject $i = 1, \dots, 200$, we generate covariates x_{i1} from a uniform distribution over the interval $[-1, 1]$, and x_{i2} from a Bernoulli distribution with probability of success as 0.5. The latent bivariate censored covariate $\mathbf{V}_i$ is generated by a linear regression model

$$\mathbf{V}_i = \begin{pmatrix} V_{i1} \\ V_{i2} \end{pmatrix} = 0.5x_{i1} + 0.5x_{i2} + \epsilon_i^v, \quad \epsilon_i^v \sim N \left(\mathbf{0}, \begin{pmatrix} 1 & 0.4 \\ 0.4 & 0.6 \end{pmatrix} \right).$$

Next, the survival time is generated from the following lognormal AFT model with heteroscedastic errors:

$$\log T_i = 0.5V_{i1} + 0.5V_{i2} - 0.5x_{i1} + 0.5x_{i2} + (1 + 0.2x_{i1})\epsilon_i, \quad \epsilon_i \sim N(0, 0.5^2),$$

where the error term is assumed to be associated with covariates. The setting makes that in the quantile regression of T , the parameter $\beta_1(\tau)$ corresponding to X_{1i} is varying with the quantile level τ . Specifically, on the mean level, the vector of true coefficients is $(\beta_0, \alpha_1, \alpha_2, \beta_1, \beta_2) =$

$(0, 0.5, 0.5, -0.5, 0.5)$. The conditional quantile function of T_i can be specified as

$$Q_{T_i}(\tau) = \exp[\beta_0(\tau) + \alpha_1(\tau)V_{1i} + \alpha_2(\tau)V_{2i} + \beta_1(\tau)X_{1i} + \beta_2(\tau)X_{2i}],$$

where the intercept $\beta_0(\tau)$, which is the inverse of the cumulative distribution function of $N(0, 0.5^2)$, varies with τ , $\alpha_1(\tau), \alpha_2(\tau), \beta_2(\tau)$ associated with V_{i1}, V_{i2}, X_{i2} , respectively are unvarying and all equal 0.5. V_{i1} is left-censored and V_{i2} is right-censored. The detection limits d_1 and d_2 for censored covariates V_{i1} and V_{i2} , respectively, are chosen to gain a expected combination of censoring rates. We assume that the censoring time C_i is covariate-dependent. Specifically, for a negative value of x_{i1} , C_i is simulated from a uniform distribution over interval $[0, C_m]$; otherwise C_i is simulated from a uniform distribution over interval $[1, C_m]$. T_i can be observed if $T_i \leq C_i$, or censored by C_i . The value of C_m is determined to achieve the desired censoring rate for T .

Based on 500 replicates, simulation results obtained by various estimation methods are summarized in terms of mean bias, standard deviation (SD), and relative efficiency (RE), where the RE is defined by the ratio of the mean square error (MSE) obtained by the benchmark to its MSE. The higher RE implies a more efficient estimator. What is more, we compute standard error (SE) based on 100 Bootstrap re-samples for methods “MI-logY” and “MI-noY”, while SEs of the estimates obtained by the shrinkage methods “MI-logY*” and “MI-noY*” are calculated using the formula in Eq (3.6).

We use CR.V1, CR.V2 and CR.T to represent the censored rates of V_1, V_2 and T , respectively. Based on our experience, the censoring rate for T would slightly affect the performance of the proposed method. To save the

Table 3.1: *Simulation results based on 500 simulation runs for the proposed method with various choices of $\mathbf{W}_i$ in the imputation model in comparison with the “full”, “CC” and “fill-d” methods, where $CR.V1=20\%$, $CR.V2=20\%$ and $CR.T=40\%$.*

method		$\tau = 0.3$					$\tau = 0.5$				
		β_0	α_1	α_2	β_1	β_2	β_0	α_1	α_2	β_1	β_2
full	bias	0.024	0.003	0.004	0.003	0.002	0.024	0.002	0.006	0.004	0.006
	SD	0.072	0.065	0.088	0.100	0.113	0.068	0.069	0.089	0.108	0.110
	RE($\times 100$)	/	/	/	/	/	/	/	/	/	/
CC	bias	0.018	0.008	0.006	-0.001	0.001	0.023	0.005	0.012	-0.003	0.004
	SD	0.100	0.120	0.129	0.125	0.148	0.099	0.120	0.132	0.131	0.145
	RE($\times 100$)	55	30	46	63	59	50	33	46	68	58
fill-d	bias	-0.060	0.104	0.103	0.013	0.025	-0.043	0.107	0.098	0.008	0.020
	SD	0.079	0.096	0.096	0.106	0.124	0.074	0.098	0.096	0.110	0.116
	RE($\times 100$)	57	21	39	87	81	70	22	43	96	89
MI-logY	bias	0.008	0.010	0.013	0.008	0.007	0.014	0.010	0.015	0.001	0.004
	SD	0.077	0.101	0.100	0.097	0.118	0.075	0.116	0.103	0.103	0.111
	SE	0.080	0.098	0.106	0.112	0.122	0.082	0.109	0.107	0.115	0.125
	RE($\times 100$)	95	41	76	104	92	88	35	74	112	99
MI-logY*	bias	0.014	0.010	0.009	0.003	0.004	0.019	0.010	0.013	-0.001	0.003
	SD	0.085	0.098	0.106	0.104	0.127	0.084	0.102	0.108	0.113	0.122
	SE	0.082	0.097	0.106	0.114	0.125	0.083	0.105	0.108	0.117	0.128
	RE($\times 100$)	77	44	69	91	80	70	45	69	93	82
MI-noY	bias	-0.034	0.027	0.021	0.006	0.042	-0.014	0.011	0.029	0.006	0.034
	SD	0.082	0.124	0.113	0.127	0.122	0.084	0.127	0.110	0.122	0.127
	SE	0.091	0.125	0.120	0.132	0.138	0.091	0.133	0.121	0.136	0.143
	RE($\times 100$)	66	27	54	70	73	72	28	54	76	74
MI-noY*	bias	-0.006	0.024	0.012	0.000	0.026	0.001	0.016	0.016	0.000	0.022
	SD	0.086	0.099	0.107	0.127	0.126	0.086	0.102	0.110	0.122	0.129
	SE	0.090	0.118	0.116	0.130	0.136	0.090	0.126	0.119	0.134	0.141
	RE($\times 100$)	69	42	61	70	73	70	44	56	76	75

Note: RE is the relative efficiency of a estimator with respect to method "full".

Table 3.2: *Simulation results based on 500 simulation runs for the proposed method with various choices of $\mathbf{W}_i$ in the imputation model in comparison with the “full”, “CC” and “fill-d” methods, where $CR.V1=20\%$, $CR.V2=40\%$ and $CR.T=40\%$.*

method		$\tau = 0.3$					$\tau = 0.5$				
		β_0	α_1	α_2	β_1	β_2	β_0	α_1	α_2	β_1	β_2
full	bias	0.023	0.004	0.008	-0.008	0.006	0.022	0.003	0.010	-0.008	0.007
	SD	0.069	0.062	0.087	0.106	0.114	0.070	0.062	0.088	0.106	0.117
	RE($\times 100$)	/	/	/	/	/	/	/	/	/	/
CC	bias	0.017	0.010	0.011	-0.006	0.008	0.020	0.006	0.024	-0.011	0.010
	SD	0.120	0.136	0.173	0.151	0.169	0.114	0.131	0.175	0.151	0.164
	RE($\times 100$)	36	21	25	50	46	40	23	25	49	51
fill-d	bias	-0.004	0.144	0.172	0.022	0.042	0.013	0.141	0.170	0.021	0.039
	SD	0.079	0.085	0.114	0.112	0.119	0.082	0.094	0.114	0.112	0.120
	RE($\times 100$)	84	14	18	87	82	79	13	19	86	86
MI-logY	bias	0.025	0.002	0.050	-0.002	-0.001	0.025	0.003	0.042	-0.006	-0.003
	SD	0.085	0.114	0.137	0.109	0.117	0.085	0.122	0.130	0.109	0.117
	SE	0.086	0.103	0.131	0.116	0.125	0.086	0.112	0.125	0.116	0.125
	RE($\times 100$)	68	30	36	96	96	69	26	42	94	100
MI-logY*	bias	0.021	0.008	0.025	-0.003	0.001	0.021	0.009	0.029	-0.010	0.003
	SD	0.095	0.105	0.137	0.121	0.134	0.092	0.104	0.138	0.123	0.131
	SE	0.089	0.105	0.133	0.120	0.131	0.089	0.111	0.128	0.122	0.130
	RE($\times 100$)	57	35	39	77	73	61	35	39	74	79
MI-noY	bias	-0.027	0.031	0.014	0.025	0.060	-0.005	0.020	0.032	0.016	0.058
	SD	0.087	0.120	0.130	0.124	0.123	0.089	0.128	0.130	0.116	0.123
	SE	0.099	0.130	0.141	0.137	0.146	0.100	0.138	0.142	0.137	0.148
	RE($\times 100$)	63	27	41	72	63	65	26	40	83	65
MI-noY*	bias	-0.007	0.030	0.003	0.016	0.035	0.002	0.021	0.014	0.006	0.040
	SD	0.091	0.111	0.133	0.131	0.129	0.094	0.118	0.134	0.126	0.133
	SE	0.098	0.125	0.139	0.135	0.146	0.100	0.133	0.141	0.136	0.148
	RE($\times 100$)	63	32	40	67	67	59	30	39	72	63

Note: RE is the relative efficiency of a estimator with respect to method ”full”.

Table 3.3: *Simulation results based on 500 simulation runs for the proposed method with various choices of $\mathbf{W}_i$ in the imputation model in comparison with the “full”, “CC” and “fill-d” methods, where $CR.V1=40\%$, $CR.V2=40\%$ and $CR.T=40\%$.*

method		$\tau = 0.3$					$\tau = 0.5$				
		β_0	α_1	α_2	β_1	β_2	β_0	α_1	α_2	β_1	β_2
full	bias	0.023	0.013	-0.001	-0.004	0.003	0.027	0.009	0.004	0.002	0.003
	SD	0.071	0.066	0.081	0.106	0.116	0.075	0.066	0.086	0.106	0.122
	RE($\times 100$)	/	/	/	/	/	/	/	/	/	/
CC	bias	0.005	0.039	-0.001	-0.006	0.017	0.009	0.049	0.006	-0.011	0.016
	SD	0.218	0.247	0.254	0.211	0.219	0.221	0.267	0.253	0.198	0.226
	RE($\times 100$)	12	7	10	25	28	13	6	12	29	29
fill-d	bias	-0.212	0.280	0.264	0.041	0.056	-0.175	0.286	0.245	0.041	0.059
	SD	0.110	0.140	0.124	0.118	0.131	0.109	0.163	0.114	0.116	0.129
	RE($\times 100$)	10	5	8	72	67	15	4	10	74	74
MI-logY	bias	-0.034	0.068	0.036	-0.002	-0.010	-0.029	0.088	0.015	-0.012	-0.019
	SD	0.119	0.161	0.138	0.111	0.124	0.128	0.185	0.134	0.113	0.126
	SE	0.108	0.153	0.139	0.122	0.131	0.114	0.179	0.137	0.125	0.135
	RE($\times 100$)	37	15	32	92	86	36	10	41	87	92
MI-logY*	bias	-0.009	0.053	0.012	-0.004	0.005	-0.007	0.069	0.012	-0.014	-0.001
	SD	0.157	0.174	0.173	0.150	0.161	0.154	0.191	0.166	0.137	0.161
	SE	0.126	0.162	0.157	0.136	0.145	0.133	0.184	0.156	0.139	0.151
	RE($\times 100$)	23	14	22	50	52	27	11	27	59	57
MI-noY	bias	-0.144	0.083	0.088	0.025	0.084	-0.097	0.073	0.098	0.017	0.074
	SD	0.121	0.172	0.142	0.140	0.154	0.117	0.174	0.136	0.142	0.154
	SE	0.140	0.186	0.163	0.167	0.176	0.139	0.196	0.163	0.169	0.178
	RE($\times 100$)	16	14	30	53	42	22	14	30	57	44
MI-noY*	bias	-0.047	0.048	0.040	0.002	0.044	-0.038	0.049	0.053	-0.005	0.040
	SD	0.156	0.181	0.188	0.157	0.184	0.143	0.188	0.179	0.162	0.184
	SE	0.152	0.185	0.174	0.167	0.177	0.151	0.198	0.177	0.173	0.183
	RE($\times 100$)	21	14	22	43	36	23	13	24	44	36

Note: RE is the relative efficiency of a estimator with respect to method "full".

space, we fix the CR.T as 40%, and compare the results with different levels of CR.V1 and CR.V2 in Tables 3.1-3.3.

The following conclusions are inferred from these tables. First, the “CC” method always produces estimates with the smallest biases in most cases and the largest standard deviations among all methods except the baseline method “full”. The standard deviation of “CC” estimator increases rapidly as the CR.V1 or CR.V2 increases. Second, although the SDs obtained from the “fill-d” method are relatively small, the bias tends to be obviously larger especially when CR.V1 or CR.V2 becomes higher. Third, despite the lower RE under higher censoring rate for V , the proposed methods lead to higher efficiency than the complete case analysis. In addition, when the number of censored or uncensored covariates (except intercept) is bigger than one, the advantages of the proposed method become obvious. Particularly, the “MI-logY” outperforms the “MI-noY” in the estimation, implying the benefits of incorporating $\log T$ in the multiple imputation algorithm. Thus in the real data analysis we focus on the methods “MI-logY” and “MI-logY*” only. Last but not least, the shrinkage approach can decrease the biases yielded from “MI-logY” and “MI-noY”, while in the meantime it increases SDs as well in most cases.

3.5 Real Data Analysis: GenIMS Data

We apply the proposed method to analyze the motivating data set from the GenIMS study in Subsection 1.4.2. Community-acquired pneumonia (CAP) is well recognized to be the main cause of severe sepsis and thus has been reported as the largest single cause of mortality among the infectious diseases (Fauci and Morens, 2012). Aspects of the pathophysiology of the

disease, reflected in some of the recently described biomarkers and genomic markers, may contribute to increased understanding of severe CAP (Steel et al., 2013).

This study included a cohort of 1418 patients who were hospitalized with CAP. The survival time of interest is the known days alive at the time of last follow up. Measurements of the biomarkers for inflammatory responses in the body and the basic demographic profile of each patient were recorded and were considered as promising prognostic factors for predicting mortality or overall survival in patients with CAP.

Among the blood samples from each patient, we focus on three cytokine assays: IL-6, TNF and IL-10 obtained on the first day of hospitalization. In the following analysis, we take the logarithmic transformation on the three biomarkers. For simplify, we still call these three transformed biomarkers as *il6*, *tnf*, and *il10*, respectively. The demographic characteristics include sex (labeled 1 for Male and 0 for female), race (1 for other races and 0 for white) and age (ranged from 18 to 102). All these demographic features were exactly observed, while three biomarkers were subject to left censoring as introduced in Subsection 1.4.2.

The censoring rates for the survival time, *il6*, *tnf*, and *il10* are 69.4%, 13.4%, 35.5% and 46.8%, respectively. 511 out of the whole cohort of patients had exactly measured biomarkers, and only 168 individuals had exact observations for both survival time and biomarkers. To make the intercept in CQR meaningful, we centralize age by subtracting the mean age of totally 1418 patients, and further centralize those censored bio-markers by subtracting their corresponding mean values based on the validation set consisting of 511 patients who had the three biomarkers exactly observed.

Based on the findings in our simulation studies, we apply the proposed

"MI-logY" and "MI-logY*" methods and apply them to analyze the data in this study, where the quantile linear imputation model in (3.4) for censored biomarkers on $\log Y$, δ and X is assumed. Figures 3.1 and 3.2 present their corresponding solution paths of regression coefficients over different level τ , in comparison with those obtained by the "CC" method, the complete case analysis based on 511 patients having exactly observed biomarkers in the validation set. Both figures demonstrate that all estimated regression coefficients vary across different quantile levels, implying the suitability of the censored quantile regression used in the analysis, regardless whether the shrinkage approach is used or not. Compared with the results from the complete case analysis, accounting for those censored data in biomarkers by imputation really produces quite different estimates of coefficients, especially for TNF. However, such difference diminishes in Figure 3.2 when employing the shrinkage approach.

To assess the adequacy of the quantile linear imputation model in the estimation procedure, we also carry out quantile nonlinear imputation, named "MIs", for the censored biomarkers on δ , X and 3-degree B-spline for $\log Y$. The resulting solution paths in Figure 3.3 show similar trends as depicted in the Figure 3.1, indicating that the linear or nonlinear imputation model does not make much difference in the estimation results. It is worth mentioning that the problem of quantile crossings appears in the imputation step of the estimation procedure. This problem becomes more severe when the spline technique is used in the nonlinear imputation of the "MIs" method as nonlinear functions are more likely to cross each other. Therefore, the estimated processes across τ in these three figures are obtained after carrying out the quantile rearrangement by R function `rearrange()` in the beginning of Step 3 or 3' of the estimation procedure.

To further examine the effects of all covariates on the quantiles of survival time, the estimated regression coefficients for quantile level $\tau = 0.05, 0.1, 0.2, 0.3, 0.35$ are summarized in Table 3.4, where the standard error (SE) of each estimated coefficient is computed based on 500 bootstrap samples. All three biomarkers tend to have negative impacts on overall survival. These results are actually supported by medical research findings on the properties of pro-inflammatory or anti-inflammatory biomarkers. Specifically, patients with high levels of the proinflammatory cytokine IL-6 and the anti-inflammatory cytokine IL-10 tend to develop severe sepsis and death Mira et al. (2008). Another pro-inflammatory cytokine TNF initiates the inflammatory process and its overproduction may increase the risk of death (Gogos et al., 2000). Biomarker il6 has significant effects at relatively lower quantile levels on overall survival, while biomarker il10 is highly or moderately significant at relatively higher quantile levels. In contrast, biomarker tnf does not exhibit any important effect at almost all quantile levels. Therefore, both cytokines il6 and il10 may be promising prognostic biomarkers for the survival of patients with CAP.

Patients' sex and age are two important demographic features for predicting survival as both of them are significant at most of all quantile levels. Given negative estimates of coefficients which are associated with sex and age at all quantile levels in Table 4, it is shown that male and/or elderly patients tend to have lower survival. In other words, mortality from CAP or sepsis of males is higher than females and it increases significantly with age. This conclusion coincides with the findings in Bernhardt et al. (2014) from the perspective of AFT regression and an earlier medical research by Mayr et al. (2014).

Table 3.4: *Estimation results obtained from the “CC”, “MI-logY”, “MI-logY*” and “MIs” methods for the GenIMS data.*

method	τ		Intercept	il6	tnf	il10	sex	race	age
CC	0.05	estimate	3.803**	-0.148	-0.528	-0.005	-0.722	0.648	-0.11**
		SE	0.437	0.114	0.376	0.235	0.513	0.893	0.018
	0.1	estimate	5.039**	-0.163	-0.357	0.067	-0.698	-0.301	-0.124**
		SE	0.443	0.134	0.471	0.236	0.517	0.896	0.021
	0.2	estimate	6.593**	-0.119	-0.548*	-0.089	-1.06**	-0.213	-0.132**
		SE	0.368	0.16	0.319	0.197	0.425	1.102	0.018
	0.3	estimate	7.895**	0.013	-0.679**	-0.116	-1.16**	-0.461	-0.137**
		SE	0.584	0.158	0.338	0.262	0.51	1.142	0.032
	0.35	estimate	8.190**	-0.03	-0.782**	-0.067	-1.017**	0.504	-0.137**
		SE	0.686	0.156	0.331	0.282	0.499	1.031	0.038
MI-logY	0.05	estimate	3.864**	-0.221**	-0.033	-0.118	-0.614*	-0.125	-0.088**
		SE	0.334	0.107	0.235	0.141	0.362	0.73	0.013
	0.1	estimate	4.924**	-0.229**	-0.021	-0.073	-0.713**	-0.247	-0.095**
		SE	0.28	0.112	0.215	0.127	0.295	0.455	0.012
	0.2	estimate	6.336**	-0.11	-0.033	-0.144*	-0.926**	-0.549	-0.099**
		SE	0.258	0.117	0.219	0.083	0.268	0.433	0.013
	0.3	estimate	7.297**	-0.002	-0.035	-0.191**	-0.787**	-0.137	-0.102**
		SE	0.288	0.121	0.224	0.092	0.264	0.487	0.018
	0.35	estimate	7.666**	0.022	0	-0.152	-0.840**	-0.196	-0.090**
		SE	0.286	0.107	0.229	0.094	0.232	0.458	0.017
MI-logY*	0.05	estimate	3.863**	-0.201**	-0.35	-0.096	-0.620*	0.208	-0.101**
		SE	0.331	0.095	0.281	0.135	0.356	0.674	0.014
	0.1	estimate	4.935**	-0.215**	-0.132	-0.031	-0.713**	-0.248	-0.116**
		SE	0.281	0.105	0.233	0.135	0.295	0.455	0.017
	0.2	estimate	6.438**	-0.11	-0.393	-0.139	-0.942**	-0.514	-0.126**
		SE	0.264	0.117	0.248	0.085	0.266	0.45	0.016
	0.3	estimate	7.649**	-0.002	-0.529*	-0.185*	-0.951**	-0.165	-0.124**
		SE	0.418	0.12	0.277	0.096	0.331	0.491	0.024
	0.35	estimate	7.889**	0.016	-0.665**	-0.144	-0.867**	0.051	-0.121**
		SE	0.395	0.104	0.296	0.099	0.248	0.553	0.029
MIs	0.05	estimate	3.742**	-0.167*	-0.188	-0.005	-0.575	-0.087	-0.088**
		SE	0.356	0.093	0.227	0.154	0.389	0.771	0.015
	0.1	estimate	4.862**	-0.215**	-0.143	-0.032	-0.756**	-0.114	-0.095**
		SE	0.283	0.108	0.202	0.125	0.306	0.469	0.011
	0.2	estimate	6.268**	-0.074	-0.169	-0.147*	-0.952**	-0.56	-0.101**
		SE	0.262	0.107	0.191	0.078	0.268	0.427	0.012
	0.3	estimate	7.098**	0.047	-0.195	-0.192**	-0.747**	-0.12	-0.097**
		SE	0.29	0.107	0.193	0.084	0.264	0.486	0.018
	0.35	estimate	7.474**	0.071	-0.193	-0.149*	-0.825**	-0.102	-0.089**
		SE	0.287	0.096	0.197	0.089	0.229	0.453	0.018

Significance level: * indicates 0.1, and ** indicates 0.05

Figure 3.1: Comparisons of solution paths for estimated regression coefficients (solid curves) along with their 95% upper and lower bounds (dashed curves) using methods “MI-logY” and “CC” for the GenIMS data.

Figure 3.2: Comparisons of solution paths for estimated regression coefficients (solid curves) along with their 95% upper and lower bounds (dashed curves) using methods “MI-logY*” and “CC” for the GenIMS data.

Figure 3.3: Comparisons of solution paths for estimated regression coefficients (solid curves) along with their 95% upper and lower bounds (dashed curves) using methods “MIs” and “CC” for the GenIMS data.

3.6 Summary

In this chapter, we have proposed a censored quantile regression model to analyze survival outcomes with a set of covariates subject to censoring. To estimate the model, we start with the development of a novel multiple imputation approach for quantiles of censored covariates conditional on the survival outcome variable and the rest of uncensored covariates, and then solve a set of martingale-based estimating equations to obtain the estimators of quantile regression coefficients. The proposed method extends the previous work of Wang and Feng (2012) to censored quantile regression. Thus it can improve model flexibility since no distribution assumption is required for the survival outcome or censored covariates, and in the meantime it is more robust to possible outlying observations in survival time variable and/or latent censored covariates.

3.7 Appendix

Let $\bar{F}(t|\mathbf{A}) = \Pr(Y \geq t|\mathbf{A})$, $\tilde{F}(t|\mathbf{A}) = \Pr(Y \leq t, \delta = 1|\mathbf{A})$, $\bar{f}(t|\mathbf{A}) = d\bar{F}(t|\mathbf{A})/dt$, $\tilde{f}(t|\mathbf{A}) = d\tilde{F}(t|\mathbf{A})/dt$, $\boldsymbol{\mu}(\mathbf{b}) = E\{\mathbf{A}N[\exp(\mathbf{A}^T\mathbf{b})]\}$ and $\tilde{\boldsymbol{\mu}}(\mathbf{b}) = E\{\mathbf{A}I[Y \geq \exp(\mathbf{A}^T\mathbf{b})]\}$. Denote

$$\tilde{s}_i(\boldsymbol{\beta}, \tau|\mathbf{A}_i, Y_i, \delta_i) = \mathbf{s}_i(\boldsymbol{\beta}, \tau|\mathbf{A}_i, Y_i, \delta_i)I[\mathbf{W}_i^T\boldsymbol{\eta}(\xi_1) < V_i < d], \quad (3.7)$$

$\bar{\boldsymbol{\eta}}_0 = (\boldsymbol{\eta}_0(u_1)^T, \dots, \boldsymbol{\eta}_0(u_{k_{1n}})^T)^T$ and $\bar{\boldsymbol{\eta}} = (\boldsymbol{\eta}(u_1)^T, \dots, \boldsymbol{\eta}(u_{k_{1n}})^T)^T$. Let

$$\phi_i(\boldsymbol{\lambda}, \boldsymbol{\eta}, u, \tau) = \tilde{s}_i(\boldsymbol{\beta}_0, \tau|(\mathbf{W}_i^T\boldsymbol{\eta}(u), \mathbf{X}_i^T)^T, Y_i, \delta_i)$$

and

$$\Phi_i(\boldsymbol{\lambda}, \bar{\boldsymbol{\eta}}, \tau) = \pi(\mathbf{W}_i; \boldsymbol{\lambda})^{-1} \int_{\xi_1}^{\pi(\mathbf{W}_i; \boldsymbol{\lambda})} \phi_i(\boldsymbol{\lambda}, \boldsymbol{\eta}, u, \tau) du$$

with fixed parameter β_0 and unknown parameters λ and η . The following mild technical conditions is further required.

- Condition 3.6.** a) $\tilde{f}(\exp(\mathbf{A}^T \mathbf{b})|\mathbf{A}) > 0$ for all $\mathbf{b} \in \mathcal{B}(\epsilon)$, where $\mathcal{B}(\epsilon) = \{\mathbf{b} \in \mathbb{R}^p : \inf_{\tau \in (0, \tau_u]} \|\mu(\mathbf{b}) - \mu\{\beta_0(\tau)\}\|_2 < \epsilon\}$.
- b) Each component of $\mu\{\beta_0(\tau)\}$ is Lipschitz for τ .
- c) Each component of $(\mu')^{-1} \tilde{\mu}'$ is uniformly bounded on $\mathbf{b} \in \mathcal{B}(\epsilon)$, where μ' and $\tilde{\mu}'$ are respective first-order derivates of $\mu(\mathbf{b})$ and $\tilde{\mu}(\mathbf{b})$ with respect to $\mathbf{b}$.
- d) $\|\mu'\{\beta_0(\tau)\}\|_2$ is uniformly bounded away from zero in $\tau \in (0, \tau_u]$.
- e) The first-order derivatives of $\bar{F}^{(1)}(t|\mathbf{A}), \bar{F}^{(2)}(t|\mathbf{A}), \tilde{F}^{(1)}(t|\mathbf{A})$ and $\tilde{F}^{(2)}(t|\mathbf{A})$ with respect to t are uniformly bounded in t and A , where $\bar{F}^{(1)}(t|\mathbf{A}) = Pr(Y \geq t|\mathbf{A}, V > d)$, $\bar{F}^{(2)}(t|\mathbf{A}) = Pr(Y \geq t|\mathbf{A}, Q_V(\xi_1|\mathbf{W}) < V \leq d)$, $\tilde{F}^{(1)}(t|\mathbf{A}) = Pr(Y \leq t, \delta = 1|\mathbf{A}, V > d)$, and $\tilde{F}^{(2)}(t|\mathbf{A}) = Pr(Y \leq t, \delta = 1|\mathbf{A}, Q_V(\xi_1|\mathbf{W}) < V \leq d)$.

Condition 3.7. $\left. \frac{\partial E\{\mathbf{s}_i(\beta, \tau|\mathbf{A}_i, Y_i, \delta_i)\}}{\partial \beta} \right|_{\beta(\tau)=\beta_0(\tau)}$ converges almost surely to a positive definite matrix $\Psi\{\beta_0(\tau)\}$ for any $\tau \in (0, \tau_u]$.

Condition 3.8. For any $\tau \in (0, \tau_u]$, $E\{\Phi_i(\lambda, \bar{\eta}, \tau)\}$ is differentiable in the neighborhood of λ_0 and $\bar{\eta}_0$ and has finite partial derivatives at points $(\lambda_0, \bar{\eta}_0)$.

What is more,

$$\begin{aligned} & \frac{1}{n} \sum_{i=n_0+1}^n \left[E\{\Phi_i(\lambda, \bar{\eta}, \tau)\} - E\{\Phi_i(\lambda_0, \bar{\eta}_0, \tau)\} \right] \\ &= \frac{1}{n} \sum_{i=n_0+1}^n \left[\left. \frac{\partial E\{\Phi_i(\lambda, \bar{\eta}, \tau)\}}{\partial \lambda} \right|_{\lambda=\lambda_0, \bar{\eta}=\bar{\eta}_0} (\lambda - \lambda_0) \right. \\ & \quad \left. + \left. \frac{\partial E\{\Phi_i(\lambda, \bar{\eta}, \tau)\}}{\partial \bar{\eta}} \right|_{\lambda=\lambda_0, \bar{\eta}=\bar{\eta}_0} (\bar{\eta} - \bar{\eta}_0) \right] + r_n, \end{aligned}$$

where $|r_n| = O(\|\lambda - \lambda_0\|_2^2 + \|\bar{\eta} - \bar{\eta}_0\|_2^2)$.

The asymptotic properties of the estimator of β can be derived from Conditions 3.6-3.8.

3.7.1 Proof of Theorem 3.1 (Uniform consistency)

Based on the imputation model assumed in (3.4), the latent variable can be written as $V_i = \mathbf{W}_i^T \boldsymbol{\eta}_0(u)$, where u is the corresponding quantile level and follows a uniform distribution over the interval from ξ_1 to its latent censoring probability. Condition 3.1 guarantees the existence of u . The latent censoring probability can be generated via logistic regression as $\pi(\mathbf{W}_i; \boldsymbol{\lambda}_0) = \frac{\exp(\mathbf{W}_i^T \boldsymbol{\lambda}_0)}{1 + \exp(\mathbf{W}_i^T \boldsymbol{\lambda}_0)}$ with latent true parameter $\boldsymbol{\lambda}_0$ and $\mathbf{W}_i$. For those censored covariates, that is $V_i < d$, we can replace them by $V_i^* = \mathbf{W}_i^T \hat{\boldsymbol{\eta}}(v)$. Among them, v is an estimated quantile level and can be generated from a uniform distribution over $(\xi_1, \hat{\pi}(\mathbf{W}_i; \hat{\boldsymbol{\lambda}}))$, in which the censoring probability can be estimated by $\hat{\pi}(\mathbf{W}_i; \hat{\boldsymbol{\lambda}}) = \frac{\exp(\mathbf{W}_i^T \hat{\boldsymbol{\lambda}})}{1 + \exp(\mathbf{W}_i^T \hat{\boldsymbol{\lambda}})}$ based on the estimate $\hat{\boldsymbol{\lambda}}$.

Let $\mathbf{A}_i^* = (V_i^*, \mathbf{X}^T)^T$ be covariate vector with the imputed value. Without loss of generality, for the censored covariate V , we assume that the first n_0 ($n_0 < n$) observations are uncensored, and the rest observations are censored. We take expectation of $\tilde{s}_i(\boldsymbol{\beta}, \tau | \mathbf{A}_i, Y_i, \delta_i)$ in (3.7), defined by $E_{(V, \mathbf{X}, T, C)} [\tilde{s}_i(\boldsymbol{\beta}, \tau | \mathbf{A}_i, Y_i, \delta_i)]$ with respect to the joint density function of $(V, \mathbf{X}, T, C)$ and by $E_{(V^*, \mathbf{X}, T, C)} [\tilde{s}_i(\boldsymbol{\beta}, \tau | \mathbf{A}_i^*, Y_i, \delta_i)]$, with respect to the joint density function of the imputed value V_i^* and $(\mathbf{X}, T, C)$.

The proposed estimation procedure is to seek a generalized solution, denoted by $\hat{\boldsymbol{\beta}}(\tau_r)$ for $r = 1, \dots, k_{2n}$, to the equation $\boldsymbol{\varrho}^*\{\boldsymbol{\beta}(\tau_r)\} = \boldsymbol{\varsigma}_{n,r}$, where

$\max_{r=1, \dots, k_{2n}} \|\mathbf{s}_{n,j}\|_2 \leq \sup_i \|A_i\|_2/n$, $0 < \tau_1 < \dots < \tau_{k_{2n}} = \tau_u$, and

$$\begin{aligned} \boldsymbol{\varrho}^*\{\boldsymbol{\beta}(\tau)\} &= \frac{1}{n} \left\{ \sum_{i=1}^{n_0} \mathbf{s}_i(\boldsymbol{\beta}, \tau | (V_i, \mathbf{X}_i^T)^T, Y_i, \delta_i) + \sum_{i=n_0+1}^n \tilde{\mathbf{s}}_i(\boldsymbol{\beta}, \tau | (V_i^*, \mathbf{X}_i^T)^T, Y_i, \delta_i) \right\} \\ &= \frac{1}{n} \sum_{i=1}^{n_0} \mathbf{A}_i \delta_i I[\log Y_i \leq \mathbf{A}_i^T \boldsymbol{\beta}(\tau)] \\ &\quad + \frac{1}{n} \sum_{i=n_0+1}^n \mathbf{A}_i^* \delta_i I[\log Y_i \leq \mathbf{A}_i^{*T} \boldsymbol{\beta}(\tau)] I[\mathbf{W}_i^T \hat{\boldsymbol{\eta}}(\xi_1) < V_i^* < d] \\ &\quad - \frac{1}{n} \sum_{i=1}^{n_0} \int_0^\tau \mathbf{A}_i I[\log Y_i \geq \mathbf{A}_i^T \boldsymbol{\beta}(u)] dH(u) \\ &\quad - \frac{1}{n} \sum_{i=n_0+1}^n \int_0^\tau \mathbf{A}_i^* I[\log Y_i \geq \mathbf{A}_i^{*T} \boldsymbol{\beta}(u)] I[\mathbf{W}_i^T \hat{\boldsymbol{\eta}}(\xi_1) < V_i^* < d] dH(u). \end{aligned}$$

As n goes to infinity, the limit of $\boldsymbol{\varrho}^*\{\boldsymbol{\beta}(\tau)\}$ is

$$E_{(V, \mathbf{X}, T, C)} [\mathbf{s}_i(\boldsymbol{\beta}, \tau | \mathbf{A}_i, Y_i, \delta_i) I(V_i > d)] + E_{(V^*, \mathbf{X}, T, C)} [\tilde{\mathbf{s}}_i(\boldsymbol{\beta}, \tau | \mathbf{A}_i^*, Y_i, \delta_i)]. \quad (3.8)$$

If all covariates can be observed, or in other words, the censored covariates have been imputed true values, (3.8) actually reduces to

$$\begin{aligned} \mathcal{S}_0(\boldsymbol{\beta}) &= E_{(V, \mathbf{X}, T, C)} [\mathbf{s}_i(\boldsymbol{\beta}, \tau | \mathbf{A}_i, Y_i, \delta_i) I(\mathbf{W}_i^T \boldsymbol{\eta}_0(\xi_1) < V_i)] \\ &:= \boldsymbol{\mu}\{\boldsymbol{\beta}(\tau)\} - \int_0^\tau \tilde{\boldsymbol{\mu}}\{\boldsymbol{\beta}(u)\} dH(u). \end{aligned} \quad (3.9)$$

By definition, we have $\mathcal{S}_0(\boldsymbol{\beta}_0(\tau)) = 0$, which is the same with the estimating equation used in Peng and Huang (2008). We split $\boldsymbol{\mu}(\mathbf{b})$ into two parts with covariate V beyond or under d as

$$\begin{aligned} \boldsymbol{\mu}(\mathbf{b}) &= E[\mathbf{A} \delta I(\log Y \leq \mathbf{A}^T \mathbf{b})] \\ &= E[\mathbf{A} \delta I(\log Y \leq \mathbf{A}^T \mathbf{b}) I(V > d)] \\ &\quad + E[\mathbf{A} \delta I(\log Y \leq \mathbf{A}^T \mathbf{b}) I(\mathbf{W}^T \boldsymbol{\eta}_0(\xi_1) < V \leq d)] \\ &:= \boldsymbol{\mu}_1(\mathbf{b}) + \boldsymbol{\mu}_2(\mathbf{b}). \end{aligned}$$

Similarly,

$$\begin{aligned}
\tilde{\boldsymbol{\mu}}(\mathbf{b}) &= E [\mathbf{A}I (\log Y \geq \mathbf{A}^T \mathbf{b})] \\
&= E [\mathbf{A}I (\log Y \geq \mathbf{A}^T \mathbf{b}) I (V > d)] \\
&\quad + E [\mathbf{A}I (\log Y \geq \mathbf{A}^T \mathbf{b}) I (\mathbf{W}^T \boldsymbol{\eta}_0(\xi_1) < V \leq d)] \\
&:= \tilde{\boldsymbol{\mu}}_1(\mathbf{b}) + \tilde{\boldsymbol{\mu}}_2(\mathbf{b}).
\end{aligned}$$

With V^* , let

$$\boldsymbol{\mu}_2^*(\mathbf{b}) = E [\mathbf{A}^* \delta I (\log Y \leq \mathbf{A}^{*T} \mathbf{b}) I (\mathbf{W}^T \hat{\boldsymbol{\eta}}(\xi_1) < V^* \leq d)]$$

and

$$\tilde{\boldsymbol{\mu}}_2^*(\mathbf{b}) = E [\mathbf{A}^* I (\log Y \geq \mathbf{A}^{*T} \mathbf{b}) I (\mathbf{W}^T \hat{\boldsymbol{\eta}}(\xi_1) < V^* \leq d)].$$

We need to pay attention to the difference between $\boldsymbol{\mu}_2^*(\mathbf{b})$ with imputed values and $\boldsymbol{\mu}_2(\mathbf{b})$,

$$\begin{aligned}
&\sup_{\mathbf{b}} \|\boldsymbol{\mu}_2^*(\mathbf{b}) - \boldsymbol{\mu}_2(\mathbf{b})\|_2 \\
&\leq \int_{\{(v,t,c,x): F_V^{-1}(\xi_1 | \mathbf{w}) < v \leq d\}} \sup_{\mathbf{b}} \left\| \mathbf{A} \delta I[t \leq \exp(\mathbf{A}^T \mathbf{b})] \right\|_2 \\
&\quad \times \left| \hat{f}_{(V^*, T, C, X)}(v, t, c, x) - f_{(V, T, C, X)}(v, t, c, x) \right| dv dt dc dx + o(1),
\end{aligned}$$

where $f_{(V, T, C, X)}$ represents the true density functions of (V, T, C, X) , and $\hat{f}_{(V^*, T, C, X)}$ denotes its estimated density function. Similarly,

$$\begin{aligned}
&\sup_{\mathbf{b}} \|\tilde{\boldsymbol{\mu}}_2^*(\mathbf{b}) - \tilde{\boldsymbol{\mu}}_2(\mathbf{b})\|_2 \\
&\leq \int_{\{(v,t,c,x): F_V^{-1}(\xi_1 | \mathbf{w}) < v \leq d\}} \sup_{\mathbf{b}} \left\| \mathbf{A} \delta I[t \geq \exp(\mathbf{A}^T \mathbf{b})] I[c \geq \exp(\mathbf{A}^T \mathbf{b})] \right\|_2 \\
&\quad \times \left| \hat{f}_{(V^*, T, C, X)}(v, t, c, x) - f_{(V, T, C, X)}(v, t, c, x) \right| dv dt dc dx + o(1).
\end{aligned}$$

Under Conditions 3.1, 3.2c) and 3.3-3.5, as the same arguments in proofs of Lemma S.1 and Theorem 2(i) in Wang and Feng (2012), we can easily get

$$\sup_{v \leq \min\{d, \mathbf{W}^T \hat{\eta}(\hat{\pi}(\mathbf{W}))\}} \left| 1 - \frac{\hat{f}(v|\mathbf{W}, v < \mathbf{W}^T \hat{\eta}(\hat{\pi}(\mathbf{W})))}{f(v|\mathbf{W}, v < d)} \right| = o_p(1)$$

when $k_{1n} \rightarrow \infty$, $n^{1/2}k_{1n}^{-1} \rightarrow 0$. Thereafter, it follows from Conditions 3.2a) and dominated convergence theorem that

$$\sup_b \|\boldsymbol{\mu}_2^*(\mathbf{b}) - \boldsymbol{\mu}_2(\mathbf{b})\|_2 = o_p(1) \quad (3.10)$$

and

$$\sup_b \|\tilde{\boldsymbol{\mu}}_2^*(\mathbf{b}) - \tilde{\boldsymbol{\mu}}_2(\mathbf{b})\|_2 = o_p(1). \quad (3.11)$$

To ease the computation complexity, we will use the following notations:

$$\mathbf{v}_n^{(1)}(\mathbf{b}) = \frac{1}{n} \sum_{i=1}^{n_0} \mathbf{A}_i \delta_i I[\log Y_i \leq \mathbf{A}_i^T \mathbf{b}] - \boldsymbol{\mu}_1(\mathbf{b}),$$

$$\tilde{\mathbf{v}}_n^{(1)}(\mathbf{b}) = \frac{1}{n} \sum_{i=1}^{n_0} \mathbf{A}_i I[\log Y_i \geq \mathbf{A}_i^T \mathbf{b}] - \tilde{\boldsymbol{\mu}}_1(\mathbf{b}),$$

$$\mathbf{v}_n^{*(2)}(\mathbf{b}) = \frac{1}{n} \sum_{i=n_0+1}^n \mathbf{A}_i^* \delta_i I[\log Y_i \leq \mathbf{A}_i^{*T} \mathbf{b}] I[\mathbf{W}_i^T \hat{\boldsymbol{\eta}}(\xi_1) < V_i^* < d] - \boldsymbol{\mu}_2(\mathbf{b}),$$

and

$$\tilde{\mathbf{v}}_n^{*(2)}(\mathbf{b}) = \frac{1}{n} \sum_{i=n_0+1}^n \mathbf{A}_i^* I[\log Y_i \geq \mathbf{A}_i^{*T} \mathbf{b}] I[\mathbf{W}_i^T \hat{\boldsymbol{\eta}}(\xi_1) < V_i^* < d] - \tilde{\boldsymbol{\mu}}_2(\mathbf{b}).$$

Looking into each term of $\boldsymbol{\varrho}^*(\boldsymbol{\beta})$ and the limits in (3.9) for each τ_r , $r = 1, \dots, k_{2n}$, we have

$$\begin{aligned} & \frac{1}{n} \sum_{i=1}^{n_0} \mathbf{A}_i N_i \left[\exp(\mathbf{A}_i^T \hat{\boldsymbol{\beta}}(\tau_r)) \right] - E \left\{ \mathbf{A} N \left[\exp(\mathbf{A}^T \boldsymbol{\beta}_0(\tau_r)) \right] I(V > d) \right\} \\ &= \mathbf{v}_n^{(1)}\{\hat{\boldsymbol{\beta}}(\tau_r)\} + \boldsymbol{\mu}_1\{\hat{\boldsymbol{\beta}}(\tau_r)\} - \boldsymbol{\mu}_1\{\boldsymbol{\beta}_0(\tau_r)\}, \end{aligned}$$

$$\begin{aligned}
& \frac{1}{n} \sum_{i=1}^{n_0} \int_0^{\tau_r} \mathbf{A}_i I \left[Y_i \geq \exp(\mathbf{A}_i^T \hat{\boldsymbol{\beta}}(u)) \right] dH(u) \\
& - E \left\{ \int_0^{\tau_r} \mathbf{A} I \left[Y \geq \exp(\mathbf{A}^T \boldsymbol{\beta}_0(u)) \right] I(V > d) dH(u) \right\} \\
& = \int_0^{\tau_r} \tilde{\mathbf{v}}_n^{(1)} \{ \hat{\boldsymbol{\beta}}(u) \} dH(u) + \int_0^{\tau_r} \left[\tilde{\boldsymbol{\mu}}_1 \{ \hat{\boldsymbol{\beta}}(u) \} - \tilde{\boldsymbol{\mu}}_1 \{ \boldsymbol{\beta}_0(u) \} \right] dH(u),
\end{aligned}$$

$$\begin{aligned}
& \frac{1}{n} \sum_{i=n_0+1}^n \mathbf{A}_i^* N_i \left[\exp(\mathbf{A}_i^{*T} \hat{\boldsymbol{\beta}}(\tau_r)) \right] I \left[\mathbf{W}_i^T \hat{\boldsymbol{\eta}}(\xi_1) < V_i^* < d \right] \\
& - E \left\{ \mathbf{A} N \left[\exp(\mathbf{A}^T \boldsymbol{\beta}_0(\tau_r)) \right] I \left(\mathbf{W}^T \boldsymbol{\eta}_0(\xi_1) < V \leq d \right) \right\} \\
& = \mathbf{v}_n^{*(2)} \{ \hat{\boldsymbol{\beta}}(\tau_r) \} + \boldsymbol{\mu}_2 \{ \hat{\boldsymbol{\beta}}(\tau_r) \} - \boldsymbol{\mu}_2 \{ \boldsymbol{\beta}_0(\tau_r) \},
\end{aligned}$$

$$\begin{aligned}
& \frac{1}{n} \sum_{i=n_0+1}^n \int_0^{\tau_r} \mathbf{A}_i^* I \left[Y_i \geq \exp(\mathbf{A}_i^{*T} \hat{\boldsymbol{\beta}}(u)) \right] I \left[\mathbf{W}_i^T \hat{\boldsymbol{\eta}}(\xi_1) < V_i^* < d \right] dH(u) \\
& - E \left\{ \int_0^{\tau_r} \mathbf{A} I \left[Y \geq \exp(\mathbf{A}^T \boldsymbol{\beta}_0(u)) \right] I \left(\mathbf{W}^T \boldsymbol{\eta}_0(\xi_1) < V \leq d \right) dH(u) \right\} \\
& = \int_0^{\tau_r} \tilde{\mathbf{v}}_n^{*(2)} \{ \hat{\boldsymbol{\beta}}(u) \} dH(u) + \int_0^{\tau_r} \left[\tilde{\boldsymbol{\mu}}_2 \{ \hat{\boldsymbol{\beta}}(u) \} - \tilde{\boldsymbol{\mu}}_2 \{ \boldsymbol{\beta}_0(u) \} \right] dH(u).
\end{aligned}$$

By the above four equations and the fact that $\mathcal{S}_0(\boldsymbol{\beta}_0(\tau)) = 0$, we obtain

$$\begin{aligned}
& \boldsymbol{\mu} \{ \hat{\boldsymbol{\beta}}(\tau_r) \} - \boldsymbol{\mu} \{ \boldsymbol{\beta}_0(\tau_r) \} \\
& = \mathbf{s}_{n,j} - \mathbf{v}_n^{(1)} \{ \hat{\boldsymbol{\beta}}(\tau_r) \} - \mathbf{v}_n^{*(2)} \{ \hat{\boldsymbol{\beta}}(\tau_r) \} \\
& + \int_0^{\tau_r} \left[\tilde{\mathbf{v}}_n^{(1)} \{ \hat{\boldsymbol{\beta}}(u) \} + \tilde{\mathbf{v}}_n^{*(2)} \{ \hat{\boldsymbol{\beta}}(u) \} \right] dH(u) \\
& + \int_0^{\tau_r} \left[\tilde{\boldsymbol{\mu}} \{ \hat{\boldsymbol{\beta}}(u) \} - \tilde{\boldsymbol{\mu}} \{ \boldsymbol{\beta}_0(u) \} \right] dH(u).
\end{aligned} \tag{3.12}$$

According to the Glivenko-Cantelli Theorem (Van Der Vaart and Wellner, 1996), $\sup_{\mathbf{b}} \|\mathbf{v}_n^{(1)}(\mathbf{b})\|_2 = o(1)$ and $\sup_{\mathbf{b}} \|\tilde{\mathbf{v}}_n^{(1)}(\mathbf{b})\|_2 = o(1)$. $\mathbf{v}_n^{*(2)}(\mathbf{b})$ can

be rewritten as

$$\begin{aligned} & \sup_{\mathbf{b}} \|\mathbf{v}_n^{*(2)}(\mathbf{b})\|_2 \\ & \leq \sup_{\mathbf{b}} \left\| \frac{1}{n} \sum_{i=n_0+1}^n \mathbf{A}_i^* N_i [\exp(\mathbf{A}_i^{*T} \mathbf{b})] I [\mathbf{W}_i^T \hat{\boldsymbol{\eta}}(\xi_1) < V_i^* < d] - \boldsymbol{\mu}_2^*(\mathbf{b}) \right\|_2 \\ & + \sup_{\mathbf{b}} \left\| \boldsymbol{\mu}_2^*(\mathbf{b}) - \boldsymbol{\mu}_2(\mathbf{b}) \right\|_2, \end{aligned}$$

where in the right hand side the first term is close to zero by the Glivenko-Cantelli Theorem, and the second term is also close to zero due to equation (3.10). Thereby, $\sup_{\mathbf{b}} \|\mathbf{v}_n^{*(2)}(\mathbf{b})\|_2 = o_p(1)$. Similarly, it follows from (3.11) that $\sup_{\mathbf{b}} \|\tilde{\mathbf{v}}_n^{*(2)}(\mathbf{b})\|_2 = o_p(1)$.

Denote $\mathcal{D}(\epsilon) = \{\boldsymbol{\mu}(\mathbf{b}) : \mathbf{b} \in \mathcal{B}(\epsilon)\}$. In line with the proof of Theorem 1 in Peng and Huang (2008), by Conditions 3.2b) and 3.6a), we obtain the result that $\boldsymbol{\mu}$ is a one-to-one map from the neighbourhood $\mathcal{B}(\epsilon)$ of $\boldsymbol{\beta}_0(\tau)$ to $\mathcal{D}(\epsilon)$. For any $\mathbf{b}$ belonging to $\mathcal{B}(\epsilon)$, there exists a function $\boldsymbol{\kappa}$ such that $\boldsymbol{\kappa}\{\boldsymbol{\mu}(\mathbf{b})\} = \mathbf{b}$. Thus,

$$\begin{aligned} & \|\tilde{\boldsymbol{\mu}}\{\hat{\boldsymbol{\beta}}(u)\} - \tilde{\boldsymbol{\mu}}\{\boldsymbol{\beta}_0(u)\}\|_2 = \|\tilde{\boldsymbol{\mu}}\{\boldsymbol{\kappa}\{\boldsymbol{\mu}(\hat{\boldsymbol{\beta}}(u))\}\} - \tilde{\boldsymbol{\mu}}\{\boldsymbol{\kappa}\{\boldsymbol{\mu}(\boldsymbol{\beta}_0(u))\}\}\|_2 \\ & = \|(\boldsymbol{\mu}'(\check{b}))^{-1} \tilde{\boldsymbol{\mu}}(\check{b})[\boldsymbol{\mu}(\hat{\boldsymbol{\beta}}(u)) - \boldsymbol{\mu}(\boldsymbol{\beta}_0(u))]\|_2, \end{aligned} \tag{3.13}$$

where $\check{b}$ is between $\boldsymbol{\mu}\{\hat{\boldsymbol{\beta}}(u)\}$ and $\boldsymbol{\mu}\{\boldsymbol{\beta}_0(u)\}$. Thereby, under Conditions 3.2b) and 3.6b)-c), applying techniques similar to those in Peng and Huang (2008) gives that the norm of the last term in equation (3.12) and $\sup_{\tau \in (0, \tau_u]} \|\boldsymbol{\mu}\{\hat{\boldsymbol{\beta}}(\tau)\} - \boldsymbol{\mu}\{\boldsymbol{\beta}_0(\tau)\}\|_2$ are both close to zero when n and k_{2n} go to infinity. It then follows from Condition 3.6d) and the Taylor expansions that $\sup_{\tau \in (0, \tau_u]} \|\hat{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}_0(\tau)\|_2 = o_p(1)$. Thus, $\sup_{\tau \in (0, \tau_u]} \|\hat{\boldsymbol{\beta}}^{(j)}(\tau) - \boldsymbol{\beta}_0(\tau)\|_2 = o_p(1)$, where $\hat{\boldsymbol{\beta}}^{(j)}(\tau)$ is the estimate based on the j -th imputed value set for $j = 1, \dots, m$. Therefore, $\sup_{\tau \in (0, \tau_u]} \|\hat{\boldsymbol{\beta}}_{MI}(\tau) - \boldsymbol{\beta}_0(\tau)\|_2 = o_p(1)$. Thus, the proof is complete.

3.7.2 Proof of Theorem 3.2 (Asymptotic normality)

Define

$$\begin{aligned}\mathbf{M}_0(\boldsymbol{\beta}, \tau) &= \frac{1}{n} \sum_{i=1}^{n_0} \mathbf{A}_i N_i [\exp(\mathbf{A}_i^T \boldsymbol{\beta}(\tau))] - \int_0^\tau \mathbf{A}_i I [Y_i \geq \exp(\mathbf{A}_i^T \boldsymbol{\beta}(u))] dH(u), \\ \mathbf{M}_1^{(j)}(\boldsymbol{\beta}, \tau) &= \frac{1}{n} \sum_{i=n_0+1}^n \mathbf{A}_i^{*(j)} N_i [\exp(\mathbf{A}_i^{*(j)T} \boldsymbol{\beta}(\tau))] I [\mathbf{W}_i^T \hat{\boldsymbol{\eta}}(\xi_1) < \mathbf{W}_i^T \hat{\boldsymbol{\eta}}(v_j) < d] \\ &\quad - \int_0^\tau \mathbf{A}_i^{*(j)} I [Y_i \geq \exp(\mathbf{A}_i^{*(j)T} \boldsymbol{\beta}(u))] I [\mathbf{W}_i^T \hat{\boldsymbol{\eta}}(\xi_1) < \mathbf{W}_i^T \hat{\boldsymbol{\eta}}(v_j) < d] dH(u),\end{aligned}$$

for $j = 1, \dots, m$. Akin to Lemma B.1 in Peng and Huang (2008), based on Conditions 3.2 and 3.6e), and Theorem 3.1, we have

$$\begin{aligned}\sup_{\tau \in (0, \tau_u]} \left\| \frac{1}{n} \sum_{i=1}^{n_0} \mathbf{A}_i N_i [\exp(\mathbf{A}_i^T \hat{\boldsymbol{\beta}}(\tau))] - \mathbf{A}_i N_i [\exp(\mathbf{A}_i^T \boldsymbol{\beta}_0(\tau))] \right. \\ \left. - [\boldsymbol{\mu}_1\{\hat{\boldsymbol{\beta}}(\tau)\} - \boldsymbol{\mu}_1\{\boldsymbol{\beta}_0(\tau)\}] \right\|_2 = o_p(n^{-\frac{1}{2}})\end{aligned}$$

and

$$\begin{aligned}\sup_{\tau \in (0, \tau_u]} \left\| \frac{1}{n} \sum_{i=1}^{n_0} \mathbf{A}_i I [Y_i \geq \exp(\mathbf{A}_i^T \hat{\boldsymbol{\beta}}(\tau))] - \mathbf{A}_i I [Y_i \geq \exp(\mathbf{A}_i^T \boldsymbol{\beta}_0(\tau))] \right. \\ \left. - [\tilde{\boldsymbol{\mu}}_1\{\hat{\boldsymbol{\beta}}(\tau)\} - \tilde{\boldsymbol{\mu}}_1\{\boldsymbol{\beta}_0(\tau)\}] \right\|_2 = o_p(n^{-\frac{1}{2}}),\end{aligned}$$

which imply

$$\sup_{\tau \in (0, \tau_u]} \left\| \mathbf{M}_0(\hat{\boldsymbol{\beta}}, \tau) - \mathbf{M}_0(\boldsymbol{\beta}_0, \tau) - E\{\mathbf{M}_0(\hat{\boldsymbol{\beta}}, \tau)\} + E\{\mathbf{M}_0(\boldsymbol{\beta}_0, \tau)\} \right\|_2 = o_p(n^{-\frac{1}{2}})$$

and

$$\begin{aligned}\mathbf{M}_0\{\hat{\boldsymbol{\beta}}^{(j)}\} &= \mathbf{M}_0(\bar{\boldsymbol{\beta}}_0) + o_p(n^{-\frac{1}{2}}) \\ &\quad + \left[\frac{\partial E_{(V, X, T, C)}\{\mathbf{M}_0(\bar{\boldsymbol{\beta}})\}}{\partial \bar{\boldsymbol{\beta}}} \Big|_{\bar{\boldsymbol{\beta}}=\bar{\boldsymbol{\beta}}_0} + o_p(1) \right] (\hat{\boldsymbol{\beta}}^{(j)} - \bar{\boldsymbol{\beta}}_0),\end{aligned}$$

where $\hat{\beta}^{(j)}$ is the estimator of $\bar{\beta} = (\beta(\tau_1)^T, \dots, \beta(\tau_{k_{2n}})^T)^T$ based on the j -th ($j = 1, \dots, m$) imputed value set with $0 < \tau_1 < \dots < \tau_{k_{2n}} = \tau_u < 1$. Parallel analysis can be found in the proof of Theorem 2(ii) of Wang and Feng (2012). Similarly,

$$\begin{aligned} \mathbf{M}_1^{(j)}\{\hat{\beta}^{(j)}\} &= \mathbf{M}_1^{(j)}(\bar{\beta}_0) + o_p(n^{-\frac{1}{2}}) \\ &+ \left[\frac{\partial E_{(V^*, X, T, C)}\{\mathbf{M}_1^{(j)}(\bar{\beta})\}}{\partial \bar{\beta}} \bigg|_{\bar{\beta}=\bar{\beta}_0} + o_p(1) \right] (\hat{\beta}^{(j)} - \bar{\beta}_0). \end{aligned}$$

From Conditions 3.1-3.5 and 3.7 and when $k_{1n} \rightarrow \infty$ and $n^{1/2}k_{1n}^{-1} \rightarrow 0$, we have

$$\frac{\partial E_{(V, X, T, C)}\{\mathbf{M}_0(\bar{\beta})\}}{\partial \bar{\beta}} \bigg|_{\bar{\beta}=\bar{\beta}_0} + m^{-1} \sum_{j=1}^m \frac{\partial E_{(V^*, X, T, C)}\{\mathbf{M}_1^{(j)}(\bar{\beta})\}}{\partial \bar{\beta}} \bigg|_{\bar{\beta}=\bar{\beta}_0} \rightarrow \Psi\{\bar{\beta}_0\}, \quad a.s.$$

Thus,

$$\hat{\beta}_{MI}(\tau) - \beta_0(\tau) = -\Psi^{-1}\{\beta_0(\tau)\} \left[\mathbf{M}_0(\beta_0, \tau) + m^{-1} \sum_{j=1}^m \mathbf{M}_1^{(j)}(\beta_0, \tau) \right] + o_p(n^{-\frac{1}{2}}).$$

Note that for any $\tau \in (0, \tau_u]$ and when $k_{1n} \rightarrow \infty$,

$$\begin{aligned} &\sup_{\|\lambda - \lambda_0\| \rightarrow 0, \|\eta - \eta_0\| \rightarrow 0} \left\| \sum_{i=n_0+1}^n \phi_i(\lambda, \eta, u, \tau) - \phi_i(\lambda_0, \eta_0, u, \tau) \right. \\ &\quad \left. - E\{\Phi_i(\lambda, \bar{\eta}, \tau)\} + E\{\Phi_i(\lambda_0, \bar{\eta}_0, \tau)\} \right\|_2 = o_p(n^{\frac{1}{2}}), \end{aligned}$$

which can be derived from Lemmas 3.2 and 3.3 in He and Shao (2000).

Define

$$\begin{aligned} \mathbf{U}_{2n}(V_i^{*(j)}, \mathbf{W}_i, \delta_i^v, \tau; \lambda_0, \eta_0) &= \left(n^{-1} \sum_{i=n_0+1}^n \frac{\partial E\{\Phi_i(\lambda, \bar{\eta}, \tau)\}}{\partial \lambda} \bigg|_{\lambda=\lambda_0, \bar{\eta}=\bar{\eta}_0} \right)^T D(V_i^{*(j)}, \mathbf{W}_i, \delta_i^v) \\ &+ \left(n^{-1} \sum_{i=n_0+1}^n \frac{\partial E\{\Phi_i(\lambda, \bar{\eta}, \tau)\}}{\partial \bar{\eta}} \bigg|_{\lambda=\lambda_0, \bar{\eta}=\bar{\eta}_0} \right)^T \bar{e}(V_i^{*(j)}, \mathbf{W}_i, \delta_i^v). \end{aligned}$$

for censored observations of V and $j = 1, \dots, m$, and

$$\begin{aligned} \mathbf{U}_{1n}(V_i, \mathbf{W}_i, \delta_i^v, \tau; \boldsymbol{\lambda}_0, \boldsymbol{\eta}_0) &= \left(n^{-1} \sum_{i=n_0+1}^n \frac{\partial E\{\Phi_i(\boldsymbol{\lambda}, \bar{\boldsymbol{\eta}}, \tau)\}}{\partial \boldsymbol{\lambda}} \bigg|_{\boldsymbol{\lambda}=\boldsymbol{\lambda}_0, \bar{\boldsymbol{\eta}}=\bar{\boldsymbol{\eta}}_0} \right)^T D(V_i, \mathbf{W}_i, \delta_i^v) \\ &\quad + \left(n^{-1} \sum_{i=n_0+1}^n \frac{\partial E\{\Phi_i(\boldsymbol{\lambda}, \bar{\boldsymbol{\eta}}, \tau)\}}{\partial \bar{\boldsymbol{\eta}}} \bigg|_{\boldsymbol{\lambda}=\boldsymbol{\lambda}_0, \bar{\boldsymbol{\eta}}=\bar{\boldsymbol{\eta}}_0} \right)^T \bar{e}(V_i, \mathbf{W}_i, \delta_i^v). \end{aligned}$$

for uncensored observations of V , where

$$\bar{e}(V_i, \mathbf{W}_i, \delta_i^v) = (e(V_i, \mathbf{W}_i, \delta_i^v, u_1), \dots, e(V_i, \mathbf{W}_i, \delta_i^v, u_{k_{1n}}))^T.$$

It follows from Conditions 3.3, 3.5 and 3.8 that

$$\begin{aligned} n\mathbf{M}_1^{(j)}(\boldsymbol{\beta}_0, \tau) &= \sum_{i=n_0+1}^n \phi_i(\hat{\boldsymbol{\lambda}}, \hat{\boldsymbol{\eta}}, u_{i(j)}, \tau) - \phi_i(\boldsymbol{\lambda}_0, \boldsymbol{\eta}_0, u_{i(j)}, \tau) + \phi_i(\boldsymbol{\lambda}_0, \boldsymbol{\eta}_0, u_{i(j)}, \tau) \\ &= o_p(n^{\frac{1}{2}}) + \sum_{i=n_0+1}^n E\{\Phi_i(\boldsymbol{\lambda}, \bar{\boldsymbol{\eta}}, \tau)\} - E\{\Phi_i(\boldsymbol{\lambda}_0, \bar{\boldsymbol{\eta}}, \tau)\} + \phi_i(\boldsymbol{\lambda}_0, \boldsymbol{\eta}_0, u_{i(j)}, \tau) \\ &= o_p(n^{\frac{1}{2}}) + \sum_{i=n_0+1}^n \left[\frac{\partial E\{\Phi_i(\boldsymbol{\lambda}, \bar{\boldsymbol{\eta}}, \tau)\}}{\partial \boldsymbol{\lambda}} \bigg|_{\boldsymbol{\lambda}=\boldsymbol{\lambda}_0, \bar{\boldsymbol{\eta}}=\bar{\boldsymbol{\eta}}_0} (\boldsymbol{\lambda} - \boldsymbol{\lambda}_0) \right. \\ &\quad \left. + \frac{\partial E\{\Phi_i(\boldsymbol{\lambda}, \bar{\boldsymbol{\eta}}, \tau)\}}{\partial \bar{\boldsymbol{\eta}}} \bigg|_{\boldsymbol{\lambda}=\boldsymbol{\lambda}_0, \bar{\boldsymbol{\eta}}=\bar{\boldsymbol{\eta}}_0} (\bar{\boldsymbol{\eta}} - \bar{\boldsymbol{\eta}}_0) \right] + \phi_i(\boldsymbol{\lambda}_0, \boldsymbol{\eta}_0, u_{i(j)}, \tau) \\ &= o_p(n^{\frac{1}{2}}) + \sum_{i=1}^{n_0} \mathbf{U}_{1n}(V_i, \mathbf{W}_i, \delta_i^v, \tau; \boldsymbol{\lambda}_0, \boldsymbol{\eta}_0) \\ &\quad + \sum_{i=n_0+1}^n \left[\mathbf{U}_{2n}(V_i^{*(j)}, \mathbf{W}_i, \delta_i^v, \tau; \boldsymbol{\lambda}_0, \boldsymbol{\eta}_0) + \phi_i(\boldsymbol{\lambda}_0, \boldsymbol{\eta}_0, u_{i(j)}, \tau) \right], \end{aligned}$$

where $u_{i(j)}$ follows a uniform distribution over $(0, \pi(\mathbf{W}_i; \boldsymbol{\lambda}_0))$ for the j -th imputed quantile level and individual $i = 1, \dots, n$. Thus the asymptotic

representation can be rewritten as

$$\begin{aligned} & n^{\frac{1}{2}}(\hat{\beta}_{MI}(\tau) - \beta_0(\tau)) \\ &= -n^{-\frac{1}{2}}\Psi^{-1}\{\beta_0(\tau)\} \left[n\mathbf{M}_0(\beta_0, \tau) + \sum_{i=1}^{n_0} \mathbf{U}_{1n}(V_i, \mathbf{W}_i, \delta_i^v, \tau; \boldsymbol{\lambda}_0, \boldsymbol{\eta}_0) \right. \\ & \quad \left. + m^{-1} \sum_{j=1}^m \sum_{i=n_0+1}^n \left[\mathbf{U}_{2n}(V_i^{*(j)}, \mathbf{W}_i, \delta_i^v, \tau; \boldsymbol{\lambda}_0, \boldsymbol{\eta}_0) + \phi_i(\boldsymbol{\lambda}_0, \boldsymbol{\eta}_0, u_{i(j)}, \tau) \right] \right] + o_p(1). \end{aligned}$$

Similar to $\mathbf{S}_n(\beta_0, \tau)$ in Peng and Huang (2008),

$$-n^{-\frac{1}{2}} \left[n\mathbf{M}_0(\beta_0, \tau) + m^{-1} \sum_{j=1}^m \sum_{i=n_0+1}^n \phi_i(\boldsymbol{\lambda}_0, \boldsymbol{\eta}_0, u_{i(j)}, \tau) \right]$$

also converges weakly to a tight Gaussian process. From Conditions 3.3 and 3.5 and by the boundedness for partial derivatives of $E\{\Phi_i(\boldsymbol{\lambda}, \bar{\boldsymbol{\eta}}, \tau)\}$ in Condition 3.8, we obtain that

$$-n^{-\frac{1}{2}} \left[\sum_{i=1}^{n_0} \mathbf{U}_{1n}(V_i, \mathbf{W}_i, \delta_i^v, \tau; \boldsymbol{\lambda}_0, \boldsymbol{\eta}_0) + m^{-1} \sum_{j=1}^m \sum_{i=n_0+1}^n \mathbf{U}_{2n}(V_i^{*(j)}, \mathbf{W}_i, \delta_i^v, \tau; \boldsymbol{\lambda}_0, \boldsymbol{\eta}_0) \right]$$

converges weakly to a tight Gaussian process as well. Therefore, it follows from the positive definiteness of Ψ that $n^{\frac{1}{2}}(\hat{\beta}_{MI}(\tau) - \beta_0(\tau))$ converges weakly to a Gaussian process. The proof of Theorem 3.2 is complete.

Chapter 4

Discussion and Future Research

This dissertation develops two-stage estimation procedures for survival data with covariates subject to complex censoring. Two topics are taken into account.

The estimation method in Chapter 2 allows the time-dependent covariate (non-terminal event time) to be censored by the response variable (terminal event time) and a working AFT model is considered for the censored covariate. The estimation procedure in Chapter 3 can tackle the CQR with multiple covariates subject to limits of detection. We utilize a working CQR for censored covariate and a working logistic regression for estimating conditional probabilities in the intermediate steps of the proposed estimation procedure.

More specifically, Chapter 2 focuses on the complicated semi-competing risks data structure, which consists of the non-terminal and terminal event times. In medical research, clinicians or researchers would like to predict the survival probability of the terminal event using the information of the intermediate event time together with other personal characteristics, thus provide more effective treatment. Using the conditional model in Chapter 2,

we can measure dynamically how large the non-terminal event time affects the terminal event time. The proposed method would perform well when the non-terminal event time carries information which is not explained by the other potential covariates. It is worth noting that we make the assumption of the conditional model regardless of the lower or upper wedge, even though the semi-competing risks data is only observed in the upper wedge $0 < V \leq T$.

Since the coefficient estimation in the AFT model is subject to the assumption $\mathbf{W}_0(\gamma_0) = 0$, we acknowledge that the proposed method may face potential bias when the assumption is violated. The estimation of the proposed conditional model still performs well if the bias of the first-stage estimator in the AFT model is within a certain range. When the rate of dependent censoring is larger than 50%, in which V is mainly censored by T rather than C , a possible alternative method is to apply the artificial censoring technique in Ding et al. (2009) to obtain estimates of γ .

The future research for this part has some directions. First, clustered data are frequently encountered in some clinical and medical investigations. Our work on conditional analysis of clustered semi-competing risks data is now ongoing. Besides, the proposed method can be extended to semi-competing risks data with multiple non-terminal events. The nonlinear fit of the non-terminal event time in the conditional model and the model checking merit further investigations.

The topic in Chapter 3 becomes more important when the censored covariates have significant effects on the survival time. The proposed approach is more efficient and robust than existing methods. For the GenIMS data, the proposed method shows a complete picture of the different effects of censored bio-markers over different quantile levels.

Based on our current work, its flexibility and robustness may facilitate future directions for analyzing more complex data. For example, topic on longitudinal survival outcomes with censored covariates is challenging but merits further investigation. Besides, it is worth noting that we use the martingale-based estimating equation in the final step of the proposed estimation procedure for the purpose of simplifying calculations and theoretical proofs. Other works for the censored quantile regression such as Portnoy (2003), Wang and Wang (2009) and Leng et al. (2013) can also be applied for such topic. To avoid the global linearity assumption in the CQR, the work of Wang and Wang (2009) can be considered, but leading to difficulties in numerical computations and statistical inferences for the CQR with censored covariates. Furthermore, A flexible imputation model is considered in the proposed method. This idea can be applied to other areas which incorporate censored or incomplete covariates.

References

- Andersen, P. K. and Gill, R. D. (1982). Cox’s regression model for counting processes: a large sample study. *The annals of statistics*, pages 1100–1120.
- Bernhardt, P. W., Wang, H. J., and Zhang, D. (2014). Flexible modeling of survival data with covariates subject to detection limits via multiple imputation. *Computational statistics & data analysis*, 69:81–91.
- Bottai, M. and Zhang, J. (2010). Laplace regression with censored data. *Biometrical Journal*, 52(4):487–503.
- Brown, B. and Wang, Y.-G. (2007). Induced smoothing for rank regression with censored survival times. *Statistics in medicine*, 26(4):828–836.
- Cardoso, F., Costa, A., Norton, L., Cameron, D., Cufer, T., Fallowfield, L., Francis, P., Gligorov, J., Kyriakides, S., Lin, N., Pagani, O., Senkus, E., Thomssen, C., Aapro, M., Bergh, J., Leo, A. D., Saghir, N. E., Ganz, P., Gelmon, K., Goldhirsch, A., Harbeck, N., Houssami, N., Hudis, C., Kaufman, B., Leadbeater, M., Mayer, M., Rodger, A., Rugo, H., Sacchini, V., Sledge, G., vant Veer, L., Viale, G., Krop, I., and Winer, E. (2012).

- 1st international consensus guidelines for advanced breast cancer (abc 1). *The Breast*, 21(3):242 – 252.
- Chen, Y.-H. (2012). Maximum likelihood analysis of semicompeting risks data with semiparametric regression models. *Lifetime data analysis*, 18(1):36–57.
- Chen, Y.-H., Chatterjee, N., and Carroll, R. J. (2009). Shrinkage estimators for robust and efficient inference in haplotype-based case-control studies. *Journal of the American Statistical Association*, 104(485):220–233.
- Chernozhukov, V., Fernández-Val, I., and Galichon, A. (2010). Quantile and probability curves without crossing. *Econometrica*, 78(3):1093–1125.
- Chiou, S. H., Kang, S., and Yan, J. (2014a). Fast accelerated failure time modeling for case-cohort data. *Statistics and Computing*, 24(4):559–568.
- Chiou, S. H., Kang, S., and Yan, J. (2014b). Fitting accelerated failure time models in routine survival analysis with r package aftgee. *Journal of Statistical Software*, 61(11):1–23.
- Chung, M., Long, Q., and Johnson, B. A. (2013). A tutorial on rank-based coefficient estimation for censored data in small-and large-scale problems. *Statistics and computing*, 23(5):601–614.
- Cox, D. R. (1972). Models and life-tables regression. *JR Stat. Soc. Ser. B*, 34:187–220.
- D’Angelo, G. and Weissfeld, L. (2008). An index approach for the cox model with left censored covariates. *Statistics in medicine*, 27(22):4502–4514.

- Ding, A. A., Shi, G., Wang, W., and Hsieh, J.-J. (2009). Marginal regression analysis for semi-competing risks data under dependent censoring. *Scandinavian Journal of Statistics*, 36(3):481–500.
- Ding, Y. (2010). *Some new insights about the accelerated failure time model*. PhD thesis, University of Michigan.
- Ding, Y. and Nan, B. (2015). Estimating mean survival time: when is it possible? *Scandinavian Journal of Statistics*, 42(2):397–413.
- Fauci, A. S. and Morens, D. M. (2012). The perpetual challenge of infectious diseases. *New England Journal of Medicine*, 366(5):454–461.
- Félix, J., Aragão, F., Almeida, J. M., Calado, F. J., Ferreira, D., Pereira, A. B., Rodrigues, R., and Rijo, J. F. (2013). Time-dependent endpoints as predictors of overall survival in multiple myeloma. *BMC cancer*, 13(1):122.
- Fine, J. P., Jiang, H., and Chappell, R. (2001). On semi-competing risks data. *Biometrika*, 88(4):907–919.
- Fu, L., Wang, Y.-G., and Bai, Z. (2010). Rank regression for analysis of clustered data: a natural induced smoothing approach. *Computational Statistics & Data Analysis*, 54(4):1036–1050.
- Gamst, A., Donohue, M., and Xu, R. (2009). Asymptotic properties and empirical evaluation of the npml in the proportional hazards mixed-effects model. *Statistica Sinica*, pages 997–1011.
- Gogos, C. A., Drosou, E., Bassaris, H. P., and Skoutelis, A. (2000). Pro-versus anti-inflammatory cytokine profile in patients with severe sepsis:

- a marker for prognosis and future therapeutic options. *The Journal of infectious diseases*, 181(1):176–180.
- Haibe-Kains, B., Desmedt, C., Loi, S., Culhane, A. C., Bontempi, G., Quackenbush, J., and Sotiriou, C. (2012). A three-gene model to robustly identify breast cancer molecular subtypes. *Journal of the National Cancer Institute*, 104(4):311–325.
- Halabi, S., Vogelzang, N. J., Ou, S.-S., Owzar, K., Archer, L., and Small, E. J. (2009). Progression-free survival as a predictor of overall survival in men with castrate-resistant prostate cancer. *Journal of clinical oncology*, 27(17):2766–2771.
- He, X. and Shao, Q.-M. (2000). On parameters of increasing dimensions. *Journal of Multivariate Analysis*, 73(1):120–135.
- Heng, D., Xie, W., Bjarnason, G., et al. (2011). Progression-free survival as a predictor of overall survival in metastatic renal cell carcinoma treated with contemporary targeted therapy. *Cancer*, 117:2637–2642.
- Hu, C. and Tsodikov, A. (2014). Joint modeling approach for semicompeting risks data with missing nonterminal event status. *Lifetime data analysis*, 20(4):563–583.
- Jin, Z., Lin, D., Wei, L., and Ying, Z. (2003). Rank-based inference for the accelerated failure time model. *Biometrika*, 90(2):341–353.
- Johnson, L. M. and Strawderman, R. L. (2009). Induced smoothing for the semiparametric accelerated failure time model: asymptotics and extensions to clustered data. *Biometrika*, 96(3):577–590.

- Kellum, J. A., Kong, L., Fink, M. P., Weissfeld, L. A., Yealy, D. M., Pinsky, M. R., Fine, J., Krichevsky, A., Delude, R. L., and Angus, D. C. (2007). Understanding the inflammatory cytokine response in pneumonia and sepsis: results of the genetic and inflammatory markers of sepsis (genims) study. *Archives of internal medicine*, 167(15):1655–1663.
- Kim, Y. and Kong, L. (2015). Estimation of c-index for cox proportional hazards model with censored biomarker covariate subject to limits of detection. *Journal of biopharmaceutical statistics*, 25(3):459–473.
- Koenker, R. (2005). *Quantile regression*. Number 38. Cambridge university press.
- Kong, S. and Nan, B. (2016). Semiparametric approach to regression with a covariate subject to a detection limit. *Biometrika*, 103(1):161–174.
- Leng, C., Tong, X., et al. (2013). A quantile regression estimator for censored data. *Bernoulli*, 19(1):344–361.
- Li, R. and Cheng, Y. (2016). Flexible association modelling and prediction with semi-competing risks data. *Canadian Journal of Statistics*, 44(3):361–374.
- Lin, D., Robins, J., and Wei, L. (1996). Comparing two failure time distributions in the presence of dependent censoring. *Biometrika*, 83(2):381–393.
- Lin, D. and Wei, L. (1991). Goodness-of-fit tests for the general cox regression model. *Statistica Sinica*, pages 1–17.
- Lin, D. Y., Wei, L.-J., and Ying, Z. (1993). Checking the cox model with cumulative sums of martingale-based residuals. *Biometrika*, 80(3):557–572.

- Lobbezoo, D. J. A., van Kampen, R. J. W., A C Voogd, M. W. D., van den Berkmortel, F., Smild, T. J., van de Wouw, A. J., Peters, F. P. J., van Riel, J. M. G. H., Peters, N. A. J. B., de Boer, M., Peer, P. G. M., and Tjan-Heijnen, V. C. G. (2015). Prognosis of metastatic breast cancer: are there differences between patients with de novo and recurrent metastatic breast cancer? *British Journal of Cancer*, 112:1445 – 1451.
- Mayr, F., Yende, S., and Angus, D. (2014). Epidemiology of severe sepsis. *Virulence*, 5(1):4–11.
- McCulloch, C. E. and Searle, S. R. (2005). *Generalized, linear, and mixed models*. John Wiley & Sons.
- Mira, J.-P., Max, A., and Burgel, P.-R. (2008). The role of biomarkers in community-acquired pneumonia: predicting mortality and response to adjunctive therapy. *Critical Care*, 12(Suppl 6):S5.
- Nan, B., Kalbfleisch, J. D., and Yu, M. (2009). Asymptotic theory for the semiparametric accelerated failure time model with missing data. *The Annals of Statistics*, pages 2351–2376.
- Nan, B. and Wellner, J. A. (2013). A general semiparametric z-estimation approach for case-cohort studies. *Statistica Sinica*, 23(3):1155.
- Parast, L., Cheng, S.-C., and Cai, T. (2011). Incorporating short-term outcome information to predict long-term survival with discrete markers. *Biometrical Journal*, 53(2):294–307.
- Peng, L. and Fine, J. P. (2006). Rank estimation of accelerated lifetime models with dependent censoring. *Journal of the American Statistical Association*, 101(475):1085–1093.

- Peng, L. and Fine, J. P. (2007). Regression modeling of semicompeting risks data. *Biometrics*, 63(1):96–108.
- Peng, L. and Huang, Y. (2008). Survival analysis with quantile regression models. *Journal of the American Statistical Association*, 103(482):637–649.
- Portnoy, S. (2003). Censored regression quantiles. *Journal of the American Statistical Association*, 98(464):1001–1012.
- Powell, J. L. (1986). Censored regression quantiles. *Journal of econometrics*, 32(1):143–155.
- Saad, E. and Katz, A. (2009). Progression-free survival and time to progression as primary end points in advanced breast cancer: often used, sometimes loosely defined. *Annals of Oncology*, 20:460–464.
- Steel, H. C., Cockeran, R., Anderson, R., and Feldman, C. (2013). Overview of community-acquired pneumonia and the role of inflammatory mechanisms in the immunopathogenesis of severe pneumococcal disease. *Mediators of Inflammation*, 2013(Article ID 490346):doi.org/10.1155/2013/490346.
- Ter-Minassian, M., Zhang, S., Brooks, N., et al. (2017). Association between tumor progression endpoints and overall survival in patients with advanced neuroendocrine tumors. *The Oncologist*, 22(2):165–172.
- Van der Vaart, A. W. (2000). *Asymptotic statistics (Vol.3)*. Cambridge University Press.
- Van Der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes*. Springer.

- Wang, H. J. and Feng, X. (2012). Multiple imputation for m-regression with censored covariates. *Journal of the American Statistical Association*, 107(497):194–204.
- Wang, H. J. and Wang, L. (2009). Locally weighted censored quantile regression. *Journal of the American Statistical Association*, 104(487):1117–1128.
- Wang, Y.-G. and Fu, L. (2011). Rank regression for accelerated failure time model with clustered and censored data. *Computational Statistics & Data Analysis*, 55(7):2334–2343.
- Wei, Y., Ma, Y., and Carroll, R. J. (2012). Multiple imputation in quantile regression. *Biometrika*, 99(2):423–438.
- Xu, J., Kalbfleisch, J. D., and Tai, B. (2010). Statistical analysis of illness–death processes and semicompeting risks data. *Biometrics*, 66(3):716–725.
- Yang, J. and Peng, L. (2016). A new flexible dependence measure for semicompeting risks. *Biometrics*, 72(3):770–779.
- Ying, Z. (1993). A large sample study of rank estimation for censored regression data. *The Annals of Statistics*, pages 76–99.
- Ying, Z., Jung, S.-H., and Wei, L.-J. (1995). Survival analysis with median regression models. *Journal of the American Statistical Association*, 90(429):178–184.
- Zeng, D., Lin, D., and Yin, G. (2005). Maximum likelihood estimation for the proportional odds model with random effects. *Journal of the American Statistical Association*, 100(470):470–483.