**Distribution Agreement**

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

Signature:

| | |
|---|---|
| Your Name | Date |

# New Statistical Methods for Complex Survival Analysis Problems

By

Bo Wei

Doctor of Philosophy

Biostatistics

---

Limin Peng, Ph.D.

---

Amita Manatunga, Ph.D.

---

Yijian Huang, Ph.D.

---

Donald Lee, Ph.D.

---

David H. Howard, Ph.D.

Accepted:

---

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

---

Date

# New Statistical Methods for Complex Survival Analysis Problems

By

Bo Wei
Sc.M., Brown University, 2016
B.S., Nanjing University, 2013

Advisor: Limin Peng, Ph.D.

An abstract of
A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in Biostatistics
2021

## Abstract

### New Statistical Methods for Complex Survival Analysis Problems

By

Bo Wei

In biomedical studies, the analysis of time-to-event data may encounter various complex problems. One such scenario is that the observation of recurrent events can be terminated by a dependent event. Another example is that treatment choice is not random, possibly outcome-dependent, and therefore standard approaches comparing treated group versus untreated group generally do not lead to valid estimates for the causal treatment effect of interest. In this dissertation, we develop new statistical methods to handle these complications in survival analysis.

In the first project, we propose two sensible adaptations of the generalized accelerated recurrence time (GART) model (Sun et al., 2016) to handle the recurrent events terminated by a dependent event. Our modeling strategies align with the rationale underlying the use of the survivors' rate function or the adjusted rate function to account for the presence of the dependent terminal event. We establish the asymptotic properties of the new estimators. Simulation studies demonstrate good finite-sample performance of the proposed methods. An application to a dataset from the Cystic Fibrosis Foundation Patient Registry (CFFPR) illustrates the practical utility of the new methods.

In the second project, we propose a new IV framework with randomly censored outcomes where the causal treatment effect is quantified as complier quantile causal effect (CQCE). Employing the special characteristic of IV and adapting the principle of conditional score, we uncover a simple weighting scheme that can be incorporated into the standard censored quantile regression procedure to estimate CQCE. We develop robust nonparametric estimation of the derived weights in the first stage, which permits stable implementation of the second stage estimation based on existing software. We establish rigorous asymptotic properties for the proposed estimator, and confirm its validity and satisfactory finite-sample performance via extensive simulations. The proposed method is applied to a dataset from the Center for International Blood and Marrow Transplant Research (CIBMTR) to evaluate the causal effect of rituximab in diffuse large B-cell lymphoma (DLBCL) patients.

In the third project, we study the IV estimation of the population quantile causal effect (PQCE) with the randomly censored data. Employing the rank similarity assumption, we propose an estimating equation based on the observed quantities. We develop a simple and easily-implemented two-step estimation procedure to solve the non-monotonous estimating equation, and propose a sample-based inference approach to avoid computation burden in resampling-based approaches. We rigorously justify the asymptotic properties for the proposed estimator. Extensive simulations and an application to a dataset from CIBMTR demonstrates the practical utility of the proposed method.

# New Statistical Methods for Complex Survival Analysis Problems

By

Bo Wei
ScM, Brown University, 2015
BS, Nanjing University, 2013

Advisor: Limin Peng, Ph.D.

A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University in
partial fulfillment of the requirements for the degree of Doctor of
Philosophy
in Biostatistics
2021

# Acknowledgement

First and foremost, I am extremely grateful to my supervisor, Dr. Limin Peng for her invaluable supervision, continuous support, and patience during my PhD study. Her insightful feedback pushes me to sharpen my thinking and bring my work to a higher level. Her immense knowledge, enormous enthusiasm, rigorousness in research and meticulousness make her my role model that will affect my future career and life. Without her invaluable help and guidance, this dissertation would never be possible.

I also want to express my sincere gratitude to Dr. Amita Manatunga for her invaluable suggestions and excellent encouragement during my Ph.D journey. I really want to thank her for offering her valuable time in improving my presentation and interview skill.

Besides my advisor and Dr. Amita Manatunga, I would like to express my deepest appreciation to my other committee members: Dr. David H. Howard, Dr. Yijian Huang, and Dr. Donald Lee for generously offering their time and guidance throughout the preparation and review of this dissertation. Their suggestions have significantly improved the work and my presentation skill.

I would also like to thank Dr. Jason P. Fine, Dr. Ying Guo, Dr. Huichuan J. Lai, Dr. Meijie Zhang, Dr. Zhuming Zhang and other faculty members that I have collaborated with. The experiences are enjoyable and I have acquired plenty knowledge by learning from them.

I thank to my friends, Dr. Liangyuan Hu, Dr. Fan Li, and Jin Ming, for their help and encouragement. I am so glad to have wonderful friends in my life. My sincere thanks also go to my Emory colleagues and friends, for accompanying me throughout this wonderful Ph.D journey.

Last but not least, I am grateful for my parents whose constant love and support

keep me motivated and confident. My accomplishments and success are because they believed in me. I also owe my deepest gratitude to Ying Cui, who is my love. I am forever thankful for the unconditional love and support throughout the pandemic and every day. I am so grateful to have a chance to share a life with you.

# Contents

# List of Figures

# List of Tables

# Chapter 1

# Introduction

## 1.1 Background

In biomedical studies, the analysis of time-to-event data may encounter various complex problems. One such scenario is that the observation of recurrent events can be terminated by a dependent event. Another example is that treatment choice is not random, possibly outcome-dependent, and therefore standard approaches comparing treated group versus untreated group generally do not lead to valid estimates for the causal treatment effect of interest. In this dissertation, we develop new statistical methods to handle these complications in survival analysis.

In the first project, we consider the survival settings with recurrent events, such as hospitalizations and infections, which are commonly encountered in longitudinal follow-up studies of chronic disease. In practice, the observation of the recurrent events may be stopped by some disease-related events, such as death. One motivating example is the Cystic Fibrosis Foundation Patient Registry (CFFPR) study. Cystic Fibrosis (CF) is one of the most common, life-shortening genetic disorders in the United States (Russell et al., 2012). Pseudomonas aeruginosa (PA) is one of the major pathogens in CF lungs, which is associated with poor clinical outcomes and greater mortality (Davies, 2002). Respiratory tract cultures are routinely obtained for identifying PA and characterizing its phenotypes (mucoid or non-mucoid). The early PA infection is usually non-mucoid and antibiotic sensitive. But recurrent of non-mucoid PA infection leads to chronic PA infection, then to mucoid PA phenotype (Mathee et al., 1999). The development of mucoid PA yet can be more complicated than this widely held paradigm (Heltshe et al., 2018). Mucoid PA is more resistant to antibiotics and more difficult to eradicate (Lyczak et al., 2002). As a result, rarely patients can go back to the non-mucoid PA infection stage once acquiring a mucoid PA infection. Under these considerations, a mucoid PA infection constitutes a dependent terminal event to the recurrent process of non-mucoid PA infections (in addition to death). The treatment strategy for the PA infections is to conduct eradication

therapy at the stage of non-mucoid PA infections. Therefore, there is a scientific interest to characterize association between the timing of non-mucoid PA infection recurrences before a mucoid PA infection and the potential risk factors. To address the complication from the dependent terminal event (i.e. mucoid PA infection), we have developed two sensible adaptations of the generalized accelerated recurrence time (GART) model (Sun et al., 2016) to provide useful alternative analyses that can offer physical interpretations while rendering extra flexibility beyond the existing work based on the accelerated failure time model.

In the second and third projects, we consider to quantify two types of causal effects, the complier causal effect and population causal effect, under the survival scenario where treatment choice decision may be outcome-dependent. Such a scenario is commonly present in observational studies. Even in the perfectly randomized experiments, the non-compliance of treatment assignments may lead to the dependency between the treatment decision and outcomes. A valid estimate of the causal treatment effects of interest may not be directly obtained by comparing the treated and untreated groups. Instrumental variable (IV) methods are important approaches to identify the causal effects in causal inference. An instrumental variable is a variable that is associated with treatment and independent of unmeasured confounding given covariates. The effect of the instrumental variable on the outcome is only through its effect on the treatment given covariates (Baiocchi et al., 2014). A valid IV could introduce the variation in the treatment but has no effect on the unmeasured confounders. This quasi-randomization from the IV can be used to identify the causal treatment effects. IV methods have been extensively studied for settings with standard uncensored outcomes. One major type of approaches concerns population (or global) causal effects of treatment defined by contrasting potential outcomes under different treatments based on the whole population. Another type of approaches studies complier (or local) causal treatment effects under the latent class framework of IV

analyses, where the IV and the treatment are both binary. That is, the whole population is divided into the latent compliance subgroups: compliers whose treatment choices always coincide with the IV, always takers who always take the treatment, never takers who always decline the treatment, and defiers whose treatment choices are always opposite to the IV. Each type of the causal effect has its own advantages and disadvantages. For example, the population causal effect has advantages in interpretation (Heckman and Urzua, 2010; Deaton, 2009), and can be applied to different types of IVs and treatment variables. In another perspective, estimating compiler treatment effect relies on fewer untestable assumptions than estimating the population (or global) causal treatment effect. Thus, compared to global causal treatment effects, complier causal treatment effects may permit more robust inference and is often of substantive interest in real applications.

Our motivating example in the second and third projects is from an analysis of a retrospective dataset from the Center for International Blood and Marrow Transplant Research (CIBMTR) was aimed to evaluate the efficacy of pre-transplant rituximab treatment in diffuse large B-cell lymphoma (DLBCL) patients for improving progression free survival, as compared to standard intervention without rituximab. It was found that patients in the rituximab group tended to be older and have had more chemotherapy regimens than patients in the standard intervention group. The selection of pre-transplant treatment may also relate to factors, such as molecular subtype of lymphoma, which potentially influence the post-transplant outcomes but were not captured in this dataset. Ignoring these unmeasured confounders, a direct comparison of the two treatment groups, even after adjusting for measured confounders, may still fail to provide a valid estimate for the causal effect of rituximab in DLBCL patients. To assess the causal treatment effect of rituximab, we address these complications by utilizing a binary IV, which, in the CIBMTR example, is the indicator of whether treatment was received after the FDA approval date of the rituximab. In the sec-

ond project, we propose a new IV framework with randomly censored outcomes to study complier (or local) causal treatment effects under the latent class framework of IV analyses, where the IV and the treatment are both binary. In the third project, we propose a censored population quantile causal effect (CPQCE) model under IV framework to quantity population causal effect with randomly censored outcomes and unmeasured confounders by employing the rank similarity assumption.

In the next section, we will present literature reviews on regression methods for recurrent events data in the presence of a dependent terminal events and IV methods for time-to-event data. At the end of this chapter, we will give an outline of this dissertation.

## 1.2 Literature Review

### 1.2.1 Existing regression methods for recurrent events data subject to a dependent terminal event

The presence of a dependent terminal event for recurrent events is common in biomedical research. 'Net' quantities that correspond to the setting without the dependent terminal event are often of interest. Net quantities, such as the marginal rate and mean function, however, are not nonparametrically identifiable in the presence of a dependent terminal event (Ghosh and Lin, 2003). Tackling such quantities generally requires additional assumptions about the association between the terminal event and the recurrent events. Various types of joint models of the recurrent events and the dependent terminal event have been studied in literature (Ghosh and Lin, 2003; Liu et al., 2004; Huang and Wang, 2004; Ye et al., 2007, among others). Ghosh and Lin (2003) considered the accelerated failure time models for both the recurrent events and the terminal event. Huang and Wang (2004) adopted the proportional hazards model for the terminal event and the proportional rate model for the recurrent events,

given a latent variable capturing the within-subject connection between the recurrent events and the terminal event. Liu et al. (2004) studied the proportional hazards models for both the terminal event and the recurrent event, and further incorporated a shared frailty to capture the dependency between the recurrent events and the terminal event.

Alternatively, one may be interested in 'crude' quantities. These quantities account for both the recurrent events and the terminal event, rather than targeting the recurrent event process for the hypothetical setting where the terminal event does not exist. Examples of such quantities include the adjusted rate function (Luo et al., 2010), which depicts the rate of recurrent events before the occurrence of the terminal event, and the survivors' rate function, which represents the rate of recurrent events conditioning on the terminal event hasn't occurred (Cook and Lawless, 1997). The interpretations of such 'crude; quantities do not assume the existence of the latent recurrent event process after the occurrence of the terminal event (which may be controversial in some practical situations). In particular, the interpretations of the survivors' rate function and the adjusted rate function bear a similar flavor to those of cause-specific hazard function and cumulative incidence function, which are popularly used for competing risks data analyses. In addition, these quantities can be estimated without imposing additional modeling of the terminal event. Many authors (Ghosh and Lin, 2002; Schaubel and Cai, 2005; Liu et al., 2004; Zeng and Cai, 2010, among others) have investigated either nonparametric or semiparametric estimation of this type of quantities. For example, Schaubel and Cai (2005) and Liu et al. (2004) studied proportional survivors' rate model and an additive survivors' rate model was proposed by Zeng and Cai (2010). Ghosh and Lin (2002) studied a proportional adjusted rate model and proposed counting-process based estimation procedures, which handle random censoring by either inverse weighting the probability of censoring or inverse weighting the survival probabilities of the terminal event time.

In the first project, we study how to address the presence of a dependent terminal event under the generalized accelerated recurrence time (GART) model (Sun et al., 2016). The GART model is a generalization of the traditional AFT modeling and offers additional flexibility to explore potential heterogeneous effects of covariates. Section 2.2 provides a brief review of the GART model. There is no existing work for dealing with a dependent terminal event under the GART model. Our work under the first topic is to fill this gap and render the above advantages of GART model.

### 1.2.2 Existing work on IV methods in time-to-event data

Instrumental variables (IV) is a common approach to handle the unobserved confounders in the estimation of the causal effect. Formally, an instrumental variable is a variable that is correlated with treatment and associated the outcome through the treatment and uncontaminated by unmeasured confounders (Baiocchi et al., 2014). A valid IV can introduce quasi-randomization characteristic to eliminate bias caused by unmeasured confounders. IV methods, initially motivated by applications in economics, have been extensively studied for the standard uncensored outcomes. There are two major types of approaches of IV methods in existing literature. One major type of approach concerns complier (or local) causal treatment effects under the latent class framework of IV analyses with binary IV and treatment variable (Imbens and Angrist, 1994; Angrist et al., 1996a; Imbens and Rubin, 1997; Abadie et al., 2002; Abadie, 2003; Cheng, Small, Tan and Ten Have, 2009, among others). Another type of IV methods is designed for evaluating the causal effect over the entire population, and can be applied to binary, discrete, and continuous treatment variables and IVs (Heckman and Robb Jr, 1985; Angrist and Imbens, 1995; Vansteelandt and Goetghebeur, 2003; Chernozhukov and Hansen, 2005, 2006, 2008, among others).

#### 1.2.2.1 Existing work on IV methods for estimating complier causal effect in time-to-event data

The setting with a binary IV and a binary treatment is frequently encountered in time-to-event data. A common example is a randomized two-arm clinical trial with treatment non-adherence issue, where a binary IV corresponds to the random group assignment, which clearly influences the actual treatment but may not be always the same as the actual treatment. The setting with binary IV and treatment can also arise in an observational study, for example, to assess a new intervention versus standard care, with a binary IV indicating a change in treatment guideline or availability of the new intervention.

Among existing methods that concern a survival outcome with binary treatment and IV, the local causal treatment effect has been commonly formulated based on the hazard of the event of interest. For example, Baker (1998) and Nie et al. (2011) studied the causal hazard difference in compliers using likelihood-based inference. Following the popular proportional hazard modeling in survival analysis, many authors (Loeys and Goetghebeur, 2003; Vansteelandt and Goetghebeur, 2003; Cuzick et al., 2007; Li and Gray, 2016; Kianian et al., 2019, among others) formulated the treatment effect as a complier causal proportional hazard ratio (CCPHR) and developed estimation procedures based on the partial or full likelihood technique or weighted estimating equations. However, the CCPHR, like the traditional hazard ratio, does not offer a straightforward physical interpretation (Reid, 1994), and is confined by the proportional hazard constraint to only confer a static view of treatment effect.

A local causal treatment effect defined on the time scale of a survival outcome may be preferred by some practitioners given its physical interpretation, but was only studied sparsely in literature. Lin et al. (2014) and Yu et al. (2015) studied the estimation of the complier location shift causal effect of treatment on a transformed time scale with randomly censored data. Their likelihood-based approaches require

a transformation model (Zeng and Lin, 2007) is assumed for each latent compliance subgroup, despite the interest lies only in the complier subgroup. A quantile treatment effect is conceptually more flexible than a location shift effect for describing how the treatment affects the potential outcome distributions (Koenker, 2005). To the best of our knowledge, a quantile causal treatment effect was considered only in the "global" setting under structural quantile regression models with control variables (Blundell and Powell, 2004; Chernozhukov et al., 2015). The available two-stage estimation procedures were also limited to handle survival data with censoring that is always known or observed.

In the second project, we made the first effort to tackle the problem of assessing complier (or *local*) quantile causal effect (CQCE) of treatment with time-to-event outcomes subject to standard random censoring. It is worth emphasizing that compared to causal average casual effect (CACE), CQCE, the causal estimand proposed in the second project, is more suitable for survival settings. This is because the presence of censoring often precludes the identifiability of average event time and hence CACE in the first place. By employing the concept of quantiles, CQCE can provide a more comprehensive picture about the causal treatment effect on the time-scale than CACE, and is less restrictive than a complier location shift effect.

#### 1.2.2.2 Existing work on IV methods for estimating population causal effect in time-to-event data

Another type of IV methods is designed for evaluating the causal effect over the entire population. Compared to the estimands of local causal effects which only work for binary IV, the estimands of population causal effect can be applied to general settings of IV, such as continuous IV. Besides, the estimands of population causal effect may have more desirable interpretations than the estimands of local causal effect (Heckman and Urzua, 2010; Deaton, 2009).

Recently, many methods have been proposed for estimating the population causal effects for time-to-event data. For example, Li et al. (2015) and Tchetgen et al. (2015) proposed methods to estimate a linear structural additive hazards model for right censored data, where the causal treatment effect corresponded to the causal difference in the hazard function. Their approaches adapted either the standard two-stage least squares (2SLS) estimation or "control function approach" for structural linear models (Wooldridge, 2010). Zheng et al. (2017) extended Li et al. (2015)'s approach to estimate the causal treatment effect on the subdistribution hazard of a competing risks outcome. Recently, many methods is considered to estimate the causal hazard ratio through the popular Cox proportional hazard model. Martínez-Camblor et al. (2019) proposed a two-stage residual inclusion procedure with an individual frailty in Cox proportional hazard model. Sørensen et al. (2019) proposed a structural Cox proportional hazard model, which includes a baseline model and a selection bias function to separately describe the effect of covariates and selection mechanism on the survival outcomes in the scenario without exposure. However, the hazard function does not have a straightforward physical interpretation (Reid, 1994), and has limitations in causal inference because of its built-in selection bias (Hernán, 2010). To avoid these issues, Huling et al. (2019) incorporated IVs in the semiparametric accelerated failure time (AFT) model, and developed a rank-based estimator of the causal effect of the treatment on the log transformed event time. All of these methods assume the time-constant causal effects, which does not hold in the heterogeneous causal effect scenario.

PQCE, which is a quantile counterpart of the PACE, can capture heterogeneous causal effects on different points of the potential outcome distribution. Like the classical quantile treatment effect, PQCE on the time scale of a survival outcome may be preferred because of its advantages in physical interpretation and flexibility in describing heterogeneous effects among the distributions of the potential outcome

(Koenker, 2005). Blundell and Powell (2007) proposed a two-stage estimator with additive models for first stage and control variables in second stage. Chernozhukov et al. (2015) proposed a two stage model, and allowed a non-additive model for the control variables. Both methods impose strong structural assumptions on the control variables, and are only applied to survival data with all observed censoring time. Hong and Tamer (2003) proposed a conditional moment inequality formulation to quantile causal treatment effect with all observed censoring time. Khan and Tamer (2009) extended Hong and Tamer (2003)'s method to allow for standard covariate dependent right censoring. However, their estimation procedures are quite complicated which involve a third order U-process. Chen (2018) proposed a sequential instrumental variable censored quantile regression procedure to estimate PCQE for censored data. In this procedure, a subsample at each quantile, for which the censoring does not affect the conditional quantile function, is constructed based on the estimators at previous quantiles. The estimator can be obtained by applying IVQR model (Chernozhukov and Hansen, 2005, 2006) in the constructed subsample. However, this approach is limited to time-to-event data with fixed censoring time. Harding and Lamarche (2012) studied the quantile regression panel duration models with endogenous covariates, which can be applied for right censored data. It is remarkable that their model is similar to our proposed model in this paper. However, our approach use different estimating equation and estimation procedure. Moreover, we can not find rigorously theoretical justification and clear estimation procedure in their work.

In the third work, we formally develop a censored population quantile causal effect model (CPQCE) model to assess PQCE on a time-to-event outcome subject to standard conditionally independent right censoring. In the CPQCE model, we propose a sequentially two-stage estimation procedure to solve the non-monotone estimating equation, and establish asymptotic theory that fully justifies the proposed estimation procedure and the sample-based inference procedure. Similar to the comparison be-

tween CACE and CQCE, PQCE, the causal estimand in the third project, is a more suitable causal estimand than population average causal effect (PACE) in the context of time-to-event data.

## 1.3 Outline

In Chapter 2, we propose two sensible adaptations of the generalized accelerated recurrence time (GART) model (Sun et al., 2016) to handle the recurrent events terminated by a dependent event. The modeling strategies align with the rationale underlying the use of the survivors' rate function or the adjusted rate function to account for the presence of the dependent terminal event. We identify and develop estimation and inference procedures, and establish the asymptotic properties of the new estimator. Simulation studies demonstrate good finite-sample performance of the proposed methods. An application to a dataset from the Cystic Fibrosis Foundation Patient Registry (CFFPR) illustrated the practical utility of the new methods.

In Chapter 3, we propose a new IV framework with randomly censored outcomes where the causal treatment effect is quantified as complier quantile causal effect (CQCE). Compared to the commonly studied complier average causal effect (CACE), CQCE has better identifiability when outcomes are subject to censoring, and can provide more dynamic insight about the potential outcome difference under different treatments. Employing the special characteristic of IV and adapting the principle of conditional score, we uncover a simple weighting scheme that can be incorporated into the standard censored quantile regression procedure to estimate CQCE. We develop robust nonparametric estimation of the derived weights in the first stage, which permits stable implementation of the second stage estimation based on existing software. We establish rigorous asymptotic properties for the proposed estimator, and confirm its validity and satisfactory finite-sample performance via ex-

tensive simulations. The proposed method is applied to a dataset from the Center for International Blood and Marrow Transplant Research (CIBMTR) to evaluate the causal effect of rituximab in diffuse large B-cell lymphoma (DLBCL) patients.

In Chapter 4, we propose a censored population quantile causal effect (CPQCE) model under instrumental variable (IV) framework to quantity the population quantile causal effect (PQCE) for the randomly censored data with unmeasured confounders. As the comparison between CQCE and CACE, PQCE has advantages in identifiability and capturing heterogeneous effects compared to population average causal effect (PACE). Employing the rank similarity assumption, an estimating equation based on the observed quantities has been provided. We develop a simple and easily-implemented two-step estimation procedure to solve the non-monotonous estimating equation, and propose a sample-based inference approach to avoid computation burden in resampling-based approaches. We rigorously justify the asymptotic properties for the proposed estimator. Extensive simulations have been conducted to confirm its validity and satisfactory finite-sample performance. An application to a dataset from the Center for International Blood and Marrow Transplant Research (CIBMTR) demonstrates the practical utility of the proposed method.

In Chapter 5, we provide a summary of our dissertation work and briefly discuss our plan of future work.

# Chapter 2

# Generalized Accelerated Recurrence Time Model in the Presence of a Dependent Terminal Event

## 2.1 Notation and Data Scenario

We first introduce notation and describe the data scenario of interest. Let $T^{(j)}$ denote the time to the $j$th recurrent event ($j = 1, 2, \cdots$), $D$ denote the time to a dependent terminal event, and $(L, R]$ denote a random observation window for the recurrent events. Define $\tilde{R} = R \wedge D$, $\delta = I(D \leq R)$, and $\boldsymbol{Z} = (1, \tilde{\boldsymbol{Z}}^\top)^\top$, where $\tilde{\boldsymbol{Z}}$ denotes a $p \times 1$ vector of covariates, $\wedge$ is the minimum operator, and $I(\cdot)$ is the indicator function. The observed counting process of recurrent events is defined as $N(t) = \sum_{j=1}^{\infty} I(L < T^{(j)} \leq t \wedge \tilde{R})$. The underlying recurrent event counting process without accounting for the presence of the terminal event is given by $N_0^*(t) = \sum_{j=1}^{\infty} I(T^{(j)} \leq t)$. The underlying recurrent event counting process that accounts for the presence of the dependent terminal event is given by $N^*(t) = \sum_{j=1}^{\infty} I(T^{(j)} \leq t \wedge D)$. Clearly, $N^*(t)$ does not jump for $t > D$, meaning it does not involve the information on the recurrent events that occur after the time $D$. We define the at-risk process as $Y(t) = I(L < t \leq \tilde{R})$, acknowledging that a subject who has experienced the terminal event would not be considered as at risk for the recurrent event. We define $S_C(t|Z) = \Pr(L < t \leq R|Z)$.

The observed data include $n$ i.i.d replicates of $(L, \tilde{R}, \delta, N, Z)$, namely, $\{(L_i, \tilde{R}_i, \delta_i, N_i, Z_i)\}_{i=1}^n$. We assume that $(L, R]$ and $N^*(t)$ are conditionally independent given $Z$.

## 2.2 A Review of the GART model

Sun et al. (2016) proposed the generalized accelerated recurrence time (GART) model for recurrent events data in the absence of the terminal event (i.e. $D = \infty$). Define $\mu_{\boldsymbol{Z}}(t) = E\{N_0^*(t)|\boldsymbol{Z}\}$ and $\tau_{\boldsymbol{Z}}(u) = \inf\{t \geq 0 : \mu_{\boldsymbol{Z}}(t) \geq u\}$. The quantity $\mu_{\boldsymbol{Z}}(t)$ represents the mean function of recurrent events, and the quantity $\tau_{\mathbf{Z}}(u)$ is the so-called time to expected frequency $u$ (Huang and Peng, 2009). Suppose $\mu_{\mathbf{Z}}(t)$ is smooth and strictly increasing. By the definition of $\tau_{\mathbf{Z}}(u)$, the expected frequency (or mean

function) of recurrent events given covariates in $\mathbf{Z}$ at time $\tau_{\mathbf{Z}}(u)$ would equal $u$. This suggests that $\tau_{\mathbf{Z}}(u)$ can be roughly viewed as the inverse function of the mean function. By its definition, $\tau_{\mathbf{Z}}(u)$ has a direct "physical" interpretation on the time-scale (Reid, 1994).

Under the GART model, covariate effects are formulated on the time to expected frequency:

$$\tau_{\mathbf{Z}}(G(u)) = \exp\left\{\boldsymbol{Z}^T\boldsymbol{\beta}_0(u)\right\}, u \in (0, U]. \tag{2.1}$$

where $G(u) = \int_0^u g(s)ds$ and $g(\cdot)$ is a known positive and continuous function, and $U$ is a positive constant in the frequency scale. The non-intercept coefficients in $\boldsymbol{\beta}_0(u)$ represent the effects of the corresponding covariates on time to expected frequence $G(u)$. When all the non-intercept coefficients in $\boldsymbol{\beta}_0(u)$ are constant over $u$ and $G(u) = u$, it can be shown that model (2.1) reduces to the accelerated failure time (AFT) model for recurrent events data (Lin et al., 1998). If the event of interest is not recurrent (i.e. $T^{(j)} = \infty$ for $j \geq 2$), then $\tau_{\mathbf{Z}}(u)$ becomes the conditional quantile function of $T^{(1)}$ given $\mathbf{Z}$, and consequently model (2.1) reduces to a quantile regression model for $T^{(1)}$.

Sun et al. (2016) showed that the GART model has an equivalent formulation in terms of the counting process:

$$E\{N(e^{Z^T\beta_0(u)})|Z\} = E\{\int_0^u Y(e^{Z^T\beta_0(s)})g(s)ds|Z\},\ u \in (0, U], \tag{2.2}$$

with $g(u) = G'(u)$. This counting process formulation of the GART model greatly facilitates the estimation of $\boldsymbol{\beta}_0(u)$. Specifically, it suggests the following stochastic integral based estimating equation:

$$n^{-1/2}\sum_{i=1}^{n}\boldsymbol{Z}_i\left\{N_i(\exp\{\boldsymbol{Z}_i^\top\boldsymbol{\beta}(u)\}) - \int_0^u Y_i(\exp\{\boldsymbol{Z}_i^\top\boldsymbol{\beta}(s)\})g(s)ds\right\} = 0 \tag{2.3}$$

As elaborated in Sun et al. (2016), the estimating equation (2.3) can be stably and effectively solved by a sequence of $L_1$-minimization problems. Desirable asymptotic properties, such as uniform consistency and weak convergence to mean-zero Gaussian process, were established for the estimator derived based on Equation (2.3).

## 2.3 The Proposed Models and Inference Procedure

### 2.3.1 An extension of the GART model based on survivors' rate function

The survivors' rate function, defined as $\lambda^S_{\mathbf{Z}}(t) \doteq E\{dN^*(t)|D \geq t, \mathbf{Z}\}/dt$, has been used as a variant of the classic rate function for accounting for the presence of the terminal event (Cook and Lawless, 1997). The interpretation of $\lambda^S_{\mathbf{Z}}(t)$ targets the subgroup with $D \geq t$ and this shares the same rationale as that adopted by cause-specific hazard which is confined to a specific failure type in a competing risks setting. Let $\Lambda^S_{\mathbf{Z}}(u) \doteq \int_0^u \lambda_{\mathbf{Z}}(t)dt$ and we shall refer it to as the cumulative survivors' rate function.

We propose an extension of the GART model in the presence of the terminal event by viewing $\Lambda^S_{\mathbf{Z}}(u)$, the integral of the survivors' rate function, as the counterpart of the mean function $\mu_{\mathbf{Z}}(u)$, which is the integral of the classic rate function. Specifically, the GART model (2.1) is transformed to

$$\tau^S_{\mathbf{Z}}(G(u)) = \exp\left\{\mathbf{Z}^T\boldsymbol{\beta}^S_0(u)\right\}, u \in (0, U], \tag{2.4}$$

where $\tau^S_{\mathbf{Z}}(u) \doteq \inf\{t \geq 0 : \Lambda^S_{\mathbf{Z}}(t) \geq u\}$ stands for the time to expected cumulative survivors' rate $u$. Here $G(u)$ is defined the same as in model (2.1).

Interestingly, we can show that model (2.4) has the same counting process formulation as in (2.2); see Proposition 1 in Appendix A:

$$E\{N(e^{Z^\top \boldsymbol{\beta}_0^S(u)})|Z\} = \int_0^u Y(e^{Z^\top \boldsymbol{\beta}_0^S(u)})g(s)ds,\ u \in (0, U]. \tag{2.5}$$

An important implication from this finding is that we can directly apply Sun et al. (2016)'s estimation procedure, theory, and inference procedures, which were originally designed for the setting without the dependent terminal event, to address the proposed model (2.4). The critical distinction is about the coefficient interpretation. In the presence of the terminal event, the non-intercept coefficients in $\boldsymbol{\beta}_0^S(u)$ represent the covariate effects on the time to cumulative survivors' rate $G(u)$. The situation discussed here is analogous to that for univariate survival setting with dependent censoring. That is, the proportional cause-specific hazard regression for dependently censored data shares the same procedure with the standard proportional hazards regression for randomly censored data (Kalbfleisch and Prentice, 2002).

Following Sun et al. (2016), we can obtain an estimator of $\boldsymbol{\beta}_0^S(\cdot)$, denoted by $\hat{\boldsymbol{\beta}}^S(\cdot)$, as a right continuous piecewise-constant function that jumps only at the grid points of $S_{L(n)} = \{0 = u_0 < u_1 < \cdots < u_{L(n)} = U\}$. We set $\exp\{\mathbf{X}_i^\intercal \hat{\boldsymbol{\beta}}^S(0)\} = 0$ for all $i$, and then obtain $\hat{\boldsymbol{\beta}}^S(u_k)$, $k = 1, 2, \ldots, L(n)$, by sequentially locate the minimizer of the $L_1$-type convex function,

$$\begin{aligned} l_k(\mathbf{h}) = &\sum_{i=1}^{n}\sum_{j=1}^{\infty} I(L_i \le T_i^{(j)} \le \tilde{R}_i)\left|\log T_i^{(j)} - \mathbf{Z}_i^\intercal \mathbf{h}\right| + \left|R^* - \left\{\sum_{i=1}^{n}\sum_{j=1}^{\infty} I(L_i \le T_i^{(j)} \le \tilde{R}_i)(-\mathbf{Z}_i)^\intercal \mathbf{h}\right\}\right| \\ &+ \left|R^* - \left\{\sum_{i=1}^{n} 2\mathbf{Z}_i^\intercal \mathbf{h} \sum_{m=0}^{k-1} Y_i(\exp\{\mathbf{Z}_i^\intercal \hat{\boldsymbol{\beta}}^S(u_m)\}) \int_{u_m}^{u_{m+1}} g(s)ds\right\}\right|, \end{aligned}$$

where $R^*$ is a very large number and $j = 1, \ldots, L(n)$. Based on the results of Sun et al. (2016), $\hat{\boldsymbol{\beta}}^S(u)$ is uniformly consistent in $u$ and weakly convergence to a mean-zero Gaussian process at the root-$n$ rate under some regularity conditions. Similarly,

the inferences about $\boldsymbol{\beta}_0^S(\cdot)$ can be carried out by bootstrapping-based procedures.

### 2.3.2 Extension of the GART model based on the adjusted rate function

The counting process $N^*(t) \doteq \sum_{j=1}^{\infty} I(T^{(j)} \leq t \wedge D)$ naturally accounts for the presence of the terminal event, and provides the base for defining the adjusted rate function. That is, the adjusted rate function can be defined as $\lambda_{\boldsymbol{Z}}^A(t) \doteq E\{dN^*(t)|\boldsymbol{Z}\}/dt$. We call $\Lambda_{\mathbf{Z}}^A(t) \doteq \int_0^u \lambda_{\mathbf{Z}}^A(t)dt$ the cumulative adjusted rate function. It is easy to see that $\Lambda_{\mathbf{Z}}^A(t) = E\{N^*(t)|\boldsymbol{Z}\}$, which reflects the expected frequency of recurrent events before the occurrence of the terminal event. In the non-recurrent event setting, $\Lambda_{\mathbf{Z}}^A(t)$ reduces to the so-called cumulative incidence function (Kalbfleisch and Prentice, 2002).

Following the strategy of using the adjusted rate function to account for the presence of the terminal event, we propose an extension of the GART model that takes the form,

$$\tau_{\boldsymbol{Z}}^A(G(u)) = \exp\{\boldsymbol{Z}^\top \boldsymbol{\beta}_0^A(u)\}, \quad u \in (0, U] \tag{2.6}$$

where $\tau_{\boldsymbol{Z}}^A(u) = \inf\{t \geq 0 : \Lambda_{\boldsymbol{Z}}^A(t) \geq u\}$, and $G(u)$ is defined in the same way as in the GART model (2.1). The non-intercept coefficients in $\boldsymbol{\beta}_0^A(t)$ can be interpreted as covariate effects on time to cumulative adjusted rate $G(u)$.

In Proposition 2 in the Appendix, we show that model (2.6) implies

$$E\left\{\sum_{j=1}^{\infty} \frac{1}{S_C(T^{(j)}|\mathbf{Z})} I(L < T^{(j)} \leq e^{Z^\top \boldsymbol{\beta}_0^A(u)} \wedge \tilde{R}|\boldsymbol{Z})\right\} = \int_0^u g(s)ds,\ u \in (0, U]. \tag{2.7}$$

By this result, we propose to estimate $\boldsymbol{\beta}_0^A(u)$ based on the the estimating equation:

$$\boldsymbol{S}_n(\boldsymbol{\beta}, u) = 0 \tag{2.8}$$

where

$$S_n(\boldsymbol{\beta}, u) = n^{-1/2} \sum_{i=1}^{n} \boldsymbol{Z}_i \left\{ \sum_{j=1}^{\infty} \frac{1}{\hat{S}_C(T_i^{(j)}|\boldsymbol{Z})} I(L < T_i^{(j)} \leq e^{\boldsymbol{Z}_i^\top \boldsymbol{\beta}(u)} \wedge \tilde{R}) - \int_0^u g(s) ds \right\} \tag{2.9}$$

and $\hat{S}_C(\cdot|\mathbf{Z})$ is a reasonable estimator of $S_C(\cdot|\mathbf{Z})$.

It could be directly estimated from non-parametric methods, such as Dabrowska estimator. When adjusting with $\mathbf{Z}$, given $L < R$, we could use Cox PH model or AFT model to estimate marginal survival function $\hat{S}_L(\cdot|\mathbf{Z})$ and $\hat{S}_R(\cdot|\mathbf{Z})$ of $L$ and $R$ given $\mathbf{Z}$ and $\hat{S}_C(\cdot|\mathbf{Z}) = \hat{S}_R(\cdot|\mathbf{Z}) - \hat{S}_L(\cdot|\mathbf{Z})$.

For the presentation simplicity, in the sequel, we assume that $L$ and $R$ are independent of $\mathbf{Z}$. In this case, $S_C(t|\mathbf{Z})$ is free of $\mathbf{Z}$ and equals $\Pr(R \geq t) - \Pr(L \geq t)$. Since $R$ is only subject to the independent censoring by $D$ and $L$ is always observed, we can estimate $\Pr(R \geq t)$ by the left-continuous version of the Kaplan-Meier estimator of $\Pr(R > t)$, denoted by $\hat{G}^R(t)$, and estimate $\Pr(L \geq t)$ by its empirical counterpart. When $L$ and $R$ are believed to be covariate-dependent, we can impose regression modeling of $L$ and $R$ given $\mathbf{Z}$ to provide a reasonable estimate for $S_C(\cdot|\mathbf{Z})$.

Note that Equation (2.8) is monotone but not continuous. Thus, an exact solution may not exist. We then define an estimator of $\boldsymbol{\beta}_0^A(u)$, $\hat{\boldsymbol{\beta}}^A(u)$, as a generalized solution to Equation (2.8), which belongs to a convex set of size $O(n^{-1})$ (Fygenson and Ritov, 1994). Following the arguments in Peng and Fine (2009), we only need to solve Equation (2.8) on a fine grid $S_{L(n)}^A = \{0 = u_0 < u_1 < \cdots < u_{L(n)} = U\}$, and then let $\hat{\boldsymbol{\beta}}^A(\cdot)$ be a right continuous piecewise-constant function that jumps only at the grid points of $S_{L(n)}^A$. We can show that locating $\hat{\boldsymbol{\beta}}^A(u_k)$ $(k = 1, \cdots, L(n))$ is equivalent to

finding the minimizer of

$$U_n(\boldsymbol{h},u) = n^{-1/2}\sum_{i=1}^{n}\sum_{j=1}^{\infty}\frac{1}{\hat{S}_C(T_i^{(j)}|\boldsymbol{Z})}I(L_i < T_i^{(j)} \le \tilde{R}_i)\big|\log T_i^{(j)} - \boldsymbol{Z}_i^\top\boldsymbol{h}\big|$$
$$+\Big|R^* - \Big\{\sum_{i=1}^{n}\sum_{j=1}^{\infty}\frac{1}{\hat{S}_C(T_i^{(j)}|\boldsymbol{Z})}I(L_i < T_i^{(j)} \le \tilde{R}_i)(-\boldsymbol{Z}_i)^\top\boldsymbol{h}\Big\}\Big| + \Big|R^* - \Big\{\sum_{i=1}^{n}2\boldsymbol{Z}_i^\top\boldsymbol{h}\int_0^u g(s)ds\Big\}\Big| \tag{2.10}$$

where $R^*$ is a sufficiently large number. This is because $\partial U_n(\boldsymbol{h},u)/\partial\boldsymbol{h}$ equals to two times $n^{1/2}S_n(\boldsymbol{\beta},u)$ when $R^*$ is chosen large enough to bound $|\sum_{i=1}^n\sum_{j=1}^\infty \frac{1}{\hat{S}_C(T_i^{(j)}|\boldsymbol{Z})}I(L_i < T_i^{(j)} \le \tilde{R}_i)(-\boldsymbol{Z}_i)^\top\boldsymbol{h}|$ and $|\sum_{i=1}^n 2\boldsymbol{Z}_i^\top\boldsymbol{h}\int_0^{u_k} g(s)ds|$. The minimization of $U_n(\boldsymbol{h},u)$ can be easily solved by using standard statistical software, such as the $l1fit()$ function in S-PLUS or the $rq()$ function in R package *quantreg*.

### 2.3.3 Asymptotic properties

We establish the uniform consistency and weak convergence of the proposed estimator $\hat{\boldsymbol{\beta}}^A(\cdot)$. We first state the regularity conditions:

(C1) There exists $v_R > 0$ such that $\Pr(R = v_R) > 0$ and $\Pr(R > v_R) = 0$. In addition, $\Pr(R > L) = 1$.

(C2) $\boldsymbol{Z}$ and $N(\tilde{R})$ are bounded.

(C3) (i) $\boldsymbol{\beta}_0^A(u)$ is Lipschitz continuous in $u \in [0,U]$; (ii) $\lambda_{\boldsymbol{Z}}^A(t) = E\{dN^*(t)\mathbf{Z}\}/dt$ is bounded above uniformly in $t$ and $\boldsymbol{Z}$.

(C4) For some $\rho_0 > 0$ and $c_0 > 0$, $\inf_{\boldsymbol{b}\in\mathcal{B}(\rho_0)}\text{eigmin}\boldsymbol{A}(\boldsymbol{b}) \ge c_0$, where $\mathcal{B}(\rho) = \{\boldsymbol{b} \in \mathbb{R}^{p+1} : \inf_{u\in[0,U]}||\boldsymbol{b} - \boldsymbol{\beta}_0(u)|| \le \rho\}$ and $\boldsymbol{A}(\boldsymbol{b}) = E\{\boldsymbol{Z}^{\otimes 2}\lambda_{\boldsymbol{Z}}^A(t)|\boldsymbol{Z}\}$. Here $||\cdot||$ is the Euclidean norm, and we define $\boldsymbol{u}^{\otimes 2} = \boldsymbol{u}\boldsymbol{u}^\top$ for a vector $\boldsymbol{u}$.

Condition (C1) is assumed to ensure the inverse weights $\{S_C(T_i^{(j)})\}^{-1}$ can be consistently estimated. This condition is usually satisfied in follow-up studies with administrative censoring or by imposing artificial truncation to the observed recur-

rent events. Conditions (C2) and (C3) are realistic assumptions; similar conditions are also adopted in Sun et al. (2016) for the GART model in the absence of the dependent terminal event. Condition (C4) implies that $S_n(\boldsymbol{\beta}, u)$ is strictly monotone in a neighborhood of $\boldsymbol{\beta}_0(u)$ for $u \in (0, U]$. This entails the identifiability of $\boldsymbol{\beta}_0(u)$ and the consistency of $\hat{\boldsymbol{\beta}}(u)$.

Under the regularity conditions (C1)-(C4), we have the following theorems:

**Theorem 2.1.** *Suppose model* (2.6) *holds for* $u \in [0, U]$. *Then under conditions C1-C4,* $\lim_{n\to\infty} \sup_{u\in[\nu,U]} ||\hat{\boldsymbol{\beta}}^A(u) - \boldsymbol{\beta_0}(u)|| \to_p 0$, *where* $0 < \nu < U$.

**Theorem 2.2.** *Suppose model* (2.6) *holds for* $u \in [0, U]$. *Then under conditions C1-C4,* $n^{1/2}\{\hat{\boldsymbol{\beta}}^A(u) - \boldsymbol{\beta_0}(u)\}$ *converge weakly to a mean zero Gaussian process for* $u \in [0, U]$ *with covariance* $\boldsymbol{\Phi}(u', u) = \boldsymbol{A}\{\boldsymbol{\beta}_0^A(u')\}^{-1} E\{\boldsymbol{\xi}(u')\boldsymbol{\xi}(u)^\top\}\boldsymbol{A}\{\boldsymbol{\beta}_0^A(u)\}^{-1}$, *where* $\boldsymbol{\xi}(u)$ *is defined in Equation* (2.15) *in Appendix B.*

The proofs of Theorems 2.1-2.2 follow the similar arguments in Peng and Fine (2009). The detailed proofs are provided in Appendix B.

### 2.3.4 Inference

To make inference on $\boldsymbol{\beta}_0(u)$, we can apply a bootstrapping procedure such as the classical resampling with replacement, or resampling perturbed estimating equation Jin et al. (2001).

Alternatively, we can also perform sample-based inference following the lines of Peng and Fine (2009). More specifically, let $\hat{\boldsymbol{\Sigma}}(u, v)$ denote a consistent plug-in estimator of $\boldsymbol{\Sigma}(u, v)$, which stands for the asymptotic covariance matrix of $\mathbf{S}_n(\boldsymbol{\beta}_0(u), u)$ and is defined in (2.18) in Appendix B. An consistent estimator $\hat{\boldsymbol{\Sigma}}(u, u)$ for $\boldsymbol{\Sigma}(u, u)$ maybe given by

$$\hat{\boldsymbol{\Sigma}}(u, u) = \frac{1}{n}\sum_{i=1}^{n}(\hat{\boldsymbol{\xi}}_{1,i}(u) - \hat{\boldsymbol{\xi}}_{2,i}(u))^{\otimes 2}$$

where $\hat{\boldsymbol{\xi}}_{1,i}(u) = \mathbf{Z}_i\{\sum_{j=1}^{\infty} \frac{1}{\hat{S}_C(T_i^{(j)}|\boldsymbol{Z})} I(L_i < T_i^{(j)} \leq e^{\boldsymbol{Z_i}^\top \hat{\boldsymbol{\beta}}^A(u)} \wedge \tilde{R}_i) - \int_0^u g(s)ds\}$, $\hat{\boldsymbol{\xi}}_{2,i}(u) = \frac{1}{n}\sum_{k=1}^{n} \boldsymbol{Z}_k \Big\{ \sum_{j=1}^{\infty} \frac{\hat{\boldsymbol{\xi}}_{\hat{S}_C,i}(T_k^{(j)})}{\hat{S}_C^2(T_k^{(j)})} \times I(L_k < T_k^{(j)} \leq e^{\boldsymbol{Z}_k^\top \hat{\boldsymbol{\beta}}^A(u)} \wedge \tilde{R}_k) \Big\}$, and $\hat{\boldsymbol{\xi}}_{\hat{S}_c,i}(T_k^{(j)}) = \hat{G}^R(T_k^{(j)}) I(T_k^{(j)} \geq \tilde{R}_i, \delta_i = 0) / \sum_{m=1}^{n} I(\tilde{R}_m \geq \tilde{R}_i) - \{I(L_i \geq t) - \frac{1}{n}\sum_{m=1}^{n} I(L_m \geq T_k^{(j)})\}$.

Then, find a symmetric and nonsigular $(p+1)\times(p+1)$ matrix $\boldsymbol{E}_n(u) = \{\boldsymbol{e}_{n,1}(u), \ldots, \boldsymbol{e}_{n,p+1}(u)\}$ such that $\hat{\boldsymbol{\Sigma}}(u,u) = \{E_n(u)\}^2$. Next, calculate $\boldsymbol{D}_n(u) = (\boldsymbol{S}_n^{-1}\{\boldsymbol{e}_{n,1}(u), u\} - \hat{\boldsymbol{\beta}}(u), \ldots, \boldsymbol{S}_n^{-1}\{\boldsymbol{e}_{n,p+1}(u), u\} - \hat{\boldsymbol{\beta}}(u))$, where $S_n^{-1}\{\boldsymbol{e}, u\}$ is defined as the solution to $S_n\{\boldsymbol{b}, u\} = \boldsymbol{e}$. Finally, we can estimate the asymptotic covariance matrix of $n^{1/2}\{\hat{\boldsymbol{\beta}}(u) - \boldsymbol{\beta}_0(u)\}$ and $n^{1/2}\{\hat{\boldsymbol{\beta}}(u') - \boldsymbol{\beta}_0(u')\}$ by $n\boldsymbol{D}_n(u')\boldsymbol{E}_n(u')^{-1}\hat{\boldsymbol{\Sigma}}(u',u)\boldsymbol{E}_n(u)^{-1}\boldsymbol{D}_n(u)^T$. With $u = u'$, the asymptotic variance matrix of $n^{1/2}\{\hat{\boldsymbol{\beta}}(u) - \boldsymbol{\beta}_0(u)\}$ can then be estimated by $n\{\boldsymbol{D}_n(u)\}^{\otimes 2}$.

## 2.4 Numerical Studies

### 2.4.1 Monte-Carlo simulations

We conduct Monte-Carlo simulations to evaluate the proposed method for the extended GART model (2.6) based on the adjusted rate function. We generate covariates, $Z_1$ and $Z_2$, respectively from $Bernoulli(0.5)$ and $Uniform(-5,5)$ distributions. Define $\eta_j = I(T^{(j)} \leq D)$. We generate $\eta_j$ $(j = 1, 2, \ldots)$ as Bernoulli random variables that satisfy $\Pr(\eta_1 = 1) = p$ and $\Pr(\eta_{j+1} = 1 | \eta_j = 1) = p$, $\Pr(\eta_{j+1} = 1 | \eta_j = 0) = 0$. The value of $p$ determines the number of recurrent events before the terminal event; setting a larger $p$ tends to generate more recurrent events before the terminal event. Define $T_{j,D} = \exp\{\frac{T^{*(j)}}{3\gamma} Z_1 + \min(0.2, \frac{T^{*(j)}}{15\gamma}) Z_2\} T^{*(j)}/\gamma$, where $\{T^{*(j)},\ j = 1, 2, \ldots\}$ are produced from a standard homogeneous Poisson process and $\gamma$ follows the $Gamma(2,2)$ distribution. For $j \geq 1$ with $\eta_j = 1$, we let $T^{(j)} = T_{j,D}$; for $j$ corresponding to the first $\eta_j = 0$, we let $D = T_{j,D}$. Under this

set-up, we can show that

$$\tau_{\mathbf{Z}}^{A}(u) = \exp\Bigg\{ \log\Big(\frac{1}{1-p}\big[\frac{2}{\{1-u(1-p)/p\}^{1/2}} - 2\big]\Big) + \frac{1}{3-3p}\big[\frac{2}{\{1-u(1-p)/p\}^{1/2}} - 2\big] Z_1 + \min\Big(0.2, \frac{1}{15-15p}\big[\frac{2}{\sqrt{1-u(1-p)/p}} - 2\big]\Big) Z_2 \Bigg\}.$$

This indicates that model (2.6) holds with $g(u) = u$. The effect of $Z_1$ on $\tau_{\mathbf{Z}}^{A}(u)$ is increasing with $u$, and the effect of $Z_2$ rises first and then becomes constant as $u$ increases. Finally, we generate $L$ as $w_1 \cdot Unif(0,1)$, where $w_1$ follows $Bernoulli(0.8)$, and generate $R$ as $w_2 \cdot Unif(L, 30) + (1 - w_2) \cdot 30$, where $w_2$ follows $Bernoulli(0.8)$.

In our simulations, we consider $p = 0.8, 0.85, 0.9, 0.95$. In each setting, we generate $1,000$ datasets with sample size 200. The estimator $\hat{\boldsymbol{\beta}}^{A}(u)$ is calculated on an equally spaced $u$-grid with 150 grid points. For $p = 0.8, 0.85, 0.9, 0.95$, the range of the $u$-grid is set as $(0, 1.5]$, $(0, 2.0]$, $(0, 2.5]$ and $(0, 3.0]$ respectively. When carrying out bootstrapping-based inference, we set the size of resampling as 100.

In Figure 2.1, we present the estimated coefficients for model (2.6) based on the method proposed in Section 3.2. We also plot the coefficient estimates obtained from applying Sun et al. (2016)'s method, which assesses the covariate effects on $\tau_{\mathbf{Z}}^{S}(u)$ (instead of $\tau_{\mathbf{Z}}^{A}(u)$). It is clearly shown that the proposed estimator $\hat{\boldsymbol{\beta}}^{A}(u)$ is virtually unbiased. Naively using Sun et al. (2016)'s method would produce biased estimates for the covariate effects on $\tau_{\mathbf{Z}}^{A}(u)$, particularly when $u$ is large. It is also noted as $p$ increases, the departure of the empirical averages of Sun et al. (2016)'s estimates from the true $\boldsymbol{\beta}_0^{A}(u)$'s decrease. This is reasonable because when $p$ is closer to 1, the terminal event is more unlikely to occur before the end of the observation (i.e. $R$). Consequently, we expect $\tau_{\mathbf{Z}}^{A}(u)$ and $\tau_{\mathbf{Z}}^{S}(u)$ would be more similar and hence the method targeting $\tau_{\mathbf{Z}}^{A}(u)$ and Sun et al. (2016)'s method which targets $\tau_{\mathbf{Z}}^{S}(u)$ would produce more agreeable results.

In Figure 2.2, we compare the estimated standard errors (SE) based on the sample-

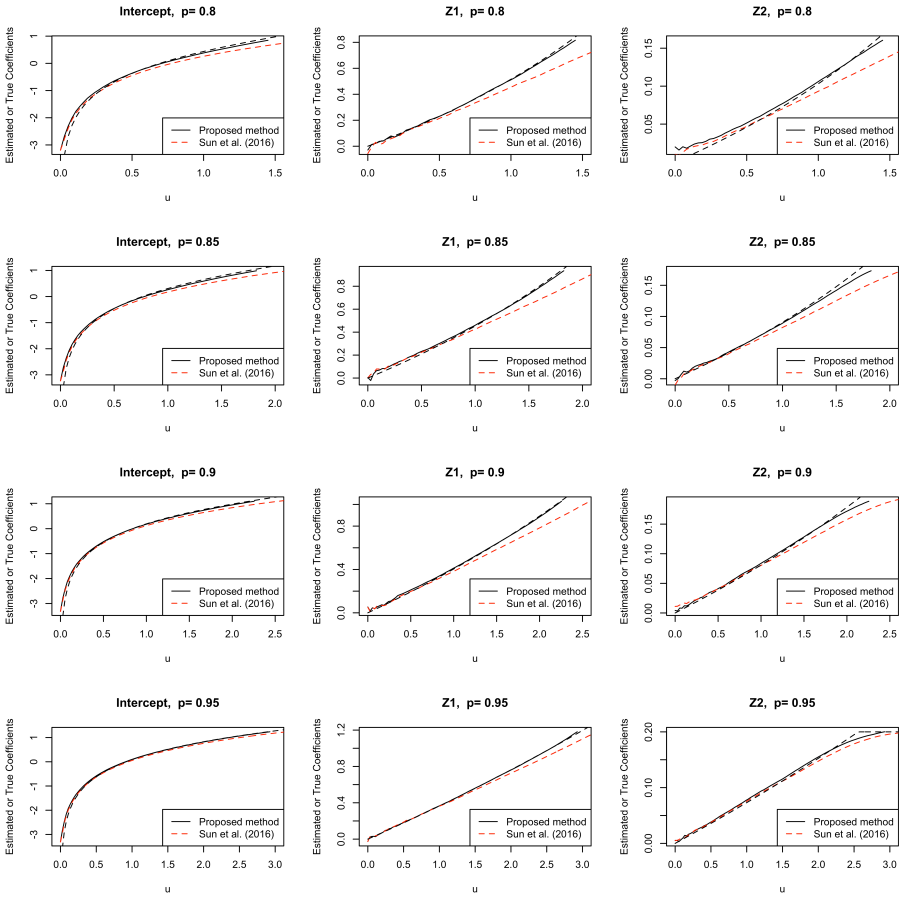


Figure 2.1: The proposed coefficient estimates for model (2.6) (solid lines), along with the true coefficients $\boldsymbol{\beta}_0^A(u)$ (dashed lines) and the coefficient estimates obtained from applying Sun et al. (2016)'s method (dotted lines) when sample size n=200

based inference procedure and those based on bootstrapping with the empirical standard deviations (SD) of the coefficient estimates. It is shown that both sample-based SEs and bootstrapping-based SEs are close to the empirical SDs in each setting except those at very small $u$'s. The bootstrapping-based SEs are slightly closer to the empirical SDs as compared to sample-based standard errors.

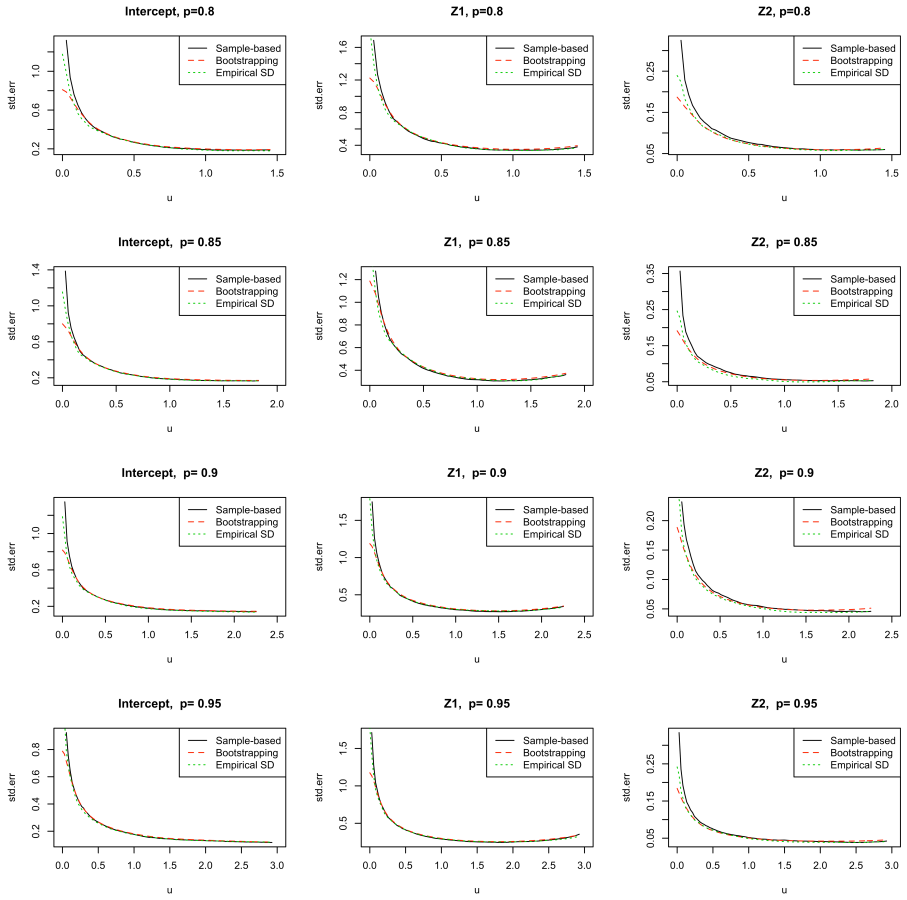


Figure 2.2: Estimated standard errors based on sample-based inference procedure (solid lines), estimated standard errors based on bootstrapping (dashed lines), and empirical standard deviations (dotted lines) with sample size n=200

We also evaluate the empirical coverage probabilities of the 95% confidence intervals (CI) constructed based on the sample-based and bootstrapping-based inference procedures. Figure 2.3 shows that the empirical coverage probabilities for the coeffi-

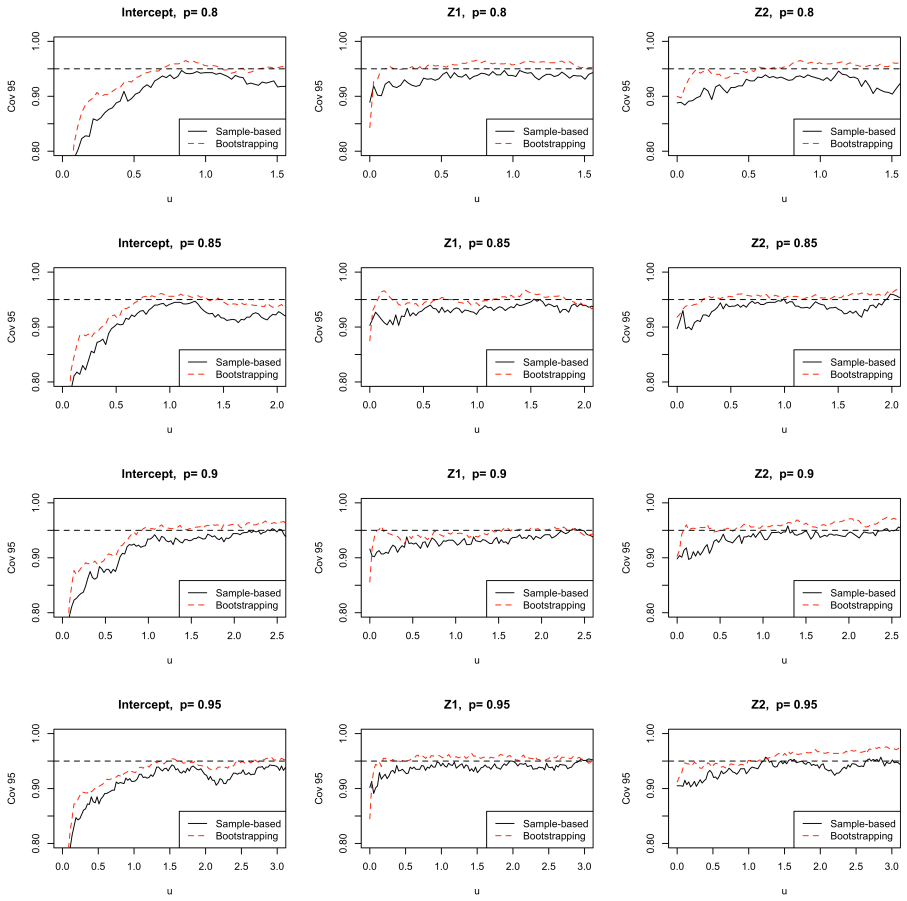


Figure 2.3: Empirical coverage probabilities of the 95% confidence intervals constructed based on the sample-based inference procedure (solid lines) and bootstrapping procedure (dashed lines) with sample size n=200

cients for $Z_1$ and $Z_2$ are fairly close to the nominal value 95%. The bootstrapping-based confidence intervals perform slightly better than the sample-based confidence intervals. For the intercept, the confidence intervals seem to be undercovered, particularly for small $u$'s. In simulations with sample size 400 (unreported here), we observe a clear improvement in the empirical coverage probabilities for the intercept.

### 2.4.2 An application to a dataset from the Cystic Fibrosis Foundation Patient Registry

Cystic Fibrosis (CF) is one of the most common, life-shortening genetic disorders with an incidence of 1:3500 in newborns in the United States (Russell et al., 2012). The leading cause of the premature death is obstructive lung disease with recurrent respiratory infections, inflammations, and structural airway damage. Pseudomonas aeruginosa (PA), a ubiquitous environmental bacterium, is one of the major pathogens in CF lungs, which is associated with poor clinical outcomes and greater mortality (Davies, 2002). Respiratory tract cultures are routinely obtained for identifying PA and characterizing its phenotypes (mucoid or non-mucoid). The early PA infection is usually non-mucoid and antibiotic sensitive. But recurrent of non-mucoid PA infection leads to chronic PA infection, then to mucoid PA phenotype (Mathee et al., 1999). The development of mucoid PA yet can be more complicated than this widely held paradigm (Heltshe et al., 2018). Mucoid PA is more resistant to antibiotics and more difficult to eradicate (Lyczak et al., 2002). As a result, rarely patients can go back to the non-mucoid PA infection stage once acquiring a mucoid PA infection. Under these considerations, a mucoid PA infection constitutes a dependent terminal event to the recurrent process of non-mucoid PA infections (in addition to death).

We apply the proposed method to a sub-dataset from the 2008 Cystic Fibrosis Foundation Patient Registry (CFFPR) data, which includes $1,974$ children who were born in or after 2000 with CF and had more than 5 years' follow-up. The objective of our analysis is to assess how several potential risk factors influence the recurrence of non-mucoid PA infections prior to the mucoid PA infection while alive. To this end, we set the time origin as the birth. We define the recurrent event time $T^{(j)}$ as the age at the $j$-th non-mucoid infection, and time to the terminal event $D$ as the age at the first mucoid PA infection or death, whichever occurred first. Age at the first CFFPR visit and age at the last follow-up visit correspond to the $L$ and $R$ respectively.

Table 2.1: Summary statistics of the potential risk factors in the CFFPR dataset ($n = 1974$)

| Potential risk factors | | $n(\%)$ |
|---|---|---|
| Sex | Female | 1024 (51.9%) |
| | Male | 950 (48.1%) |
| F508del | Heterogeneous | 1274(64.5%) |
| | Homogeneous/Other | 700 (35.5%) |
| Meconuim ileus | Yes | 534 (72.9%) |
| | No | 1440 (27.1%) |
| Pancreatic insufficiency status | Insufficient | 1810 (91.7%) |
| | Sufficient | 164 (8.3%) |

In our dataset, a total of $3,459$ non-mucoid PA infections before mucoid PA infections were documented, and 472 subjects experienced mucoid PA infections during the follow-up. There are 14 subjects who died before the first mucoid PA infection. Within each subject, the number of non-mucoid PA infections before the first mucoid PA infection range from 0 to 19, with mean and median equal to 1.75 and 1 respectively. We consider risk factors including sex (coded as $Sex = 1$ if female and 0 otherwise), patient's CFTR genotype (coded as $F508/Other = 1$ if F508del heterogeneous and 0 otherwise) meconium ileus (MI) status (coded as $MI = 1$ if having the diagnosis of MI and 0 otherwise, and pancreatic insufficiency status (coded as $Pancreat = 1$ if pancreatic insufficient and 0 otherwise). Table 2.1 provides a summary of these potential risk factors.

We first fit our dataset the extended GART model based on the adjusted rate function, model (2.6), with $g(u) = 1$. In Figure 2.4, we plot the estimated regression coefficients with 95% pointwise confidence intervals (CI). The intercept coefficient estimates represent the estimated log time to cumulative adjusted rate (or alternatively, expected frequency of non-mucoid PA infection before mucoid PA infection and death) for the reference group, which included CF boys with homozygous F508del mu-

tations who had no MI and were pancreatic sufficient. For example, the estimated intercept coefficient plot suggests that the expected frequency of non-mucoid PA infection before mucoid PA infection while alive reaches 1 approximately at the age of 4.5 years.

The non-intercept coefficients estimates represent the estimated effects of covariates on $\tau_{\mathbf{Z}}^{A}(u)$.For example, the estimated effects of MI at $u = 1$ suggests the estimated time to cumulative adjusted rate $u = 1$ between CF boys with homozygous F508del mutations who had MI and were pancreatic sufficient is 0.6 times of the time to cumulative adjusted rate $u = 1$ for reference group. Negative estimates indicate quicker progression to non-mucoid PA infection recurrence in the presence of mucoid PA infection and death. From Figure 2.4, it is observed that the coefficients for $Sex$ and $F508$ are mostly small with the 95% CIs fully covering zero. This suggests that gender and F508 genotype may have little effect on the acquisition and the recurrence of non-mucoid PA infections. The coefficients for $MI$ and $Pancreat$ are all negative and moreover the upper bounds of the corresponding 95% CIs are mostly below zero. This indicates that CF children with MI or pancreatic insufficiency tend to have more rapid recurrence of non-mucoid PA infections compared to those without MI or pancreatic insufficiency. This finding is consistent with our expectation because MI and pancreatic insufficiency are generally known to be associated with worse prognosis of CF outcomes.

We also plot the coefficient estimates for the extended GART model (2.4) based on the survivors' rate function. As justified in Section 3, we obtain the coefficient estimates from implementing Sun et al. (2016)'s method while treating the mucoid PA infection as a part of the random observation window (i.e. setting $R$ as the age at the first mucoid PA infection or death if either of these events occurred, otherwise the age at the last follow-up visit).The intercept coefficient estimate represents the estimated log time to cumulative survivors' rate for the reference group, which included CF boys

with homozygous F508del mutations who had no MI and were pancreatic sufficient. For example, the estimated intercept coefficient plot suggests that the expected frequency of non-mucoid PA infection before mucoid PA infection among the subjects that who are free of mucoid-PA infections and death reaches 1 approximately at the age of 4.05 years. The coefficient estimates are quite similar to those for model (2.6) except for the estimated coefficients for $MI$. In Figure 2.4, the coefficient estimates for $MI$ suggest that given mucoid PA and death haven't occurred, the timing of non-mucoid PA infection may be similar between CF children with MI and those without MI, while the MI phenotype seems to have a significant negative impact on time to cumulative adjusted rate of non-mucoid infections. The discrepancy relating to $MI$'s effect can be explained by appropriately understanding the distinction between $\tau^S_{\mathbf{Z}}(u)$ and $\tau^A_{\mathbf{Z}}(u)$. More specifically, the MI phenotype is generally associated with poorer clinical and survival outcomes in CF patients (Sawyer et al., 1994; Oliveira et al., 2002). Based on this dataset, a simple log-rank test suggests worse mucoid PA infection free survival (i.e. $\Pr(D > t)$) for the MI group compared to the non-MI group ($p = 0.036$). We also use the Peng and Huang (2008)'s model to assess the effects of MI on the mucoid PA infections and death at different quantiles. The results shows patients with MI have higher risk of mucoid PA infections and death. As discussed in Luo et al. (2010), the adjusted rate function and survivors' rate function are not identical to the rate function. Specifically, we may expect that the survivors in the MI group tend to be "stronger" or "less fragile" than the survivors in the non-MI group (where "survivors" in the context of this example refer to as subjects who haven't died and developed mucoid PA infection). Consequently, the comparison of the survivors' timing of recurrent non-mucoid PA infections between the MI group and the non-MI group would be shifted in favor towards the MI group. This may result in the observed attenuated effect of $MI$ on $\tau^S_{\mathbf{Z}}(u)$.

Overall, the presented analyses of the CFFPR dataset provide alternative views of

risk factors for recurrent nonmucoid PA infections under the GART framework. The possible dependent termination by mucoid PA infection and death are appropriately handled and interpreted based on the proposed models and estimation methods.

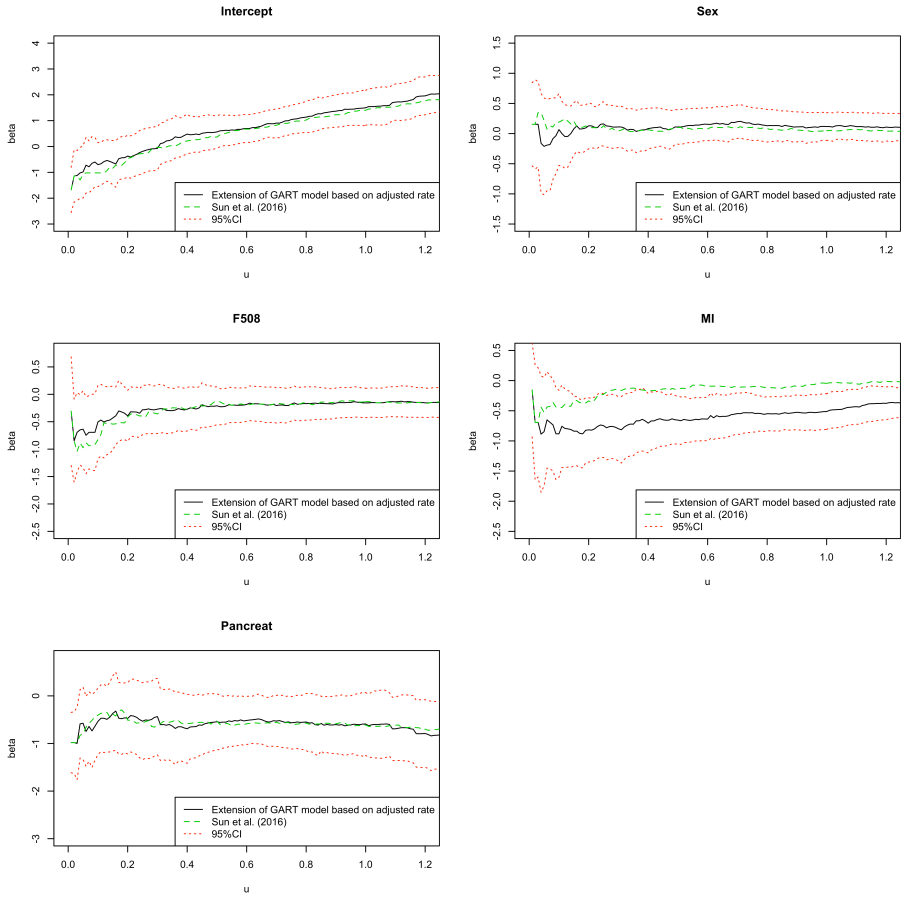


Figure 2.4: CFFPR data example: coefficient estimates (solid lines) and 95% pointwise CIs (dotted lines) for the extended GART model (6) based on the adjusted rate function, and the coefficient estimates for the extended GART model (4) based on the survivors' rate function (dashed lines)

## 2.5 Remarks

In this paper, we investigate two extensions of the generalized accelerated recurrence time (GART) model for recurrent events data with a dependent terminal event.

We adapt the GART modeling based on survivors' rate function and adjusted rate function, which are established crude quantities for accommodating the presence of the terminal event in recurrent events settings. Through our investigation, we find that directly applying the existing GART method that assumes a random observation window (Sun et al., 2016) exactly renders the estimates for the extended GART model (2.4) based on the survivors' rate function. Such a connection is analogous to the situation where one applies the standard Cox regression to dependently censored non-recurrent event data. The procedure would remain valid as long as the interpretation of coefficient estimates is tuned towards cause-specific proportional hazards regression.

The survivors' rate function and adjusted rate function are the extensions of cause-specific hazard rate and cumulative incidence rate in competing risk data. Similar to the comparison between cause-specific hazard rate and cumulative incidence rate in Lau et al. (2009), the extension of GART model based on the survivors' rate function should be better for studying the etiology of the disease since the survivors' rate function denotes the instantaneous rate of the recurrent event in those subjects who are currently event free. The extension of GART model based on adjusted rate function is more preferable if the primary interest is to estimate the effect of covariates on the incidence of the recurrent event or evaluate actual risks and prognosis (Koller et al., 2012).

For simplicity, we assume the independence between $(L, R)$ and $Z$ to simplify the estimation of $S_c(\cdot)$, which is needed in the proposed extension of GART model based on the adjusted rate function. Similar to Peng and Fine (2009), this assumption can be relaxed providing some reasonable modeling of the relationship between $(L, R)$ and $\mathbf{Z}$.

## 2.6 Appendix

### 2.6.1 Appendix A: Justification of the Counting Process Formulations of Model (4) and Model (6)

The following are some regularity conditions:

(B0) $(L, R)$ and $T^{(j)}$ are independent given $\mathbf{Z}$;

(B1) $(L, R)$ and $D$ are independent given $\mathbf{Z}$;

(B2) $\boldsymbol{\beta}_0^S(u)$ is continuously differentiable.

(B3) $S_C(e^{\mathbf{Z}^\top \boldsymbol{\beta}_0^S(u)}|\mathbf{Z}) > 0$ and $\Pr\{e^{\mathbf{Z}^\top \boldsymbol{\beta}_0^S(u)} < D|\mathbf{Z}\} > 0$ for $\mathbf{Z} \in \mathbf{Z}$ and $u \in (0, U]$

*Proposition A1.* Under conditions (B0)-(B3), model (4) and model (5) are equivalent.

*Proof of Proposition A1:* Given the random observation window assumptions in (B0) and (B1), we get

$$
\begin{aligned}
& E\{dN(e^{\mathbf{Z}^\top \boldsymbol{\beta}_0^S(u)})|\mathbf{Z}\} \\
&= E\{\sum_{j=1}^{\infty} I(e^{\mathbf{Z}^\top \boldsymbol{\beta}_0^S(u)} \le T^{(j)} < e^{\mathbf{Z}^\top \boldsymbol{\beta}_0^S(u+du)}, e^{\mathbf{Z}^\top \boldsymbol{\beta}_0^S(u)} \le D, L < e^{\mathbf{Z}^\top \boldsymbol{\beta}_0^S(u)} \le R)|\mathbf{Z}\} \quad (2.11) \\
&\qquad \cdot e^{\mathbf{Z}^\top \boldsymbol{\beta}_0^S(u)} \mathbf{Z}^\top d\boldsymbol{\beta}_0^S(u) \\
&= E\{\sum_{j=1}^{\infty} I(e^{\mathbf{Z}^\top \boldsymbol{\beta}_0^S(u)} \le T^{(j)} < e^{\mathbf{Z}^\top \boldsymbol{\beta}_0^S(u+du)}, e^{\mathbf{Z}^\top \boldsymbol{\beta}_0^S(u)} \le D|\mathbf{Z}\} \Pr\{L < e^{\mathbf{Z}^\top \boldsymbol{\beta}_0^S(u)} \le R)|\mathbf{Z}\} \\
&\qquad \cdot e^{\mathbf{Z}^\top \boldsymbol{\beta}_0^S(u)} \mathbf{Z}^\top d\boldsymbol{\beta}_0^S(u) \\
&= E\{dN^*(e^{\mathbf{Z}^\top \boldsymbol{\beta}_0^S(u)})|e^{\mathbf{Z}^\top \boldsymbol{\beta}_0^S(u)} \le D, \mathbf{Z}\} \Pr\{e^{\mathbf{Z}^\top \boldsymbol{\beta}_0^S(u)} \le D|\mathbf{Z}\} S_C(e^{\mathbf{Z}^\top \boldsymbol{\beta}_0^S(u)}|\mathbf{Z}) \\
&= d\Lambda_{\mathbf{Z}}^S(e^{\mathbf{Z}^\top \boldsymbol{\beta}_0^S(u)}) \Pr\{e^{\mathbf{Z}^\top \boldsymbol{\beta}_0^S(u)} \le D|\mathbf{Z}\} S_C(e^{\mathbf{Z}^\top \boldsymbol{\beta}_0^S(u)}|\mathbf{Z}) \quad (2.12)
\end{aligned}
$$

and

$$
\begin{aligned}
&E\{Y(e^{\mathbf{Z}^\top \boldsymbol{\beta}_0^S(u)})g(u)du\} \\
&= E\{I(L < e^{\mathbf{Z}^\top \boldsymbol{\beta}_0^S(u)} \le R)I(e^{\mathbf{Z}^\top \boldsymbol{\beta}_0^S(u)} \le D)|\mathbf{Z}\}g(u)du \\
&= E\{I(L < e^{\mathbf{Z}^\top \boldsymbol{\beta}_0^S(u)} \le R)|\mathbf{Z}\}E\{I(e^{\mathbf{Z}^\top \boldsymbol{\beta}_0^S(u)} < D)|\mathbf{Z}\}g(u)du \\
&= S_C(e^{\mathbf{Z}^\top \boldsymbol{\beta}_0^S(u)}|\mathbf{Z})\Pr\{e^{\mathbf{Z}^\top \boldsymbol{\beta}_0^S(u)} < D|\mathbf{Z}\}g(u)du
\end{aligned} \tag{2.13}
$$

Given assumption (B3), by comparing (2.12) and (2.13), we can easily see that model (4) implies model (5) and, on the other hand, model (5) implies model (4). This completes the proof of Proposition A1.

*Proposition A2.* Under conditions (B0)–(B3), model (6) implies (7).

*Proof of Proposition A2:* Define $T^{*(j)} = I(T^{(j)} \le D) \times T^{(j)} + I(T^{(j)} > D) \times \infty$. Given the random observation window assumptions, (B0) and (B1), and by the definition of $T^{*(j)}$, we get

$$
\begin{aligned}
&E\{\sum_{j=1}^{\infty} \frac{1}{S_C(T^{(j)}|\mathbf{Z})} I(L < T^{(j)} \le e^{\mathbf{Z}^\top \boldsymbol{\beta}_0^A(u)} \wedge \tilde{R})|\mathbf{Z}\} \\
&= E\{\sum_{j=1}^{\infty} \frac{1}{S_C(T^{*(j)}|\mathbf{Z})} I(L < T^{*(j)} \le e^{\mathbf{Z}^\top \boldsymbol{\beta}_0^A(u)} \wedge R)|\mathbf{Z}\} \\
&= E\left\{E\{\sum_{j=1}^{\infty} \frac{1}{S_C(T^{*(j)}|\mathbf{Z})} I(L < T^{*(j)} \le R)I(T^{*(j)} \le e^{\mathbf{Z}^\top \boldsymbol{\beta}_0^A(u)})|T^{*(j)}, \mathbf{Z}\}|\mathbf{Z}\right\} \\
&= E\left\{E\{\sum_{j=1}^{\infty} \frac{1}{S_C(T^{*(j)}|\mathbf{Z})} S_C(T^{*(j)}|\mathbf{Z})I(T^{*(j)} \le e^{\mathbf{Z}^\top \boldsymbol{\beta}_0^A(u)})|T^{*(j)}, \mathbf{Z}\}|\mathbf{Z}\right\} \\
&= E(N^*(e^{\mathbf{Z}^\top \boldsymbol{\beta}_0^A(u)})|\mathbf{Z}) \\
&= \mu_{\mathbf{Z}}(e^{\mathbf{Z}^\top \boldsymbol{\beta}_0^A(u)})
\end{aligned} \tag{2.14}
$$

By the definition of $\tau_{\mathbf{Z}}(\cdot)$, model (6) implies $\mu_{\mathbf{Z}}(e^{\mathbf{Z}^\top \boldsymbol{\beta}_0^A(u)}) = G(u)$ and hence (7) holds.

### 2.6.2 Appendix B: Proofs of Theorem 2.1 and Theorem 2.2

Define $G^R(t) = Pr(R \geq t)$, $G^L(t) = Pr(L \geq t)$, $N_i^R(t) = I(\tilde{R}_i \leq t, \delta_i = 0)$, $Y_i^R(t) = I(\tilde{R}_i \geq t)$, $Y_i^{(j)}(t) = I(T_i^{(j)} \geq t)$, $y^R(t) = Pr(\tilde{R} \geq t)$, $\lambda^R(t) = \lim_{\Delta \to 0} \Pr(\tilde{R} \in (t, t+\Delta), \delta = 0 | \tilde{R} \geq t)/\Delta$, $\Lambda^R(t) = \int_0^t \lambda^R(s) ds$, $M_i^R(t) = N_i^R(t) - \int_0^t Y_i^R(s) d\Lambda^R(s)$, and $\boldsymbol{\xi}_{S_C,i}(t) = G^R(t) \int_0^t y^R(s)^{-1} dM_i^R(s) - \{I(L_i \geq t) - Pr(L \geq t)\}$ . Define

$$\boldsymbol{\xi}_i(u) = \boldsymbol{\xi}_{1,i}(u) - \boldsymbol{\xi}_{2,i}(u), \tag{2.15}$$

where $\boldsymbol{\xi}_{1,i}(u) = \mathbf{Z}_i \{\sum_{j=1}^{\infty} \frac{1}{S_C(T_i^{(j)} | \boldsymbol{Z})} I(L_i < T_i^{(j)} \leq e^{\boldsymbol{Z_i}^\top \boldsymbol{\beta}_0^A(u)} \wedge \tilde{R}_i) - \int_0^u g(s) ds\}$, and $\boldsymbol{\xi}_{2,i}(u) = E_{(L,R,D,Z,\bar{T})} \{\mathbf{Z} \sum_{j=1}^{\infty} \boldsymbol{\xi}_{S_C,i}(T^{(j)}) I(L < T^{(j)} \leq e^{\boldsymbol{Z}^\top \boldsymbol{\beta}_0^A(u)} \wedge \tilde{R}) / S_C^2(T^{(j)})\}$, where $\bar{T} = (T^{(1)}, T^{(2)}, \ldots)$ and $E_{(L,R,D,\mathbf{Z},\bar{T})}$ means expectation w.r.t. $(L, R, D, \mathbf{Z}, \bar{T})$.

We also assume the following regularity conditions (also stated in the main manuscript):

(C1) There exists $v_R > 0$ such that $\Pr(R = v_R) > 0$ and $\Pr(R > v_R) = 0$. In addition, $\Pr(R > L) = 1$.

(C2) $\boldsymbol{Z}$ and $N(\tilde{R})$ are bounded.

(C3) (i) $\boldsymbol{\beta}_0^A(u)$ is Lipschitz continuous in $u \in [0, U]$; (ii) $\lambda_{\boldsymbol{Z}}^A(t) = E\{dN^*(t)\mathbf{Z}\}/dt$ is bounded above uniformly in $t$ and $\boldsymbol{Z}$.

(C4) For some $\rho_0 > 0$ and $c_0 > 0$, $\inf_{\boldsymbol{b} \in \mathcal{B}(\rho_0)} \text{eigmin} \boldsymbol{A}(\boldsymbol{b}) \geq c_0$, where $\mathcal{B}(\rho) = \{\boldsymbol{b} \in \mathbb{R}^{p+1} : \inf_{u \in [0,U]} ||\boldsymbol{b} - \boldsymbol{\beta}_0^A(u)|| \leq \rho\}$ and $\boldsymbol{A}(\boldsymbol{b}) = E\{\boldsymbol{Z}^{\otimes 2} \lambda_{\boldsymbol{Z}}^A(t) | \boldsymbol{Z}\}$. Here $||\cdot||$ is the Euclidean norm, and we define $\boldsymbol{u}^{\otimes 2} = \boldsymbol{u}\boldsymbol{u}^\top$ for a vector $\boldsymbol{u}$.

*Proof of Theorem 2.1:* Define $\boldsymbol{S}_n^G(\boldsymbol{b}, u) = n^{-1/2} \sum_{i=1}^n \boldsymbol{Z}_i \big[ \sum_{j=1}^{\infty} \frac{1}{S_C(T_i^{(j)})} I(L_i < T_i^{(j)} \leq e^{\boldsymbol{Z}_i^\top \boldsymbol{b}} \wedge \tilde{R}) - \int_0^u g(s) ds \big]$, $\widetilde{\boldsymbol{S}}_n(\boldsymbol{b}, u) = n^{-1/2} \sum_{i=1}^n \boldsymbol{Z}_i \big[ E\{N^*(e^{\boldsymbol{Z}_i^\top \boldsymbol{b}})\} - \int_0^u g(s) ds \big]$, and $\boldsymbol{\mu}(\boldsymbol{b}, u) = E\{n^{-1/2} \widetilde{\boldsymbol{S}}_n(\boldsymbol{b}, u)\}$. Hereafter we use $\sup_{\boldsymbol{b}}$ or $\sup_u$ to denote supremum taken over $\boldsymbol{b} \in \mathbb{R}^{p+1}$ or $u \in [\nu, U]$ respectively.

First, by condition (C1), for every $r > 0$, we have $\sup_{t < v^R} |\hat{S}_C(t) - S_C(t)| =$

$o(n^{-1/2+r}), a.s.$. Coupled with conditions (C2), it implies that

$$\sup_{\boldsymbol{b},u} ||n^{-1/2}S_n(\boldsymbol{b},u) - n^{-1/2}S_n^G(\boldsymbol{b},u)|| = o(n^{-1/2+r}),\ a.s.$$

Define $\mathcal{F} = \{\boldsymbol{Z}_i\big(\sum_{j=1}^{\infty}\frac{1}{S_C(T_i^{(j)})}I(L_i < T_i^{(j)} \leq e^{\boldsymbol{Z}_i^\top \boldsymbol{b}} \wedge \tilde{R}) - \int_0^u g(s)ds\big), \boldsymbol{b} \in \mathbb{R}^{p+1}, u \in [0,U]\}$. The function class $\mathcal{F}$ is Donsker (Vaart and Wellner, 1996) and thus Glivenko-Cantelli (van der Vaart and Wellner 1996) because the class of indicator function is Dnosker and both $\boldsymbol{Z_i}$ and $1/S_C(T_i^{(j)})$ are uniformly bounded. It then follows from the Clivenko-Cantelli theorem that $\sup_{\boldsymbol{b},u} ||n^{-1/2}S_n^G(\boldsymbol{b},u) - \boldsymbol{\mu}(\boldsymbol{b},u)|| = o(1), a.s.$. Therefore,

$$\sup_{\boldsymbol{b},u} ||n^{-1/2}S_n(\boldsymbol{b},u) - \boldsymbol{\mu}(\boldsymbol{b},u)|| = o(1), a.s. \tag{2.16}$$

Secondly, for any $\boldsymbol{w} \in \mathbb{R}^{p+1}$ satisfying $||\boldsymbol{w}||^2 = 1$, $\boldsymbol{w}^\top \boldsymbol{\mu}(\boldsymbol{\beta}_0^A(u) + \boldsymbol{w}\delta, u)$ is increasing in $\delta$. Then for $\delta > \rho_0$,

$$\boldsymbol{w}^\top[\boldsymbol{\mu}(\boldsymbol{\beta}_0^A(u) + \boldsymbol{w}\delta, u) - \boldsymbol{\mu}(\boldsymbol{\beta}_0^A(u), u)] \geq \boldsymbol{w}^\top[\boldsymbol{\mu}(\boldsymbol{\beta}_0^A(u) + \boldsymbol{w}\rho_0, u) - \boldsymbol{\mu}(\boldsymbol{\beta}_0^A(u), u)] \geq 0$$

From the Cauchy-Schwarz inequality, we have

$$\begin{aligned}
&||\boldsymbol{\mu}(\boldsymbol{\beta}_0^A(u) + \boldsymbol{w}\delta, u) - \boldsymbol{\mu}(\boldsymbol{\beta}_0^A(u), u)||^2 \cdot ||\boldsymbol{w}||^2 \\
&\quad \geq (\boldsymbol{w}^\top[\boldsymbol{\mu}(\boldsymbol{\beta}_0^A(u) + \boldsymbol{w}\delta, u) - \boldsymbol{\mu}(\boldsymbol{\beta}_0^A(u), u)])^2 \\
&\quad \geq (\boldsymbol{w}^\top[\boldsymbol{\mu}(\boldsymbol{\beta}_0^A(u) + \boldsymbol{w}\rho_0, u) - \boldsymbol{\mu}(\boldsymbol{\beta}_0^A(u), u)])^2 \\
&\quad = (\boldsymbol{w}^\top \boldsymbol{A}(\boldsymbol{\beta}_0^A(u) + \boldsymbol{w}\rho^*)\boldsymbol{w}\rho_0)^2 \geq c_0^2\rho_0^2
\end{aligned}$$

where $\rho^* \in [0, \rho_0]$. Since $\boldsymbol{\beta}_0^A(u) + \boldsymbol{w}\rho^* \in \mathcal{B}(\rho_0)$, the last above inequality follows from the Condition (C4). Therefore, we have $\inf_{\boldsymbol{b} \notin \mathcal{B}(\rho_0)} ||\boldsymbol{\mu}(\boldsymbol{b},u) - \boldsymbol{\mu}(\boldsymbol{\beta}_0^A(u), u)|| \geq c_0\rho_0$.

Next, simple algebraic manipulation shows that

$$\begin{aligned}\boldsymbol{\mu}(\widehat{\boldsymbol{\beta}}^A(u),u)-\boldsymbol{\mu}(\boldsymbol{\beta}_0^A(u),u) =& n^{-1/2}\boldsymbol{S}_n(\widehat{\boldsymbol{\beta}}^A(u),u)-\boldsymbol{\mu}(\boldsymbol{\beta}_0^A(u),u)\\ &-[n^{-1/2}\boldsymbol{S}_n(\widehat{\boldsymbol{\beta}}^A(u),u)-\boldsymbol{\mu}(\widehat{\boldsymbol{\beta}}^A(u),u)]\end{aligned}$$

By the definitions of $\widehat{\boldsymbol{\beta}}^A(\cdot)$ and $\boldsymbol{\beta}_0^A(\cdot)$, we have $\boldsymbol{S}_n(\widehat{\boldsymbol{\beta}}^A(u),u)=o(n^{-1/2})$, a.s., and $\boldsymbol{\mu}(\boldsymbol{\beta}_0^A(u),u)=0$. From (2.16), we have that

$$\boldsymbol{\mu}(\widehat{\boldsymbol{\beta}}^A(u),u)-\boldsymbol{\mu}(\boldsymbol{\beta}_0^A(u),u)=o(1),a.s. \tag{2.17}$$

and thus there exists $N_0>0$ such that for $n>N_0$, $\sup_u||\boldsymbol{\mu}(\widehat{\boldsymbol{\beta}}^A(u),u)-\boldsymbol{\mu}(\boldsymbol{\beta}_0^A(u),u)||<c_0^2\rho_0^2/2$ with probability 1. Consequently, $\{\widehat{\boldsymbol{\beta}}^A(u):u\in[0,U]\}\in\mathcal{B}(\rho_0)$ with probability 1 when $n$ is large enough. Note that

$$\sup_u||\widehat{\boldsymbol{\beta}}^A(u)-\boldsymbol{\beta}_0^A(u)||=\sup_u||\boldsymbol{A}(\check{\boldsymbol{\beta}}^A(u))^{-1}[\boldsymbol{\mu}(\widehat{\boldsymbol{\beta}}^A(u),u)-\boldsymbol{\mu}(\boldsymbol{\beta}_0^A(u),u)]||,$$

where $\check{\boldsymbol{\beta}}^A(u)$ is between $\widehat{\boldsymbol{\beta}}^A(u)$ and $\boldsymbol{\beta}_0^A(u)$. Therefore $\check{\boldsymbol{\beta}}^A(u)\in\mathcal{B}(\rho_0)$ for $n$ large enough. Uniform consistency follows from (2.17) and (C4).

**Lemma 2.1.** *For any positive sequence $\{d_n\}_{n=1}^\infty$ satisfying $d_n\to 0$,*

$$\begin{aligned}\lim_{n\to\infty}\sup_{\boldsymbol{b},\boldsymbol{b}'\in\mathcal{B}(\rho_0),\|\boldsymbol{b}-\boldsymbol{b}'\|\le d_n}\Bigg\|&n^{-1/2}\sum_{i=1}^n\boldsymbol{Z}_i\Big\{\sum_{j=1}^\infty\frac{1}{S_C(T_i^{(j)})}\big\{I(L_i<T_i^{(j)}\le e^{\boldsymbol{Z}_i^\top\boldsymbol{b}}\wedge\tilde{R}_i)\\ &-I(L_i<T_i^{(j)}\le e^{\boldsymbol{Z}_i^\top\boldsymbol{b}'}\wedge\tilde{R}_i)\big\}\Big\}-n^{1/2}\{\boldsymbol{\mu}(\boldsymbol{b},u)-\boldsymbol{\mu}(\boldsymbol{b}',u)\}\Bigg\|=0,a.s.\end{aligned}$$

*Proof of Lemma 2.1:* This lemma can be proved by using the results in Alexander (1984) and the arguments for theorem 1 of Lai and Yang (1988). The main step is to

show that there exists $G_0 > 0$ such that

$$\mathrm{Var}\left(\boldsymbol{Z}_i\Big\{\sum_{j=1}^{\infty}\frac{1}{S_C(T_i^{(j)})}\big\{I(L_i < T_i^{(j)} \le e^{\boldsymbol{Z}_i^\top \boldsymbol{b}} \wedge \tilde{R}_i) - I(L_i < T_i^{(j)} \le e^{\boldsymbol{Z}_i^\top \boldsymbol{b}'} \wedge \tilde{R}_i)\big\}\Big\}\right) \le G_0\|\boldsymbol{b} - \boldsymbol{b}'\|$$

This follows from the uniform boundedness of $\lambda_{\mathbf{Z}}^A(t)$ and the boundedness of $1/S_C(\cdot)$, $\mathbf{Z}$ and $\mathcal{B}(\rho_0)$, which are implied by conditions (C1)–(C4).

*Proof of Theorem 2.2:* From Pepe (1991), we have

$$\sup_{t\in[0,v^R)} ||n^{1/2}\{\widehat{G}^R(t) - G^R(t)\} - n^{-1/2}\sum_{i=1}^{n} G^R(t)\int_0^t y^R(s)^{-1}dM_i^R(s)|| \to_{a.s.} 0.$$

and therefore

$$\begin{aligned}\sup_{t\in[0,v^R)} ||n^{1/2}\{\widehat{S}_C(t) - S_C(t)\} - n^{-1/2}\sum_{i=1}^{n}\big[G^R(t)\int_0^t y^R(s)^{-1}dM_i^R(s) - \{I(L_i \ge t) \\ -Pr(L \ge t)\}\big]|| \\ = \sup_{t\in[0,v^R)} ||n^{1/2}\{\widehat{S}_C(t) - S_C(t)\} - n^{-1/2}\sum_{i=1}^{n}\boldsymbol{\xi}_{S_C,i}(t)|| \to_{a.s.} 0.\end{aligned}$$

In addition, we have $\sup_{t\in[0,v^R)} \|\widehat{S}_C(t) - S_C(t)\| \to_{a.s.} 0$. Based on these results, it follows from standard asymptotic arguments and an applicationsof the Glivenko-

Cantelli Theorem that

$$
\begin{aligned}
&\boldsymbol{S}_n\{\boldsymbol{\beta}_0^A(u), u\} \\
&\quad = \boldsymbol{S}_n^G\{\boldsymbol{\beta}_0^A(u), u\} + [\boldsymbol{S}_n\{\boldsymbol{\beta}_0^A(u), u\} - \boldsymbol{S}_n^G\{\boldsymbol{\beta}_0^A(u), u\}] \\
&\quad = n^{-1/2}\sum_{i=1}^n \boldsymbol{\xi}_{1,i}(u) - n^{-1/2}\sum_{i=1}^n \boldsymbol{Z}_i \left\{ \sum_{j=1}^{\infty} \frac{\hat{S}_C(T_i^{(j)}) - S_C(T_i^{(j)})}{\hat{S}_C(T_i^{(j)}) \cdot S_C(T_i^{(j)})} I(L_i < T_i^{(j)} \le e^{\boldsymbol{Z}_i^\top \boldsymbol{\beta}_0^A(u)} \wedge \tilde{R}_i) \right\} \\
&\qquad \times I(L_i < T_i^{(j)} \le e^{\boldsymbol{Z}_i^\top \boldsymbol{\beta}_0^A(u)} \wedge \tilde{R}_i) \Big\} \\
&\quad \approx n^{-1/2}\sum_{i=1}^n \boldsymbol{\xi}_{1,i}(u) \\
&\qquad - n^{-1}\sum_{i=1}^n \boldsymbol{Z}_i \Big\{ \sum_{j=1}^{\infty} \frac{n^{-1/2}\sum_{k=1}^n \boldsymbol{\xi}_{S_C,k}(T_i^{(j)})}{S_C^2(T_i^{(j)})} \times I(L_i < T_i^{(j)} \le e^{\boldsymbol{Z}_i^\top \boldsymbol{\beta}_0^A(u)} \wedge \tilde{R}_i) \Big\} \\
&\quad = n^{-1/2}\sum_{i=1}^n \boldsymbol{\xi}_{1,i}(u) \\
&\qquad - n^{-1/2}\sum_{k=1}^n \frac{1}{n}\sum_{i=1}^n \boldsymbol{Z}_i \Big\{ \sum_{j=1}^{\infty} \frac{\boldsymbol{\xi}_{S_C,k}(T_i^{(j)})}{S_C^2(T_i^{(j)})} \times I(L_i < T_i^{(j)} \le e^{\boldsymbol{Z}_i^\top \boldsymbol{\beta}_0^A(u)} \wedge \tilde{R}_i) \Big\} \\
&\quad = n^{-1/2}\sum_{i=1}^n \{\boldsymbol{\xi}_{1,i}(u) - \boldsymbol{\xi}_{2,i}(u)\} \doteq n^{-1/2}\sum_{i=1}^n \boldsymbol{\xi}_i(u),
\end{aligned}
$$

where $\approx$ denotes asymptotic equivalence uniformly in $u \in [\nu, U)$.

We can show that $\mathcal{F}^* = \{\boldsymbol{\xi}_{1,i}(u), u \in [0, U)\}$ and $\mathcal{F}^{**} = \{\boldsymbol{\xi}_{2,i}(u), u \in [0, U)\}$ are Donsker. First, given the Lipschitz continuity of $\boldsymbol{\beta}_0^A(\cdot)$, we can show $\mathcal{F}^*$ is Donsker by applying the similar arguments for $\mathcal{F}$ and using the fact that the Donsker property preserves under Lipschitz transformation. The Donsker property of $\mathcal{F}^{**}$ follows similarly. Since the Donsker property preserves under addition and subtraction, we can apply the Donsker theorem to $n^{-1/2}\sum_{i=1}^n \boldsymbol{\xi}_i(u)$. It then follows that $\boldsymbol{S}_n(\boldsymbol{\beta}_0^A(u), u)$ converges weakly to a mean zero Gaussian process with covariance matrix

$$
\boldsymbol{\Sigma}(u', u) = E\{\boldsymbol{\xi}(u'), \boldsymbol{\xi}(u)\} \tag{2.18}
$$

Next, simple algebric manipulations show that $\boldsymbol{S}_n(\hat{\boldsymbol{\beta}}^A(u), u) - \boldsymbol{S}_n(\boldsymbol{\beta}_0^A(u), u) =$

$(\mathbf{I}) + (\mathbf{II})$, where

$$(\mathbf{I}) = n^{-1/2}\sum_{i=1}^{n}\boldsymbol{Z}_i\Big\{\sum_{j=1}^{\infty}\frac{1}{S_C(T_i^{(j)})}\big\{I(L_i < T_i^{(j)} \le e^{\boldsymbol{Z}_i^\top\hat{\boldsymbol{\beta}}^A(u)}\wedge\tilde{R}_i) - I(L_i < T_i^{(j)} \le e^{\boldsymbol{Z}_i^\top\boldsymbol{\beta}_0^A(u)}\wedge\tilde{R}_i)\big\}\Big\}$$

and

$$(\mathbf{II}) = n^{-1/2}\sum_{i=1}^{n}\boldsymbol{Z}_i\Big\{\sum_{j=1}^{\infty}\big\{\frac{1}{\hat{S}_C(T_i^{(j)})} - \frac{1}{S_C(T_i^{(j)})}\big\}\big\{I(L_i < T_i^{(j)} \le e^{\boldsymbol{Z}_i^\top\hat{\boldsymbol{\beta}}^A(u)} \wedge \tilde{R}_i) - I(L_i < T_i^{(j)} \le e^{\boldsymbol{Z}_i^\top\boldsymbol{\beta}_0^A(u)} \wedge \tilde{R}_i)\big\}\Big\}$$

From Lemma 1 and the uniform consistency of $\hat{\boldsymbol{\beta}}^A(u)$, we have $(I) \approx n^{1/2}[\boldsymbol{\mu}\{\hat{\boldsymbol{\beta}}^A(u), u\} - \boldsymbol{\mu}\{\boldsymbol{\beta}_0^A(u), u\}]$. Since $\sup_i\{\hat{S}_C(T_i^{(j)})^{-1} - S_C(T_i^{(j)})^{-1}\} = o_p(1)$, we can see that $\boldsymbol{S}_n(\hat{\boldsymbol{\beta}}^A(u), u) - \boldsymbol{S}_n(\boldsymbol{\beta}_0^A(u), u)$ is dominated by $(\mathbf{I})$. Taylor expansions of $\mu(\boldsymbol{b})$ around $\boldsymbol{b} = \boldsymbol{\beta}_0^A(u)$ along with the facts that $\hat{\boldsymbol{\beta}}^A(u)$ uniformly converges to $\boldsymbol{\beta}_0^A(u)$ gives that

$$\boldsymbol{S}_n(\hat{\boldsymbol{\beta}}^A(u), u) - \boldsymbol{S}_n(\boldsymbol{\beta}_0^A(u), u) = [\boldsymbol{A}\{\boldsymbol{\beta}_0^A(u)\} + \epsilon_n(u)] \cdot n^{1/2}\{\hat{\boldsymbol{\beta}}^A(u) - \boldsymbol{\beta}_0^A(u)\}$$

where $sup_u\|\epsilon_n(u)\| \to 0$. Given $\boldsymbol{S}_n(\hat{\boldsymbol{\beta}}^A(u), u) \approx 0$, this further implies that $n^{1/2}\{\hat{\boldsymbol{\beta}}_0^A(u) - \boldsymbol{\beta}_0^A(u)\} = -\boldsymbol{A}\{\boldsymbol{\beta}_0^A(u)\}^{-1}\boldsymbol{S}_n(\boldsymbol{\beta}_0^A(u), u) + \epsilon_n^*(u)$, where $sup_u\|\epsilon_n^*(u)\| \to 0$. It follows that

$$n^{1/2}\{\hat{\boldsymbol{\beta}}_0^A(u) - \boldsymbol{\beta}_0^A(u)\} \approx n^{-1/2}\sum_{i=1}^{n}\boldsymbol{A}\{\boldsymbol{\beta}_0^A(u)\}^{-1}\boldsymbol{\xi}_i(u)$$

It follows from the Donsker Theorem (Vaart and Wellner, 1996) that $n^{1/2}\{\hat{\boldsymbol{\beta}}_0(u) - \boldsymbol{\beta}_0(u)\}$ converges weakly to a mean zero Gaussian process for $u \in [0, U]$ with covariance

$$\boldsymbol{\Phi}(u', u) = \boldsymbol{A}\{\boldsymbol{\beta}_0^A(u')\}^{-1}E\{\boldsymbol{\xi}(u')\boldsymbol{\xi}(u)^\top\}\boldsymbol{A}\{\boldsymbol{\beta}_0^A(u)\}^{-1}. \tag{2.19}$$

# Chapter 3

# Estimation of Complier Causal Quantile Effects with a Binary Instrumental Variable and Censored Data

## 3.1 Potential Outcomes Framework and Assumptions

Let $D$ and $V$ denote a binary treatment indicator and a binary IV respectively. Let $D_v$ denote the potential treatment selection given $V = v$, $T_d$ denote the potential survival time given $D = d$, and $T_{vd}$ denote the potential survival time with $V = v$ and $D = d$, where $v = 0, 1$ and $d = 0, 1$. Following the terminology of Angrist et al. (1996a), subjects are classified into four latent subgroups: *compliers* (i.e. $D_1 > D_0$), *always takers* (i.e. $D_1 = D_0 = 1$), *never takers* (i.e. $D_1 = D_0 = 0$), and *defiers* (i.e. $D_1 < D_0$). Let $\mathbf{X}$ represent the $d-$dimensional vector of covariates, and $\perp\!\!\!\perp$ represent statistical independence.

We adopt the following standard IV assumptions:

(A1) Independence of IV: $(T_{00}, T_{01}, T_{10}, T_{11}, D_0, D_1) \perp\!\!\!\perp V|\mathbf{X}$.

(A2) Exclusion of IV: $P(T_{1d} = T_{0d}|\mathbf{X}) = 1$ for $d = 0, 1$.

(A3) First stage: $0 < P(V = 1|\mathbf{X}) < 1$ and $P(D_1 = 1|\mathbf{X}) > P(D_0 = 1|\mathbf{X})$.

(A4) Monotonicity: $P(D_1 \geq D_0|\mathbf{X}) = 1$.

The potential outcome framework and assumptions described above are rather standard in IV literature (Abadie et al., 2002, for example). By assumption (A1), the IV, $V$, mimics a random assignment conditional on $\mathbf{X}$. Assumption (A2) requires that the variation in $V$ only affects the potential survival time through its effects on the treatment $D$. This assumption also implies $T_0 = T_{00} = T_{10}$ and $T_1 = T_{01} = T_{11}$. Assumption (A3) guarantees that $D$ and $V$ are correlated conditional on $\mathbf{X}$, and that each subject can have $V = 0$ or $V = 1$ with a non-zero probability conditional on $\mathbf{X}$. Assumptions (A4) excludes the existence of *defiers*.

## 3.2 The Proposed Method

### 3.2.1 A causal censored quantile regression model

In this work, we study a complier quantile regression model specified as

$$Q_{T_d}(\tau|\mathbf{X}, D_1 > D_0) = \exp\{\beta_d(\tau)d + \boldsymbol{\beta}_x(\tau)^T X\},\ \tau \in (0, \tau_U),\ d = 0, 1 \tag{3.1}$$

where $Q_{T_d}(\tau|\mathbf{X}, D_1 > D_0) = \inf\{t : \Pr(T_d \leq t|\mathbf{X}, D_1 > D_0) \geq \tau\}$, and $0 < \tau_U < 1$. It is easy to show that the treatment coefficient $\beta_d(\tau)$ in model (3.1) satisfies

$$\beta_d(\tau) = Q_{\log T_1}(\tau|\mathbf{X}, D_1 > D_0) - Q_{\log T_0}(\tau|\mathbf{X}, D_1 > D_0).$$

This implies a causal interpretation that $\beta_d(\tau)$ represents the difference in complier's $\tau$-th quantile of the potential survival time (in the logarithm scale) between the scenario where the treatment is received versus the scenario where the treatment is not received, given the covariates in $\mathbf{X}$. By this interpretation, we shall refer $\beta_d(\tau)$ to the complier $\tau$-th quantile causal effect of treatment, denoted by CQCE($\tau$). Compared to the commonly used complier average causal effect (CACE), namely $E(T_1 - T_0|\mathbf{X}, D_1 > D_0)$, CQCE($\tau$) is a more flexible venue to depict the causal treatment effect in compliers, a subgroup that does not have any treatment preference and serves to reveal the treatment efficacy of a substantive interest. In the presence of censoring to $T$, CQCE($\tau$) can be identifiable when CACE is not.

Consider a special case, one-sided compliance case, where subjects with $V = 0$ have no access to the treatment (i.e. $\Pr(D_0 = 0|X) = 1$). The one-sided compliance assumption is reasonable in many practical settings. For example, in a randomized clinical trial comparing a new treatment versus placebo, where IV is chosen as the random group assignment, the one-sided compliance means that patients assigned to the placebo group has no access to the new treatment. In an observational study

comparing a new inpatient treatment versus standard care, where IV is chosen as the approval status of the new treatment at hospital admission, this assumption means that patients have no access. We can further show that

$$\beta_d(\tau) = Q_{\log T_1}(\tau|\mathbf{X}, D=1) - Q_{\log T_0}(\tau|\mathbf{X}, D=1);$$

see Proposition 3.1 in Section 3.5.1 of the Appendix C. This means, CQTE($\tau$) in the one-sided compliance case has an alternative simpler interpretation as the quantile causal treatment effect for the treated population (i.e. $D=1$).

To address the interest in $\beta_d(\tau)$, our key finding is that under assumptions (A1) and (A2), model (3.1) is equivalent to

$$Q_T(\tau|D, \mathbf{X}, D_1 > D_0) = \exp\{\beta_d(\tau)D + \boldsymbol{\beta}_X^T X\}, \tau \in (0,1), \tag{3.2}$$

where $T = D \times T_1 + (1-D) \times T_0$ and $Q_{\log T}(\tau|D, \mathbf{X}, D_1 > D_0) = \inf\{t : \Pr(\log T \leq t|D, \mathbf{X}, D_1 > D_0) \geq \tau\}$. The justification for this result is provided in Proposition 3.2 in Section 3.5.1 of the Appendix C. Note that under model 3.2,

$\beta_d(\tau)$ is linked with the conditional quantile of $T$, the observed survival time (in the absence of censoring), rather than the potential survival time $T_d$. Hence, model 3.2 provides a more convenient venue to estimate $\beta_d(\tau)$ than model 3.1.

### 3.2.2 Estimation procedure with randomly censored data

Suppose the event time of interest $T$ is subject to right censoring time by $C$. Define $W = \min\{T, C\}$ and $\delta = I(T \leq C)$. The observed data consist of $n$ i.i.d. replicates of $\mathbf{O} \doteq (W, \delta, D, \mathbf{X}, V)$ denoted by $\mathbf{O}_i \doteq \{(W_i, \delta_i, D_i, \mathbf{X}_i, V_i)\}_{i=1}^n$. We adopt the following censoring assumption:

(A5) $C$ is independent of $T$ given $(V, D, \mathbf{X})$, and $C$ is independent of $V$ given $\mathbf{X}$.

By this assumption, $T$ is subject to the standard conditionally independent censoring. The conditional independence between $C$ and $V$ is consistent with the independence of IV assumption, (A1).

Let $\mathbf{U} = (W, \delta, D, \mathbf{X})$, $\mathbf{Z} = (D, \mathbf{X})$ and $\boldsymbol{\beta} = (\beta_d, \boldsymbol{\beta}_x)$. Define $F_T(t|\mathbf{Z}) = \Pr(T \leq t|\mathbf{Z})$, $\Lambda_T(t|\mathbf{Z}) = -\log\{1-F_T(t|\mathbf{Z})\}$, $N(t) = I(W \leq t, \delta = 1)$ and $M(t) = N(t)-\Lambda_T(t\wedge W|\mathbf{Z})$. We use the subscript $_i$ to denote sample analogues, and $\boldsymbol{\beta}_0(\cdot)$ to represent the true coefficient function in model (3.1).

To estimate $\boldsymbol{\beta}_0(\cdot)$, a key observation from applying the result of Peng and Huang (2008) for censored quantile regression is that under model (3.2),

$$E\left\{\mathbf{Z}\left(N(\exp\{\boldsymbol{\beta}_0(\tau)^T\mathbf{Z}) - \int_0^\tau I[W \geq \exp\{\boldsymbol{\beta}_0(u)^T\mathbf{Z}\}]dH(u)\right)\middle| D_1 > D_0\right\} = 0.$$

This immediately implies

$$E\left\{I(D_1 > D_0)\mathbf{Z}\left(N(\exp\{\boldsymbol{\beta}_0(\tau)^T\mathbf{Z}) - \int_0^\tau I[W \geq \exp\{\boldsymbol{\beta}_0(u)^T\mathbf{Z}\}]dH(u)\right)\right\} = 0. \tag{3.3}$$

Equation (3.3) cannot be directly translated into an estimating equation because $D_1$ and $D_0$ cannot be observed at the same time and thus the complier indicator $I(D_1 > D_0)$ is not observable. Following the spirit of conditional score principal for handling measurement errors (Stefanski and Carroll, 1987), we consider utilizing the conditional expectation of $I(D_1 > D_0)\mathbf{Z}\Big(N(\exp(\boldsymbol{\beta}_0(\tau)^\top\mathbf{Z}) - \int_0^\tau I[W \geq \exp\{\boldsymbol{\beta}_0(u)^\top\mathbf{Z}\}]dH(u)\Big)$ given $\mathbf{U}$ to construct an estimating equation for $\boldsymbol{\beta}_0(\cdot)$. Here, the observed $\mathbf{U}$ plays the same role as the complete and sufficient statistic used by Stefanski and Carroll (1987) to transform their original score function that involves unknown covariates to a conditional score function free of unknown covariates. With $\kappa_v(\mathbf{U}) = \Pr(D_1 > D_0|\mathbf{U})$, it is easy to see that the resulting "conditional score" equals $\kappa_v(\mathbf{U})\mathbf{Z}\Big(N(\exp(\boldsymbol{\beta}_0(\tau)^\top\mathbf{Z}) - \int_0^\tau I[W \geq \exp\{\boldsymbol{\beta}_0(u)^\top\mathbf{Z}\}]dH(u)\Big)$, which does not

depend on the unobservable $D_0$ and $D_1$, and has expectation zero, i.e.

$$E\left\{\kappa_v(\mathbf{U})\mathbf{Z}\left(N(\exp\{\boldsymbol{\beta}_0(\tau)^T\mathbf{Z}) - \int_0^\tau I[W \geq \exp\{\boldsymbol{\beta}_0(u)^T\mathbf{Z}\}]dH(u)\right)\right\} = 0, \quad (3.4)$$

Equation 3.4 suggests a simple weighting scheme that transforms Peng and Huang (2008)'s estimating equation to produce a valid estimate for CQCE($\tau$). The weight $\kappa_v(\mathbf{U})$, by its definition, represents an observed surrogate of the unobseravable complier indicator $I(D_1 > D_0)$.

With further manipulations, we can show that

$$\kappa_v(\mathbf{U}) = 1 - \frac{D(1-v(\mathbf{U}))}{1-\pi(\mathbf{X})} - \frac{(1-D)v(\mathbf{U})}{\pi(\mathbf{X})}, \quad (3.5)$$

where $v(\mathbf{U}) = \Pr(V = 1|\mathbf{U})$ and $\pi(\mathbf{X}) = \Pr(V = 1|\mathbf{X})$; see Proposition 3.3 in Section 3.5.1 of the Appendix C.

Suppose $\kappa_{v,i} \doteq \kappa_v(\mathbf{U}_i)$ is known, an estimating equation for $\boldsymbol{\beta}_0(\cdot)$ is given by

$$n^{-1/2}\sum_{i=1}^n \kappa_{v,i}\mathbf{Z}_i\left(N_i[\exp\{\boldsymbol{\beta}(\tau)^T\mathbf{Z}_i\}] - \int_0^\tau I[W_i \geq \exp\{\boldsymbol{\beta}^T(u)\mathbf{Z}_i\}]dH(u)\right) = 0. \quad (3.6)$$

In practice, $\kappa_{v,i}$ is usually unknown. To estimate $\kappa_{v,i}$, parametric modeling of the IV, such as logistic regression modeling of $V$ given $\mathbf{U}$ or $\mathbf{X}$, can be used. Such a procedure is easy to implement but can induce considerable bias if the assumed models are misspecified, as suggested by the simulation studies in Section 3.3.1. To overcome this issue, we propose nonparametric estimation of $\kappa_{v,i}$. Specifically, let $\mathbf{Y} = (W, \mathbf{X}^\top)^\top$, and define $v_{ij}(\mathbf{Y}) = \Pr(V = 1|\mathbf{Y}, \delta = i, D = j)$ and $v(\mathbf{U}_i) = \sum_{j=0}^1\sum_{k=0}^1 I(\delta_i = j, D_i = k)v_{jk}(\mathbf{Y}_i)$. Let $\mathcal{K}^*_{\sigma_1}(\mathbf{u})$ and $\mathcal{K}^{**}_{\sigma_2}(\mathbf{u})$ denote two kernel functions with bandwidths $\sigma_1$ and $\sigma_2$ respectively, subject to the technical conditions (C7)-(C8) in Section 3.2.3. For example, these kernel functions can be the multiplicative second order Epanechnikov kernels, $\mathcal{K}^*_{\sigma_1}(\mathbf{u}) = \prod_{i=1}^q \frac{3}{4}\frac{1}{\sigma_1}(1-\frac{u_i}{\sigma_1})^2 I(|\frac{u_i}{\sigma_1}| \leq 1)$ and

$\mathcal{K}^{**}_{\sigma_2}(\mathbf{u}) = \prod_{i=1}^{q} \frac{3}{4}\frac{1}{\sigma_2}(1-\frac{u_i}{\sigma_2})^2 I(|\frac{u_i}{\sigma_2}| \leq 1)$, with $\mathbf{u} = (u_1, \cdots, u_q)^\top$. Asymmetric kernels may be considered to adjust for tail problems when needed. When all components of $X$ are continuous, we can estimate $\pi(\mathbf{x})$ and $v_{jk}(\mathbf{u})$ respectively by

$$\hat{\pi}(\mathbf{x}) = \frac{\sum_{i=1}^{n} \mathcal{K}^{*}_{\sigma_1}(\mathbf{x}-\mathbf{X}_i)V_i}{\sum_{i=1}^{n} \mathcal{K}^{*}_{\sigma_1}(\mathbf{x}-\mathbf{X}_i)}, \quad \text{and} \quad \hat{v}_{jk}(\mathbf{y}) = \frac{\sum_{i=1}^{n} I(\delta_i = j, D_i = k)\mathcal{K}^{**}_{\sigma_2}(\mathbf{y}-\mathbf{Y}_i)V_i}{\sum_{i=1}^{n} I(\delta_i = j, D_i = k)\mathcal{K}^{**}_{\sigma_2}(\mathbf{y}-\mathbf{Y}_i)}.$$

Subsequently we estimate $v(\mathbf{U}_i)$ by $\hat{v}(\mathbf{U}_i) = \sum_{j=0}^{1}\sum_{k=0}^{1} I(\delta_i = j, D_i = k)\hat{v}_{jk}(\mathbf{U}_i)$. When $X$ involves discrete components, we can obtain kernel estimates stratified on the discrete covariates and then combine them into nonparametric estimates for $\pi(\mathbf{X}_i)$ and $v(\mathbf{U}_i)$.

To determine the bandwidths, $\sigma_{1,n}$ and $\sigma_{2,n}$, we can use cross-validation. For example, let $\hat{\pi}_\sigma(\mathbf{X})$ denote the estimator of $\pi(\mathbf{X})$ obtained from the training dataset with bandwidth $\sigma$. We may choose $\sigma_{1,n}$ as the $\sigma$ that minimizes $\sum_{i\in\text{test set}} |V_i - \hat{\pi}_\sigma(\mathbf{X}_i)|$ or some other appropriate criterion. The bandwidth $\sigma_{2,n}$ can be selected in a similar way.

A non-parametric estimator of $\kappa_{v,i}$ is then given by

$$\hat{\kappa}_{v,i} = 1 - \frac{D_i(1-\hat{v}(\mathbf{U}_i))}{1-\hat{\pi}(\mathbf{X}_i)} - \frac{(1-D_i)\hat{v}(\mathbf{U}_i)}{\hat{\pi}(\mathbf{X}_i)}.$$

Note that $\kappa_{v,i} = \Pr(D_1 > D_0|\mathbf{U}_i)$ and thus should be bounded between 0 and 1, we propose to further adjust $\hat{\kappa}_{v,i}$ by $\tilde{\kappa}_v = \min(\max(\hat{\kappa}_{v,i}, c_{l,n}), c_{u,n})$, where $c_{l,n}$ and $c_{u,n}$ are positive constants approaching 0 and 1 respectively as $n$ increases.

Replacing $\kappa_{v,i}$ in equation (3.6) by $\tilde{\kappa}_{v,i}$ , the proposed estimating equation is given by

$$n^{1/2} S_n(\boldsymbol{\beta}, \tau) = 0 \tag{3.7}$$

where

$$S_n(\boldsymbol{\beta}, \tau) = n^{-1} \sum_{i=1}^{n} \tilde{\kappa}_{v,i} \mathbf{Z}_i \left( N_i[\exp\{\boldsymbol{\beta}^T(\tau)\mathbf{Z}_i\}] - \int_0^{\tau} I[W_i \geq \exp\{\boldsymbol{\beta}^T(u)\mathbf{Z}_i\}] dH(u) \right).$$

It is important to note that equation (3.7) can be viewed as a weighted version of Peng and Huang (2008)'s equation for standard censored quantile regression. The specific procedure to carry out the proposed estimation procedure is described as follows:

Step 1 Specify a fine $\tau$-grid $S_{L(n)} = \{0 = \tau_0 < \tau_1 < \cdots < \tau_{L(n)} = \tau_U < 1\}$, and define the proposed estimator $\hat{\boldsymbol{\beta}}(\tau)$ as a piecewise-constant right-continuous function that only jumps at the grid points on $S_{L(n)}$.

Step 2 Calculate $\hat{\pi}(\mathbf{X}_i)$ and $\hat{v}(\mathbf{U}_i)$ with bandwidth $\sigma_1$ and $\sigma_2$ selected by cross-validation, and then compute $\tilde{\kappa}_{v,i}, i = 1, \ldots, n$.

Step 3 Obtain $\hat{\boldsymbol{\beta}}(\tau_k), k = 1, \ldots, L(n)$ based on equation (3.7). This can be implemented by using the R function $crq : fit : peng()$ or the SAS procedure QUANTLIFE, with the time-to-event outcomes, covariates, and weights specified by $(W_i, \delta_i), \mathbf{Z}_i$ and $\tilde{\kappa_{v,i}}$ respectively.

### 3.2.3 Asymptotic properties

In this section, we establish the uniform consistency and weakly convergence of the proposed estimator $\hat{\boldsymbol{\beta}}(\cdot)$. Beyond Peng and Huang (2008)'s results on a similar estimating function with the fixed weight 1, we need to properly account for the variation induced by the estimated weight $\tilde{\kappa}_{v,i}$ in the stochastic integral process. This may considerably complicate the derivation and the justification of the limit process of the proposed estimator. The nonparametric nature of $\tilde{\kappa}_{v,i}$ adds another layer of complexity, as the asymptotic properties of $\tilde{\kappa}_{v,i}$ are rather delicate in comparison to parametric

modeling of the weight, owing to the estimated weight converging at a slower rate than usual parametric weight estimates.

Before presenting the regularity conditions used in the asymptotic properties, we firstly introduce some new notation to simplify the presentation.

Denote $F(t|\mathbf{Z}) = Pr(W \leq t|\mathbf{Z}, D_1 > D_0)$, $\bar{F}(t|\mathbf{Z}) = 1 - F(t|\mathbf{Z})$,$\tilde{F}(t|\mathbf{Z}) = Pr(W \leq t, \delta = 1|\mathbf{Z}, D_1 > D_0)$. Let $f(t|\mathbf{Z}) = dF(t|\mathbf{Z})/dt$, $\bar{f}(t|\mathbf{Z}) = d\bar{F}(t|\mathbf{Z})/dt$ and $\tilde{f}(t|\mathbf{Z}) = d\tilde{F}(t|\mathbf{Z})/dt$. Define $\boldsymbol{\mu}_c(\boldsymbol{b}) = E\{\kappa_v \mathbf{Z} N(\exp \mathbf{Z}^\top \boldsymbol{b})\} = E\{I(D_1 > D_0)\mathbf{Z} N(\exp \mathbf{Z}^\top \boldsymbol{b})\}$, $\tilde{\boldsymbol{\mu}}_c(\boldsymbol{b}) = E[\kappa_v \mathbf{Z} I\{W \geq \exp(\mathbf{Z}^\top \boldsymbol{b})\}] = E[I(D_1 > D_0)\mathbf{Z} I\{W \geq \exp(\mathbf{Z}^\top \boldsymbol{b})\}]$, $\boldsymbol{\nu}_n(\boldsymbol{b}) = n^{-1}\sum_{i=1}^n \kappa_{v,i}\mathbf{Z}_i N_i\{\exp(\mathbf{Z}_i^\top \boldsymbol{b})\} - \boldsymbol{\mu}_c(\boldsymbol{b})$ and $\tilde{\boldsymbol{\nu}}_n(\boldsymbol{b}) = n^{-1}\sum_{i=1}^n \kappa_{v,i}\mathbf{Z}_i I\{W_i \geq \exp(\mathbf{Z}_i^\top \boldsymbol{b})\} - \tilde{\boldsymbol{\mu}}_c(\boldsymbol{b})$. Let $\mathbf{B}(\mathbf{b}) = \mathrm{E}[I(D_1 > D_0)\mathbf{Z}^{\otimes 2}\tilde{f}\{\exp(\boldsymbol{Z}^\top \boldsymbol{b}|\mathbf{Z})\}\exp(\boldsymbol{Z}^\top \boldsymbol{b})]$, $\mathbf{J}(\mathbf{b}) = \mathrm{E}[I(D_1 > D_0)\mathbf{Z}^{\otimes 2}\bar{f}\{\exp(\boldsymbol{Z}^\top \boldsymbol{b}|\mathbf{Z})\}\exp(\boldsymbol{Z}^\top \boldsymbol{b})]$. For $d > 0$, define $\mathcal{B}(d) = \{\boldsymbol{b} \in \mathbb{R}^p : \inf_{\tau \in (0,\tau_U]} \|\boldsymbol{\mu}_c(\boldsymbol{b}) - \boldsymbol{\mu}_c(\boldsymbol{\beta}_0(\tau))\| \leq d\}$. Besides, we define $\boldsymbol{\alpha}_0(\tau) = \boldsymbol{\mu}_c\{\boldsymbol{\beta}_0(\tau)\}$ and $\hat{\boldsymbol{\alpha}}(\tau) = \boldsymbol{\mu}_c(\hat{\boldsymbol{\beta}}(\tau))$. Define $\mathcal{A}(d) = \{\boldsymbol{\mu}_c(\boldsymbol{b}) : \boldsymbol{b} \in \mathcal{B}(d)\}$.

For simplify the notation, let $v, \hat{v}, \pi, \hat{\pi}$ denote $v(\mathbf{U}), \hat{v}(\mathbf{U}), \pi(\mathbf{X})$ and $\hat{\pi}(\mathbf{X})$ respectively. Similarly, let $v_i, \hat{v}_i, \pi_i, \hat{\pi}_i$ denote $v(\mathbf{U}_i), \hat{v}(\mathbf{U}_i), \pi(\mathbf{X}_i)$ and $\hat{\pi}(\mathbf{X}_i)$ respectively. Moreover, let $f_{\mathbf{X}}(\cdot)$ denote the density function of $\mathbf{X}$, and let $f_{\mathbf{Y}}^{ij}(\cdot)$ denote the density function of $\mathbf{Y} = (W, \mathbf{X}^\top)^\top$ conditional on $\delta = i$ and $D = j$ for $i, j \in \{0, 1\}$. Define $\mathbf{m}(\mathbf{U}, \tau) = \mathbf{Z}[N(\exp\{\mathbf{Z}\boldsymbol{\beta}_0(\tau)\}) - \int_v^\tau I(W \geq \exp\{\mathbf{Z}^\top \boldsymbol{\beta}_0(u)\})dH(u)]$, $\mathbf{m}_{ij}(\mathbf{U}, \tau) = I(\delta = i, D = j)\mathbf{m}(\mathbf{U}, \tau)$, $\mathbf{a}(\mathbf{U}, \tau) = \mathbf{m}(\mathbf{U}, \tau) \cdot \left(\frac{(1-D)\cdot v}{\pi^2} - \frac{D\cdot(1-v)}{(1-\pi)^2}\right)$, and

$$\mathbf{H}(\mathbf{X}, \tau) = E\{\mathbf{a}(\mathbf{U}, \tau)|\mathbf{X}\} = \mathrm{E}\left[\mathbf{m}(\mathbf{U}, \tau) \cdot (-\frac{D \cdot (1 - V)}{(1 - \pi)^2} + \frac{(1 - D) \cdot V}{(1 - \pi)^2})|\mathbf{X}\right].$$

Let $\mathcal{X}$ and $\mathcal{U}$ denote the support of $\mathbf{X}$ and the support of $\mathbf{U}$ respectively. For a vector $\mathbf{v}$, we let $v^{(j)}$ denote the $j$-th component of $\mathbf{v}$, and $\|v\|$ denote the Euclidean norm of $\mathbf{v}$. Let $o_I(a_n)$ (or $O_I(a_n)$) denote a function of $\tau$ with its supremum over $\tau \in I$ being $o_p(a_n)$ (or $O_p(a_n)$). Without loss of generality, we assume that $\tau_1, \cdots, \tau_{L_1}$ are equally spaced between 0 and $\tau_U$. Let $a_n = \|S_{L(n)}\|$ and $b_n = a_n/(1 - \tau_U)$.

We assume the following regularity conditions:

(C1) (a) Each discrete component of $\mathbf{X}$ takes on finite values, and (b) conditional on $D, \delta$ and the discrete components of $\mathbf{X}$, $(W, \mathbf{X}_c)$ has a support equal to a product of compact intervals and has a density at least third order continuously differentiable and bounded away from zero and infinity. Here $\mathbf{X}_c$ denotes the subvector of $\mathbf{X}$ corresponding to continuous covariates.

(C2) (a) Each component of $E(I(D_1 > D_0)\mathbf{Z}N[\exp\{\mathbf{Z}^\top \boldsymbol{\beta}_0(\tau)\}])$ is a Lipschitz function of $\tau$, and (b) $\tilde{f}(t|\mathbf{Z})$ and $f(t|\mathbf{Z})$ are bounded uniformly in $t$ and $\mathbf{Z}$; (c) there exists $C_H > 0$ such that $\|H(\mathbf{X}, \tau) - H(\mathbf{X}, \tau')\| \leq C_H|\tau - \tau'|$ for all $\mathbf{x} \in \mathcal{X}$ and $0 < \tau < \tau' < \tau_U$.

(C3) (a) $\tilde{f}\{\exp(\mathbf{Z}^\top \boldsymbol{b})|\mathbf{Z}\} > 0$ for all $\boldsymbol{b} \in \mathcal{B}(d_0)$; (b) $\mathrm{E}(I(D_1 > D_0)\mathbf{Z}^{\otimes 2}) > 0$; (c) each component of $\mathrm{E}[I(D_1 > D_0)\mathbf{Z}^{\otimes 2}\bar{f}\{\exp(\boldsymbol{Z}^\top \boldsymbol{b}|\mathbf{Z})\}\exp(\boldsymbol{Z}^\top \boldsymbol{b})] \times (\mathrm{E}[I(D_1 > D_0)\mathbf{Z}^{\otimes 2}\tilde{f}\{\exp(\boldsymbol{Z}^\top \boldsymbol{b}|\mathbf{Z})\}\exp(\boldsymbol{Z}^\top \boldsymbol{b})])^{-1}$ is uniformly bounded in $\boldsymbol{b} \in \mathcal{B}(d_0)$; (d) each component of $\exp(\mathbf{Z}^\top \mathbf{b})\mathbf{B}(\mathbf{b})^{-1}$ is uniformly bounded in $\mathbf{b} \in \mathcal{B}(d_0)$ and $\mathbf{Z}$.

(C4) For some $c_0 > 0$, $\inf_{\mathbf{b} \in \mathcal{B}(d_0)} \mathrm{eigmin}\mathrm{E}(I(D_1 > D_0)\mathbf{Z}^{\otimes 2}\tilde{f}[\exp\{\boldsymbol{Z}^\top \boldsymbol{\beta}_0(\tau)|\mathbf{Z}\}]\exp\{\boldsymbol{Z}^\top \boldsymbol{\beta}_0(\tau)\}) > c_0$ for any $v \in (0, \tau_U]$, where $\mathrm{eigmin}(\cdot)$ denotes the minimum eigenvalue of a matrix.

(C5) (a) For some $c > 0$,$\kappa_v(\mathbf{U}) > c$ almost surely; (b) for some $0 < c_1 < c_2 < 1$, $c_1 < \pi(\mathbf{X}) < c_2$ almost surely.

(C6) $c_{l,n} = o(n^{-1/2})$ and $1 - c_{u,n} = o(n^{-1/2})$

(C7) (a) There is a positive integer $\Delta$, such that $\mathcal{K}^*_{\sigma_1}(\mathbf{u})$ and $\mathcal{K}^{**}_{\sigma_2}(\mathbf{u})$ are differentiable of order $\Delta$ and the derivatives of order $\Delta$ are Lipschitz in a bounded support. $\mathcal{K}^*_{\sigma_1}(\mathbf{u})$ and $\mathcal{K}^{**}_{\sigma_2}(\mathbf{u})$ have bounded support; (b) $\mathcal{K}^*_{\sigma_k}(\mathbf{u}) = 1$ for $k = 1, 2$; (c) for some positive integers $s_1$ and $s_2$, $\int \mathcal{K}^*_{\sigma_k}(\mathbf{u})[\otimes_{l=1}^{j}\mathbf{u}]d\mathbf{u} = 0$ for all $j < s_k$, where $k = 1, 2$, and $\otimes_{l=1}^{j}\mathbf{u}$ stands for executing $j$ times Kronecker product on $\mathbf{u}$.

(C8) (a) $v(\cdot)$ and $\pi(cdot)$ are at least $p$-th order continuous differentiable; (b)$p \geq s_i$; (c) $n\sigma_{i,n}^{2p} \to 0$ and $\frac{n\sigma_{i,n}^{2d+2}}{(\log n)^2} \to \infty$.

Condition (C1) implies the boundedness of $\mathbf{X}$ and $\mathbf{Z}$, and the positiveness and boundedness of the density of $W$ or $\mathbf{X}$, which are reasonable in practice. Conditions (C2) and (C3) are similar to the assumptions adopted in Peng and Huang (2008), and impose realistic assumptions for covariates, and underlying coefficient processes and density functions. Like in Peng and Huang (2008), condition (C4) is the key assumption to ensure the identifiablilty of $\{\boldsymbol{\beta}_0(\tau), \tau \in (0, \tau_U]\}$. By condition (C5), $\kappa_v(\mathbf{U})$ and $\pi(\mathbf{X})$ are bounded away from 0 and 1 almost surely. Condition (C6) implies that truncating $\hat{\kappa}_v$ by $\mathcal{I} = (c_{l,n}, c_{u,n})$ would only lead to negligible impact on the asymptotic results with $c_{l,n}$ and $c_{u,n}$ approaching 0 and 1 respectively. Conditions (C7)-(C8) are similar to the regularity conditions in Newey (1994) for kernel estimators. With these conditions, we are able to strengthen Newey (1994)'s result for the kernel estimators of full means based on $\hat{\pi}(\mathbf{x})$ and $\hat{v}(\mathbf{x})$. By condition (C8), we require $v(\cdot)$ and $\pi(\cdot)$ should be smooth enough, and bandwidths satisfy $\sigma_{k,n} = o(n^{-1/(2p)} \wedge (\log n)^{1/(d+1)} n^{-1/(2d+2)})$, $k = 1, 2$.

We establish the uniform consistency and weakly convergence of the proposed estimator in the following theorems:

**Theorem 3.1** (Uniform consistency)**.** *Under the Condition (C1)-(C8) and* $\lim_{n\to\infty} \|S_L\| = 0$*, we have* $\sup_{[v,\tau_U]} \|\hat{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}_0(\tau)\| \to_p 0$*, where* $0 < v < \tau_U$*.*

**Theorem 3.2** (Weakly convergence)**.** *Under the Conditions (C1)-(C8) and* $\lim_{n\to\infty} n^{1/2}\|S_L\| = 0$*, we have* $n^{1/2}\{\hat{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}_0(\tau)\}$ *weakly converges to a Gaussian process for* $\tau \in [v, \tau_U]$*, where* $0 < v < \tau_U$*.*

To prove Theorem 3.1, following the rationale presented in Peng (2012), the key step is to show that using $\kappa_{v,i}$ in place of $\hat{\kappa}_{v,i}$ in $\mathbf{S}_n(\boldsymbol{\beta}, \tau)$ only induces a small variation to $\mathbf{S}_n(\boldsymbol{\beta}, \tau)$, which converges to 0 uniformly in $\tau$. Such a result is essentially

entailed by the uniform convergence of the kernel estimator of $\kappa_v(\mathbf{U})$ (Newey, 1994; Hansen, 2008). However, it is much more challenging to investigate the asymptotic distribution of $\hat{\boldsymbol{\beta}}(\cdot)$. For example, one substantial challenge is about how to mitigate the bias of the kernel weight estimate,which can be larger than $O(n^{-1/2})$ to achieve the weak convergence of $\hat{\boldsymbol{\beta}}(\cdot)$ at the root-$n$ rate. Another complication is that the variability induced to $n^{-1/2}\mathbf{S}_n(\boldsymbol{\beta}_0, \tau)$ by $\tilde{\kappa}_{v,i}$ is not asymptotically negligible and takes a complex form involving stochastic integrals. To overcome these difficulties, in the proof of Theorem 2, we first delicately strengthen the uniform convergence result of Peng and Huang (2008). Combining this new result with the properties of the kernel estimate $\tilde{\kappa}_{v,i}$ and empirical process arguments, we are able to establish a smooth approximation to the change in the discontinuous $n^{-1/2}\mathbf{S}_n(\boldsymbol{\beta}, \tau)$as $\boldsymbol{\beta}$ is slightly moved away from $\boldsymbol{\beta}_0$. We show that this smooth approximation may not be affected by the bias of $\tilde{\kappa}_{v,i}$ (see Lemma 3.2). This plays an important role to justify an elegant relationship between $n^{\hat{\boldsymbol{\beta}}(\tau)-\boldsymbol{\beta}_0(\tau)}$ and $n^{-1/2}\mathbf{S}_n(\boldsymbol{\beta}_0, \tau)$ via a production integration. We also establish a stronger uniform version of Newey (1994)'s result for kernel estimators of full means. This consequently allows us to derive a uniform i.i.d. sum representation of $n^{1/2}\mathbf{S}_n(\boldsymbol{\beta}_0, \tau)$ (see Lemma 3.3). This finding, coupled with the link between $n^{1/2}\{\hat{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}_0(\tau)\}$ and $n^{1/2}\mathbf{S}_n(\boldsymbol{\beta}_0, \tau)$, implies the weak convergence of $n^{1/2}\{\hat{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}_0(\tau)\}$ to a Gaussian process. The detailed proofs of Theorems 3.1 and 3.2 are provided in section 3.5.2 of Appendix D.

### 3.2.4 Inference

Given the complexity in the asymptotic distribution of $\hat{\boldsymbol{\beta}}(\tau)$ derived in Theorem 3.2, we propose to make inference on $\boldsymbol{\beta}_0(\cdot)$ using bootstrapping. For example, to estimate the asymptotic variance of $\hat{\boldsymbol{\beta}}(\tau)$, we first resample $n$ samples from the observed dataset $\{\mathbf{O}_i\}_{i=1}^n$ with replacement, and then obtain an estimate, $\hat{\boldsymbol{\beta}}^*(\tau)$ by applying the same estimation procedure described in Section 3.2.2 to the resampled dataset.

By repeating this procedure for many times, we can get a large number of resampled estimates, denoted by $\{\hat{\boldsymbol{\beta}}_b^*(\tau)\}_{b=1}^B$. The variance of $\hat{\boldsymbol{\beta}}(\tau)$ can be estimated by the empirical variance of $\{\hat{\boldsymbol{\beta}}_b^*(\tau)\}_{b=1}^B$.

As suggested by the asymptotic studies, for the proposed IV estimator, the first-order term of the Edgeworth expansion involves $\mathbf{B}\{\boldsymbol{\beta}_0(\tau)\}^{-1}$, which would approach $\infty$ as the proportion of compliers goes to zero. This means, given a weak IV that yields a low proportion of compliers, the first-order term of the Edgeworth expansion would be large and dominated by higher-order terms for a given sample size. Thus, the Edgeworth expansion may not provide a good finite-sample approximation to the distribution of interest. Consequently, unstable and inaccurate estimation and bootstrap inference may be expected when the proportion of compliers is small and the sample size is not large enough. This issue shares a similar spirit with the drawback of the two-stage least squares estimator in the global IV estimation setting (Horowitz, 2001).

Given the uniform consistency and weak convergence established in Theorems 3.1 and 3.2, we can further conduct second-stage inference to test the overall significance and the constancy of CQCE($\tau$) or other regression coefficients over $\tau \in ]v, \tau_U]$. The specific inference procedure can follow similar lines of existing work on censored quantile regression (Peng and Huang, 2008, for example).

## 3.3 Numerical Studies

### 3.3.1 Monte-Carlo simulations

Extensive simulations are conducted to assess the finite-sample performance of the proposed method. To satisfy assumptions (A1)-(A4), the data are generated as follows:

1 Generate $\mathbf{X} = (X_1, X_2)$, where $X_1 \sim Unif(0, 1)$ and $X_2 \sim Bernoulli(0.5)$.

2 Generate the latent compliance subgroup membership from a multinomial distribution with $p_c = 2/3$, $p_a = p_n = 1/6$, where $p_c$. $p_a$ and $p_n$ denote proportion of compliers, proportion of always takers, and proportion of never takers respectively.

3 Given $\mathbf{X}$, generate $V$ from $Bernoulli(\pi(\mathbf{X}))$, where

$$\pi(\mathbf{X}) = \frac{exp(0.1 * X_2 + X_1^2 + X_1X_2 + \epsilon)}{1 + exp(0.1 * X_2 + X_1^2 + X_1X_2 + \epsilon)}.$$

Here $\epsilon \sim Normal(0, 0.5^2)$, and controls the variation of $\pi(\mathbf{X})$ from a logistic regression model.

4 Determine $D$ based on $V$ and the latent compliance subgroup membership from Step 2.

5 Generate the survival time $T$ separately for compliers and non-compliers from models specified below.

6 Independently draw censoring time from a uniform distribution to generate about 30% censoring.

In Step 5, we consider two scenarios for generating survival time $T$:

(A)

$$T = \begin{cases} \exp(-0.2X_1 - 0.3X_2 + 0.5D + \epsilon_c), & D_1 > D_0 \\ \exp(-0.1X_1 - 0.2X_2 + 0.2D + \epsilon_{nc}), & \text{otherwise} \end{cases}, \tag{3.8}$$

where $\epsilon_c \sim Extremevalue(1)$ and $\epsilon_{nc} \sim Normal(0, 0.5^2)$.

(B)

$$T = \begin{cases} \exp(\log\tau - 0.2X_1 - 0.3X_2 + 0.5 \times \exp(0.3\tau) \times D), & D_1 > D_0 \\ \exp(-0.1X_1 - 0.2X_2 + 0.2D + \epsilon_{nc}), & \text{otherwise} \end{cases}, \tag{3.9}$$

where $\tau \sim Unif(0,1)$ and $\epsilon_{nc} \sim Normal(0, 0.5^2)$.

We can show that in scenario (A),

$$Q_{\log T}(\tau|D, \mathbf{X}, D_1 > D_0) = \log(-\log(1-\tau)) - 0.2X_1 - 0.3X_2 + 0.5D \tag{3.10}$$

holds, and in scenario (B),

$$Q_{\log T}(\tau|D, \mathbf{X}, D_1 > D_0) = \log(\tau) - 0.2X_1 - 0.3X_2 + 0.5 \times \exp(0.3\tau)D \tag{3.11}$$

holds. Note that the CCQTE $\beta_d(\tau)$ is constant over $\tau$ under model (3.10), and varies with $\tau$ under model (3.11).

For each scenario, we consider two different sample sizes, 1000 and 2000. In each setting, we generate 1000 simulated datasets and choose $B = 250$ as the number of bootstrapping. To obtain the $\tilde{\kappa}_{v,i}$ in the proposed method, we adopt the second order of Epanechnikov kernel as the kernel functions in our estimation procedure. For computational simplicity, we let bandwidths $\sigma_{1,n}$ and $\sigma_{2,n}$ both equal $\sigma$, where $\sigma$ is chosen from $\{0.1, 0.2, \cdots, 0.9\}$ and minimizes $\sum_{i\in\text{test set}} |V_i - \hat{v}(\mathbf{U}_i)| + \sum_{i\in\text{test set}} |V_i - \hat{\pi}(\mathbf{X}_i)|$ respectively. We set $c_{l,n} = 10/n$ and $c_{u,n} = 1 - 10/n$. Except for the proposed method, we consider the naive as-treated method which directly applies Peng and Huang (2008) to the whole dataset, an oracle benchmark method, which applies Peng and Huang (2008)'s method to the true latent complier group, and a modified proposed method with weights, $\kappa_{v,i}$'s estimated by performing logistic regression of $V$ over $\mathbf{X}$ or $\mathbf{Y}$.

Figure 3.1 and Figure 3.2 show the simulation results from scenarios (A) and (B) respectively, with $n = 1000$. The results include the average estimated coefficients, average variance estimates along with empirical variances, and coverage probabilities of 95% confidence intervals based on the proposed method, the naive method and the benchmark method for $\tau \in [0.1, 0.6]$. Not surprisingly, the estimates from the bench-

mark method are the best among all three methods in both settings. Impressively, the performance of the proposed method is fairly close to that of the benchmark method. For scenario (A), based on the proposed method, the coefficient estimates are very close to true coefficients, the empirical coverage probabilities of 95% CI are close to the nominal level, and the bootstrapping-based variance estimates agree with the empirical variances quite well. In contrast, the naive as-treated method produces substantially biased estimates, and the resulting empirical coverage probabilities deviate from the nominal level. This suggests that ignoring treatment endogeneity by the unobserved confounders, as in the naive as-treated method, can lead to problematic estimation and inference. In Figure 3.1, we also note that, when the weights in our two-stage estimation procedure are calculated based on logistic regression models that are misspecified, the resulting estimation and empirical coverage probabilities of 95% CI can be considerably biased. This observation highlights the importance of using the proposed nonparametric weight estimates, $\tilde{\kappa}_{v,i}$'s.

We have similar observations for scenario (B). The results are presented in Figure 3.1, suggesting that the proposed method also works well when $CQCE(\tau)$is $\tau$-varying. That is, the empirical estimation biases and the departures of the empirical coverage probabilities from 95%, though slightly larger than those in scenario (A), are still rather minimal and moreover decrease with the sample size. We also observe a very good agreement between the estimated variances and empirical variance

Figure 3.3 and Figure 3.4 show the simulation results with $n = 2000$ for scenarios (A) (B). We note that the biases of the proposed estimator further diminish as the sample size increases. The performance regarding the variance estimation and confidence intervals also improves with the sample size.

Next, we consider scenarios (C) and (D), which are the same as scenarios (A) and (B) respectively, except that the proportion of compliers, $p_c$, is reduced from 2/3 to 1/3, and the proportion of always takers and the proportion of never takers are both

set as 1/3. The results with $n = 1000$ and $n = 2000$ are presented in Figure 3.5-3.8. As expected, we observe that the empirical biases and variances from the proposed estimator increase as the proportion of compliers decreases to 1/3. Nevertheless, the resulting estimates are still fairly close to the true values, and have smaller biases than the naive methods. The empirical variances and the estimated variances still agree well with each other, and their agreement improves with $n$.

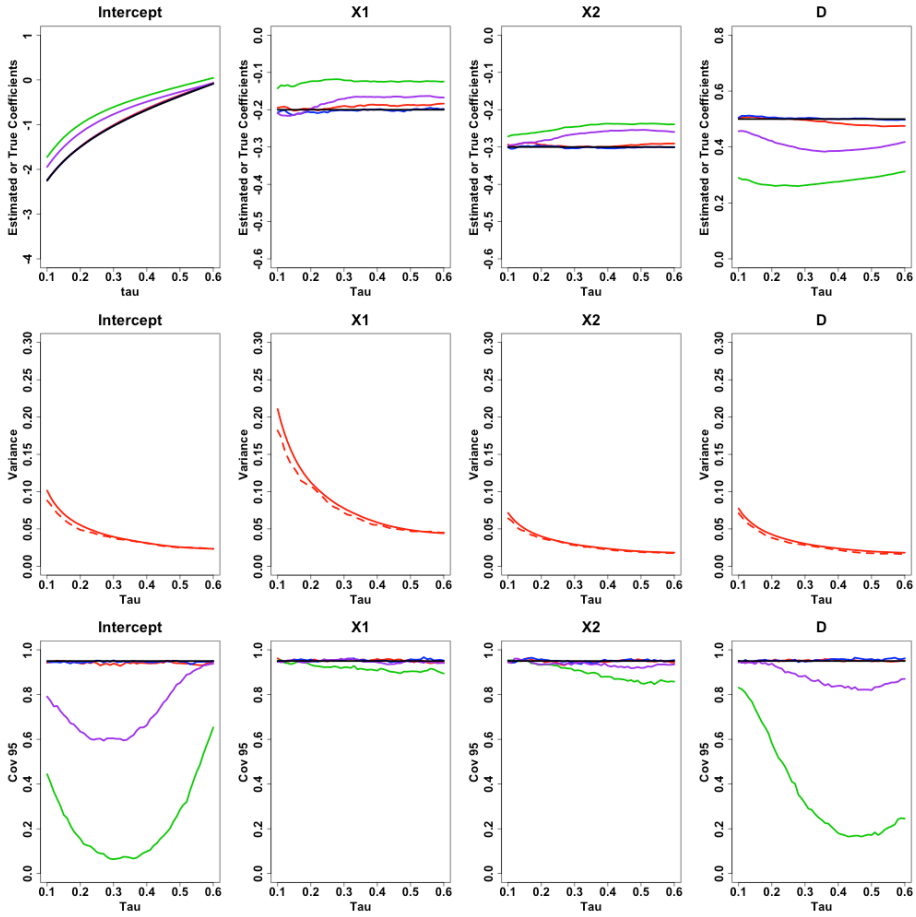


Figure 3.1: The simulation results in scenario (A) with sample size n=1000. In the first and the third rows, black lines correspond to true or nominal values; red lines correspond to the proposed method; blue lines correspond to the benchmark method; green lines correspond to the naive as-treated method; purple lines correspond to a modified proposed method with weights estimated by logistic regression. In the second row, red solid lines represent the average variance estimates from the proposed method, and red dashed lines represent the empirical variances of the proposed estimator.

### 3.3.2 An application to a dataset from the Center for International Blood and Marrow Transplant Research

Diffuse large B-cell lymphoma (DLBCL) is the most common aggressive non-Hodgkin lymphoma (NHL). A dataset from the Center for International Blood and Marrow Transplant Research (CIBMTR) includes 986 DLBCL patients between 18 to 76 years old with autologous hematopoietic stem cell transplantation (TX) between 1996 and 2003. Among these 986 patients, 174 patients used rituximab in the pre-transplant treatment, while other 812 patients did not. Excluding 38 patients who have missing data, Table 3.1 presents the summary statistics of covariates and the endpoint of interest, which is the composite of progressive lymphoma post-transplant, lymphoma recurrence, and death from any cause. One specific interest from analyzing this dataset is to investigate whether the pre-transplant rituximab treatment can improve the progression free survival (PFS), defined as time to the composite endpoint, in DLBCL patients. The overall rate of censoring to PFS is 36%.

A complication is that selecting rituximab as a pre-transplant treatment may be influenced by the observed and unobserved patient characteristics and risk factors. For example, Table 1 shows that patients in the rituximab group tended to be older, and have received more chemotherapy regimens than patients in the control group. This suggests that treatment selection may be confounded by factors that potentially influence the PFS. Some of these factors, for example, molecular subtype of lymphoma, were not captured by this dataset (Zheng et al., 2017).

Controlling for the covariates captured in this dataset, including age, number of chemotherapy regimens, disease status, and Karnofsky performance score, Fenske et al. (2009) reported that using rituximab in the pre-transplant treatment may improve the PFS in DLBCL patients. We apply the proposed IV method to further account for the impact of unmeasured confounders in the estimation of causal treatment effect. We adopt the IV variable suggested in Zheng et al. (2017), which is

the indicator of whether the treatment started after FDA approval of the use of rituximab on November 26, 1997. During the relatively short study period from 1996 to 2003, there were no major changes in clinical practice or technical improvements for the autologous hematopoietic stem cell transplantation, except for the FDA approval of rituximab. Therefore, it is reasonable to assume that the adopted IV affects the PFS outcome only through the use of rituximab, and is independent of unmeasured confounders. These then justify the independence of IV assumption (A1) and the exclusion assumption (A2). Note that DLBCL patients could not receive the pre transplant rituximab treatment before the FDA approval of rituximab. Consequently, this confers a one-sided compliance case, and the monotonicity assumption (A4) is automatically satisfied.

Table 2 3.2 presents the summary statistics stratified by the IV variable. From Table 2, we observe a strong positive association between using rituximab in the pre-transplant treatment and starting treatment after the FDA approval. This further justifies the validity of the adopted IV. Age and Karnofsky performance scores are similar between patients who began their treatments before FDA approval of rituximab and those who started their treatments after the FDA approval. The distributions of disease status and number of chemotherapy regimens however may differ between these two groups of patients. This suggests the need to include these factors as covariates in our model (3.1). Note that the rituximab became available only after the FDA approval date. Thus, this example renders a one-sided compliance IV setting. As discussed in Section 3.2.1, the proposed estimate for $\beta_d$ can be interpreted as the estimated causal quantile treatment effect for the treated population.

We apply the proposed method, the modified proposed method, and the as-treated analysis based on the method of Peng and Huang (2008), all of which adjust for all the observed covariates. To implement the proposed method, we adopt the second order Epanechnikov kernel and choose bandwidths $\sigma_{1,n}$ and $\sigma_{2,n}$ in the same way as in

simulation studies, except that we select from $\{0.02, 0.04, \ldots, 0.2\}$. The results on the treatment coefficients are presented in Figure 3.9, and results on the other coefficients can be found in Figure 3.10. By Figure 3.9, the estimated CQCE(0.5) equals 1.42, which, combined with the corresponding intercept coefficient estimate, 3.39, in Figure 3.10 indicates a clinically significant treatment effect on median PFS time. That is, for rituximab treated patients with covariates set at their reference levels (i.e. younger than 55 years, having 0 or 1 chemotherapy regimens, with CR1 disease status and karnofsky score $< 90$), receiving rituximab may have prolonged their median PFS from 26.67 ($= \exp(3.39)$) months to 122.73 ($= \exp(3.39 + 1.42)$) months. Similarly, the estimate for CQCE(0.25), 0.56, suggests a causal effect of rituximab in prolonging the 25th percentile of the same potential PFS from 15.3 months to 26.9 months for treated patient with reference covariate values. This shows that assessing CQCE($\tau$) based on Figure 3.9 can depict a comprehensive picture of the causal treatment effect of rituximab.

As noted from Figure 3.9, the proposed method reveals a stronger beneficial effect of rituximab than the as-treated analysis. By the proposed method, CQCE($\tau$) of rituximab, is significantly above 0 across all $\tau$'s between 0.1 and 0.6, while by the as-treated analysis, the positive effect of rituximab is not significant on the quantiles of PFS time with $\tau < 0.3$. This suggests that the naive as-treated analysis may underestimate the effect of rituximab for the patients with poor PFS. This result is interesting and also sensible. A reasonable explanation is that sicker patients are more likely to choose rituximab treatment, as suggested by Table 3.1. Because sicker patients tend to have more rapid disease progression, implying shorter PFS, ignoring the unmeasured sickness related confounders, as in the as-treated analysis, as in the as-treated analysis, can lead to an attenuated positive effect of rituximab, particularly on the lower quantiles of the potential PFS. The modified proposed method with weights estimated from logistic regression suggests stronger effects of rituximab

than does the as-treated analysis, but still fails to identify the beneficial effects of rituximab at lower quantiles of PFS with $\tau < 0.27$. Such a discrepancy may reflect the estimation bias from incorrectly assuming a logistic regression model for the IV, a phenomenon clearly demonstrated by the simulation studies. Our finding about rituximab's effect generally agrees with the results in Zheng et al. (2017) that the mean effect of rituximab was underestimated in the standard analysis without addressing the bias from the endogenous treatment selection.

## 3.4 Remarks

In this work, we quantify causal treatment effect by complier quantile causal effect (CQCE), which is a meaningful counterpart of the complier average causal effect (CACE) that has been commonly studied in standard IV literature. For a time-to-event outcome subject to censoring, CQCE is identifiable under weaker conditions than CACE, which generally cannot be estimated with bounded censoring. CQCE also offers greater flexibility in depicting the causal treatment effect than other causal estimands in survival analysis, such as CCPHR and complier location shift effect. We develop a simple and rigorously justified two-stage estimation procedure, and elaborate how it can readily be implemented by existing software. The ease of implementation should facilitates future applications of the proposed method.

To apply the proposed method, a prerequisite step is to identify a binary IV for which (A1)-(A4) are plausible assumptions. One also needs to make sure the resulting interpretation of CQCE is scientifically relevant. Readers may refer to Baiocchi et al. (2014) for discussions and examples of reasonable IVs.

Like many other IV approaches, the proposed method faces challenges when the selected IV is a weak IV that is characterized by a low proportion of compliers in the present IV setting (Sovey and Green, 2011). As explained in Section 3.2.4, when a

weak IV is used, the proposed estimation and the associated bootstrap inference can produce unstable and inaccurate results. This is confirmed by our simulation studies. Another potential problem from a weak IV is high sensitivity to the violation of the independence of IV assumption, as discussed in Baiocchi et al. (2014). That is, when the IV has only a minor association with unmeasured confounders, the resulting estimation bias can be exacerbated when the IV is weak. Therefore, caution is warranted when a weak IV is suspected.

## 3.5 Appendix

### 3.5.1 Appendix C: Propositions and their proofs

**Proposition 3.1.** *Under assumptions (A1) and (A2), in the one-sided compliance case, where subjects with* $V = 0$ *have no access to the treatment (i.e.* $Pr(D_0 = 0|X) = 1$*),* $\beta_d(\tau) = Q_{\log T_1}(\tau|\mathbf{X}, D = 1) - Q_{\log T_0}(\tau|\mathbf{X}, D = 1)$

*Proof.* First, $\Pr(D_0 = 0|X) = 1$ implies that subjects with $D = 1$ must belong to the complier group. Given the fact that $T = D \times T_1 + (1 - D) \times T_0$, we have It follows that

$$\begin{aligned} Q_{\log T}(\tau|D = 1, \mathbf{X}, D_1 > D_0) &= \inf\{t : \Pr(\log T_1 \leq t|D = 1, \mathbf{X}, D_1 > D_0) \geq \tau\} \\ &= \inf\{t : \Pr(\log T_1 \leq t|D = 1, \mathbf{X}) \geq \tau\} \\ &= Q_{\log T_1}(\tau|\mathbf{X}, D = 1). \end{aligned}$$

At the same time,

$$
\begin{aligned}
&Q_{\log T}(\tau|D=0,\mathbf{X},D_1>D_0)=\inf\{t:\Pr(\log T_0\le t|D=0,\mathbf{X},D_1>D_0)\ge\tau\}\\
=&\inf\{t:\Pr(\log T_0\le t|V=0,\mathbf{X},D_1=1)\ge\tau\}\\
=&\inf\{t:\Pr(\log T_0\le t|V=1,\mathbf{X},D_1=1)\ge\tau\}\\
=&\inf\{t:\Pr(\log T_0\le t|D=1,\mathbf{X})\ge\tau\}\\
=&Q_{\log T_0}(\tau|\mathbf{X},D=1).
\end{aligned}
$$

The second equality follows from the one-sided compliance constraint, $\Pr(D_0 = 0|X) = 1$, and the third equality is ensured by assumption (A1). Assumption (A1), combined with $\Pr(D_0 = 0|X) = 1$, further implies the fourth equality.

Therefore, $\beta_d(\tau)$ is equivalent to $Q_{\log T_1}(\tau|\mathbf{X}, D = 1) - Q_{\log T_0}(\tau|\mathbf{X}, D = 1)$. □

**Proposition 3.2.** *Under assumptions (A1) and (A2), model* (3.1) *and model* (3.2) *are equivalent.*

**Proof of Proposition 3.2:** Given the fact that $T = D \times T_1 + (1 - D) \times T_0$, we have

$$
\begin{aligned}
Q_{\log T}(\tau|D=1,\mathbf{X},D_1>D_0)&=\inf\{t:\Pr(\log T\le t|D=1,\mathbf{X},D_1>D_0)\ge\tau\}\\
&=\inf\{t:\Pr(\log T_1\le t|D=1,\mathbf{X},D_1>D_0)\ge\tau\}
\end{aligned}
$$

Since $V = D$ for the compliers, then $\inf\{t : \Pr(\log T_1 \le t|D = 1, \mathbf{X}, D_1 > D_0) \ge \tau\} = \inf\{t : \Pr(\log T_1 \le t|V = 1, \mathbf{X}, D_1 > D_0) \ge \tau\}$. Under assumption (A1) and (A2), we have

$$
\begin{aligned}
\inf\{t:\Pr(\log T_1\le t|V=1,\mathbf{X},D_1>D_0)\ge\tau\}&=\inf\{t:\Pr(\log T_1\le t|\mathbf{X},D_1>D_0)\ge\tau\}\\
&=Q_{\log T_1}(\tau|\mathbf{X},D_1>D_0)
\end{aligned}
$$

Similarly, we have $Q_{\log T}(\tau|D = 0, \mathbf{X}, D_1 > D_0) = Q_{\log T_0}(\tau|\mathbf{X}, D_1 > D_0)$.

These imply that $Q_{\log T}(\tau|D=d,\mathbf{X},D_1>D_0)=Q_{\log T_d}(\tau|\mathbf{X},D_1>D_0)$, $d=0,1$. Thus, model (3.1) and model (3.2) are equivalent.

**Proposition 3.3.** *Under assumption (A1)-(A5), suppose* $\kappa_v(\mathbf{U})=\Pr(D_1>D_0|\mathbf{U})$, *then* $\kappa_v(\mathbf{U})=1-\frac{D(1-v(\mathbf{U}))}{1-\pi(\mathbf{X})}-\frac{(1-D)v(\mathbf{U})}{\pi(\mathbf{X})}$, *where* $v(\mathbf{U})=Pr(V=1|\mathbf{U})$ *and* $\pi(\mathbf{X})=Pr(V=1|\mathbf{X})$.

*Proof.* Recall $\mathbf{U}=(W,\delta,D,\mathbf{X})$. Note that $D(1-V)$ only differs from zero only if $D=1$ and $V=0$. By the monotonicity assumption, $D_0=1$ implies $D_1=1$. Then

$$
\begin{aligned}
&E(D(1-V)|\mathbf{U})=Pr(D(1-V)=1|\mathbf{U})=Pr(D_1=D_0=1,V=0|\mathbf{U})\\
&=Pr(D_1=D_0=1|\mathbf{U})Pr(V=0|D_1=D_0=1,W_1=min(T_1,C),\delta=I(T_1<C),\mathbf{X})\\
&=Pr(D_1=D_0=1|\mathbf{U})Pr(V=0|\mathbf{X}).
\end{aligned}
$$

□

The last equality holds because assumptions (A1) and (A5) imply that $V$ is independent of $(D_1,D_0,T_1,T_0)$ and $C$ conditional on $\mathbf{X}$. Similarly, we can show that$E((1-D)V|\mathbf{U})=Pr(D_1=D_0=0|\mathbf{U})Pr(V=1|\mathbf{X})$.. Therefore

$$
\begin{aligned}
&1-\frac{D(1-v(\mathbf{U}))}{1-\pi(\mathbf{X})}-\frac{(1-D)v(\mathbf{U})}{\pi(\mathbf{X})}\\
=&E\Big\{1-\frac{D(1-V)}{Pr(V=0|\mathbf{X})}-\frac{(1-D)V}{Pr(V=1|\mathbf{X})}\mathbf{U}\Big\}\\
=&1-Pr(D_1=D_0=1|\mathbf{U})-Pr(D_1=D_0=0|\mathbf{U})=Pr(D_1>D_0|\mathbf{U}).
\end{aligned}
$$

The last equality follows from the monotonicity assumption.

## 3.5.2 Appendix D: Proofs of Theorem 3.1 and Theorem 3.2

### 3.5.2.1 Technical lemmas and their proofs

**Lemma 3.1.** *Under the Conditions (C1)-(C8) and* $\lim_n \|S_L\| = o(n^{-1/2})$,

$$n^{1/4} \sup_{\tau \in (0,\tau_U]} \|\boldsymbol{\mu}_c(\tilde{\boldsymbol{\beta}}_n(\tau)) - \boldsymbol{\mu}_c(\boldsymbol{\beta}_0(\tau))\| \to_p 0$$

*Proof.* Define $\boldsymbol{\gamma}_{n,j} = n^{-1}\sum_{i=1}^n (\hat{\kappa}_{v,i} - \kappa_{v,i})\mathbf{Z}_i[N_i(\exp\{\mathbf{Z}_i^\top \hat{\boldsymbol{\beta}}_n(\tau_j)\}) - \int_0^{\tau_j} I(W_i \geq \exp\{\mathbf{Z}_i^\top \hat{\boldsymbol{\beta}}_n(u)\})dH(u)]$. Because $\sup_i |\hat{\kappa}_{v,i} - \kappa_{v,i}| = o_p(n^{-1/4})$ and the boundedness of $\mathbf{Z}$, we have $\sup_{1\leq j\leq L(n)} \|\boldsymbol{\gamma}_{n,j}\| = o_p(n^{-1/4})$.

Given $\hat{\boldsymbol{\beta}}(\tau)$ is defined as a generalized solution of model (3.2), we have

$$n^{-1}\sum_{i=1}^n \hat{\kappa}_{v,i}\mathbf{Z}_i N_i(\exp\{\mathbf{Z}_i^\top \hat{\boldsymbol{\beta}}_n(\tau_j)\}) = n^{-1}\sum_{i=1}^n \hat{\kappa}_{v,i}\mathbf{Z}_i - \int_0^{\tau_j} I(W_i \geq \exp\{\mathbf{Z}_i^\top \hat{\boldsymbol{\beta}}_n(u)\})dH(u) + \boldsymbol{\xi}_{n,j}$$

where $\max_{j=1,2,\cdots,M} \|\boldsymbol{\xi}_{n,j}\| \leq \sup_i \|\hat{\kappa}_{v,i} Z_i\|/n \leq \sup_i \|Z_i\|/n = o(n^{-1/4})$. After some algebraic manipulations, we have

$$\begin{aligned} n^{1/4}\{\boldsymbol{\mu}_c(\hat{\boldsymbol{\beta}}_n(\tau_j)) - \boldsymbol{\mu}_c(\boldsymbol{\beta}_0(\tau_j))\} =& -n^{1/4}\boldsymbol{\nu}_n(\hat{\boldsymbol{\beta}}_n(\tau_j)) + n^{1/4}\int_0^{\tau_j} \tilde{\boldsymbol{\nu}}_n(\hat{\boldsymbol{\beta}}_n(u))dH(u) \\ &+ \sum_{k=1}^{j}\int_{\tau_{k-1}}^{\tau_k} n^{1/4}[\tilde{\boldsymbol{\mu}}_c(\hat{\boldsymbol{\beta}}_n(u)) - \tilde{\boldsymbol{\mu}}_c(\boldsymbol{\beta}_0(u))]dH(u) \\ &+ n^{1/4}\{\boldsymbol{\xi}_{n,j} - \boldsymbol{\gamma}_{n,j}\}, \end{aligned}$$

where $\sup_i n^{1/4}\|\boldsymbol{\xi}_{n,j} - \boldsymbol{\gamma}_{n,j}\| = o(1)$ a.s.

From the Law of Iterated Logarithm for empirical process on Vapnik-Črvonenkis (VC) (Alexander et al., 1989), $\sup_{\mathbf{b}} n\boldsymbol{\nu}_n(\mathbf{b}) = \sum_{i=1}^n \kappa_{v,i}\mathbf{Z}_i N_i\{\exp(\mathbf{Z}_i^\top \mathbf{b})\} - \boldsymbol{\mu}_c(\mathbf{b}) = O(n^{1/2}(\log\log n))$, which implies $\sup_{\mathbf{b}} n^{1/4}\boldsymbol{\nu}_n(\mathbf{b}) = o(1)$. Similarly, we have $\sup_{\mathbf{b}} n^{1/4}\tilde{\boldsymbol{\nu}}_n(\mathbf{b}) = o(1)$. Therefore, as $n$ is sufficient large, we have $\sup_i n^{1/4}\|\boldsymbol{\xi}_{n,j} - \boldsymbol{\gamma}_{n,j}\| < C_2$ and $\sup_{\mathbf{b}} n^{1/4}\|\{\boldsymbol{\nu}_n(\mathbf{b}) + \int_0^{\tau_j} \tilde{\boldsymbol{\nu}}_n(\hat{\boldsymbol{\beta}}_n(u))dH(u)\}\| < C_1$, where $C_1$ and $C_2$ can be arbi-

trary small constants. From C2(a), there is a constant $C_3$, such that $\|\boldsymbol{\mu}_c\{\boldsymbol{\beta}_0(\tau')\} - \boldsymbol{\mu}_c\{\boldsymbol{\beta}_0(\tau)\}\| = C_3|\tau'-\tau|$. From C3(c), there is a constant $C_4$, such that $\|\{\mathbf{B}(\mathbf{b})\}^{-1}\mathbf{J}(\mathbf{b})\mathbf{x}\| \leq C_4\|\mathbf{x}|$ for any $\mathbf{x} \in R^{p+1}$ and $\mathbf{b} \in B(d_0)$.

Let $\tilde{a}_n = n^{1/4}a_n$ and $b_n = a_n/(1-\tau_U)$. Since $a_n = o(n^{-1/2})$, we have $\tilde{a}_n = o(1)$ and $b_n = o(1)$. Define $\varepsilon_0 = C_3\tilde{a}_n$, $\varepsilon_1 = C_1 + \epsilon_0 C_4 b_n + C_2 + C_3\tilde{a}_n$, and $\varepsilon_l = C_1 + (\sum_{k=0}^{l-1}\varepsilon_k)C_4b_n + C_2 + C_3\tilde{a}_n$ for $l = 2, \cdots, L-1$. It can be shown $\varepsilon_l$ is increased with $l$. and $\varepsilon_l = (1+C_4b_n)^(l-1)(C_1+\varepsilon_0C_4b_n+C_2+C_3\tilde{a}_n)$. Since $\tilde{a}_n = o(1)$ and $L = \tau_U/a_n$, we get $\lim_{n\to\infty}(1+C_4b_n)^{L-1} = \exp\exp\{C_4\tau_U/(1-\tau_U)\}$. Thus with $C_1$ chosen small enough for $n \geq$ some $N_0$, we have $\varepsilon_l \leq x\exp\{\tau_U/(1-\tau_U)\}C_1 < d_0$ for $l = 1, \ldots, L$

Following the similar arguments in the proof of Theorem 1 in Peng and Huang (2008), we can show that $\varepsilon_{L-1}$ is the upper bound for $\sup_{(0,\tau_U]} n^{1/4}\|\boldsymbol{\mu}_c\{\hat{\boldsymbol{\beta}}_n(\tau)\} - \boldsymbol{\mu}_c\{\boldsymbol{\beta}_0(\tau)\}\|$ with probability 1. Since $\tilde{a}_n = o(1)$ and $b_n = o(1)$, and $C_1$ and $C_2$ can be arbitrarily small, we have $\sup_{(0,\tau_U]} n^{1/4}\|\boldsymbol{\mu}_c\{\hat{\boldsymbol{\beta}}_n(\tau)\} - \boldsymbol{\mu}_c\{\boldsymbol{\beta}_0(\tau)\}\| = o_p(1)$. □

**Lemma 3.2.** *For any sequence* $\{\tilde{\boldsymbol{\beta}}_n(\tau), \tau \in (0,\tau_U]\}_{n=1}^{\infty}$ *that satisfies* $n^{1/4}\sup_{\tau\in(0,\tau_U]}\|\boldsymbol{\mu}_c(\tilde{\boldsymbol{\beta}}_n(\tau)) - \boldsymbol{\mu}_c(\boldsymbol{\beta}_0(\tau))\| \to_p 0$, *then*

$$\sup_{\tau\in(0,\tau_U]} \|n^{-1/2}\sum_{i=1}^{n}\hat{\kappa}_{v,i}\mathbf{Z}_i[N_i(\exp\{\mathbf{Z}_i^{\top}\tilde{\boldsymbol{\beta}}(\tau)\}) - N_i(\exp\{\mathbf{Z}_i^{\top}\boldsymbol{\beta}_0(\tau)\})] - n^{1/2}[\boldsymbol{\mu}_c\{\tilde{\boldsymbol{\beta}}(\tau)\} - \boldsymbol{\mu}_c\{\boldsymbol{\beta}_0(\tau)\}]\| \to_p 0. \tag{3.12}$$

*Proof.* By the assumption that $n^{1/4}\sup_{\tau\in(0,\tau_U]}\|\boldsymbol{\mu}_c(\tilde{\boldsymbol{\beta}}_n(\tau)) - \boldsymbol{\mu}_c(\boldsymbol{\beta}_0(\tau))\| \to_p 0$ , we have $\tilde{\boldsymbol{\beta}}_n(\tau) \in \mathcal{B}(d_0), \tau \in (0,\tau_U]$. We first note that

$$n^{-1/2}\sum_{i=1}^{n}\hat{\kappa}_{v,i}\mathbf{Z}_i[N_i(\exp\{\mathbf{Z}_i^{\top}\tilde{\boldsymbol{\beta}}_n(\tau)\}) - N_i(\exp\{\mathbf{Z}_i^{\top}\boldsymbol{\beta}_0(\tau)\})] - n^{1/2}[\boldsymbol{\mu}_c\{\tilde{\boldsymbol{\beta}}_n(\tau)\} - \boldsymbol{\mu}_c\{\boldsymbol{\beta}_0(\tau)\}]$$
$$= \mathrm{I}_n(\tau) + \mathrm{II}_n(\tau), \tag{3.13}$$

where

$$\begin{aligned}\mathrm{I}_n(\tau) =& n^{-1/2}\sum_{i=1}^{n}\kappa_{v,i}\mathbf{Z}_i\Big\{N_i(\exp\{\mathbf{Z}_i^\top\tilde{\boldsymbol{\beta}}_n(\tau)\}) - N_i(\exp\{\mathbf{Z}_i^\top\boldsymbol{\beta}_0(\tau)\})\Big\}\\ &- n^{1/2}\Big\{\boldsymbol{\mu}_c\{\tilde{\boldsymbol{\beta}}_n(\tau)\} - \boldsymbol{\mu}_c\{\boldsymbol{\beta}_0(\tau)\}\Big\},\end{aligned}$$

and

$$\mathrm{II}_n(\tau) = n^{-1/2}\sum_{i=1}^{n}(\hat{\kappa}_{v,i} - \kappa_{v,i})\mathbf{Z}_i[N(\exp\{\mathbf{Z}_i^\top\tilde{\boldsymbol{\beta}}_n(\tau)\}) - N(\exp\{\mathbf{Z}_i^\top\boldsymbol{\beta}_0(\tau)\})].$$

We first show $\sup_{\tau\in(0,\tau_U]}\|\mathrm{I}_n(\tau)\| \to_p 0$ by using the results of Alexander et al. (1984) and Lai and Ying (1988). Let $\phi(\mathbf{b},\tau) = \kappa_{v,i}\{N_i(\exp\{\mathbf{Z}_i^\top\mathbf{b}\}) - N_i(\exp\{\mathbf{Z}_i^\top\boldsymbol{\beta}_0(\tau)\})\}$, $\sigma^2(\mathbf{b},\tau) = \mathrm{Var}(\phi_i(\mathbf{b},\tau))$,

$$\Phi_{n,\nu} = \sup_{\tau\in(0,\tau_U],\sup_{\tau\in(0,\tau_U]}\sigma^2(\mathbf{b},\tau)\le\nu}\Big|n^{-1/2}\sum_{i=1}^{n}\{\phi(\mathbf{b},\tau) - E(\phi(\mathbf{b},\tau))\}\Big|,$$

and

$$\begin{aligned}\tilde{\Phi}_n = \sup_{\tau\in(0,\tau_U]}\Big|&n^{-1/2}\sum_{i=1}^{n}\kappa_{v,i}[N_i(\exp\{\mathbf{Z}_i^\top\tilde{\boldsymbol{\beta}}_n(\tau)\}) - N_i(\exp\{\mathbf{Z}_i^\top\boldsymbol{\beta}_0(\tau)\})]\\ &- n^{1/2}[\mu_c^{(2)}\{\tilde{\boldsymbol{\beta}}(\tau)\} - \mu_c^{(2)}\{\boldsymbol{\beta}_0(\tau)\}],\end{aligned}$$

where $\boldsymbol{\mu}_c^{(2)}\{\mathbf{b}\} = E\{\kappa_v N(\exp\mathbf{Z}^\top\mathbf{b})\} = E\{I(D_1 > D_0)N(\exp\mathbf{Z}^\top\mathbf{b})\}$.

We can show that the entropy condition (2.4) in Lai and Ying (1988) with $0 < r < 2$ is satisfied by the function class $\mathcal{F}_\nu = \{\phi(\mathbf{b},\tau) - E(\phi(\mathbf{b},\tau)) : \tau\in(0,\tau_U], \sup_{\tau\in(0,\tau_U]}\sigma^2(\mathbf{b},\tau)\le\nu\}$. By (2.6) in Lai and Ying (1988), there exists some constant $C_* > 0$ such that if $C_*\nu^{(2-r)/4}$ and $n^{(r-2)/2(r+2)} < s$, then

$$P(\Phi_{n,\nu}\ge s) \le 5\exp\{-s^2\nu^{-1}g(2s\cdot n^{-1/2}v^{-1})/4\}, \tag{3.14}$$

where $g(t) = 2t^{-2}\{(1+t)\log(1+t) - t\}$.

Given $\sup_{\tau\in(0,\tau_U]} \|\boldsymbol{\mu}_c(\tilde{\boldsymbol{\beta}}_n(\tau)) - \boldsymbol{\mu}_c(\boldsymbol{\beta}_0(\tau))\| = o_p(1)$, we can show that $\sup_{\tau\in(0,\tau_U]} \sigma^2(\tilde{\boldsymbol{\beta}}_n(\tau), \tau) \to_p 0$ by using similar arguments for Lemma 1 of Peng and Huang (2008). This means for any $\epsilon > 0$ and $s > 0$, there exists $N_{\epsilon,s,1} > 0$ such that for $n \geq N_{\epsilon,s,1}$,

$$P(\sup_{\tau\in(0,\tau_U]} \sigma^2(\tilde{\boldsymbol{\beta}}_n(\tau), \tau) > \nu_{\epsilon,s}) < \epsilon/2,$$

where $\nu_{\epsilon,s}$ satisfies $C_*\nu_{\epsilon,s}^{(2-r)/4} \leq s$ and $5\exp\{-C_1 s^2 \nu_{\epsilon,s}^{-1}/4\} < \epsilon/2$, with $C_1$ being a fixed constant between 0 and 1.

At the same time, there exist $N_{\epsilon,s,2} > 0$ such that for $n \geq N_{\epsilon,s,2}$, it holds that $n^{(r-2)/2(r+2)} < s$ and $g(2s \cdot n^{-1/2}\nu_{\epsilon,s}^{-1}) > C_1$. Then by (3.14),

$$P(\Phi_{n,\nu_{\epsilon,s}} \geq s) \leq 5\exp\{-C_1 s^2 \nu_{\epsilon,s}^{-1}/4\} < \epsilon/2.$$

For $n > \max(N_{\epsilon,s,1}, N_{\epsilon,s,2})$, we have

$$P(\tilde{\Phi}_n \geq s) \leq P(\Phi_{n,\nu_{\epsilon,s}} \geq s) + P(\sup_{\tau\in(0,\tau_U]} \sigma^2(\tilde{\boldsymbol{\beta}}_n(\tau), \tau) > \nu_{\epsilon,s}) < \epsilon/2 + \epsilon/2 = \epsilon.$$

This implies $\tilde{\Phi}_n \to_p 0$. Note that $\tilde{\Phi}_n$ corresponds to the supremum of the second component of $\mathrm{I}_n(\tau)$ over $\tau \in (0, \tau_U]$. Given the boundedness of $\mathbf{Z}$, we can similarly show the uniform convergence to 0 for the other components of $\mathrm{I}_n(\tau)$. Therefore, $\sup_{\tau\in(0,\tau_U]} \|\mathrm{I}_n(\tau)\| \to_p 0$.

To assess $\mathrm{II}_n(\tau)$, we note that

$$\begin{aligned}\sup_{\tau\in(0,\tau_U]}\|\mathrm{II}_n(\tau)\| &= \sup_{\tau\in(0,\tau_U]}\|n^{-1/2}\sum_{i=1}^{n}(\frac{\hat{\kappa}_{v,i}-\kappa_{v,i}}{\kappa_{v,i}})\kappa_{v,i}\mathbf{Z}_i[N(\exp\{\mathbf{Z}_i^\top\tilde{\boldsymbol{\beta}}_n(\tau)\})-N(\exp\{\mathbf{Z}_i^\top\boldsymbol{\beta}_0(\tau)\})]\| \\ &\le \sup_{\tau\in(0,\tau_U]} n^{-1/2}\sum_{i=1}^{n}|\frac{\hat{\kappa}_{v,i}-\kappa_{v,i}}{\kappa_{v,i}}|\cdot\|\kappa_{v,i}\mathbf{Z}_i[N(\exp\{\mathbf{Z}_i^\top\tilde{\boldsymbol{\beta}}_n(\tau)\})-N(\exp\{\mathbf{Z}_i^\top\boldsymbol{\beta}_0(\tau)\})]\| \\ &\le \sup_i n^{\frac{1}{4}}|\frac{\hat{\kappa}_{v,i}-\kappa_{v,i}}{\kappa_{v,i}}|n^{-\frac{3}{4}}\sup_{\tau\in(0,\tau_U]}\sum_{i=1}^{n}\|\kappa_{v,i}\mathbf{Z}_i[N(\exp\{\mathbf{Z}_i^\top\tilde{\boldsymbol{\beta}}_n(\tau)\})-N(\exp\{\mathbf{Z}_i^\top\boldsymbol{\beta}_0(\tau)\})]\|\end{aligned} \tag{3.15}$$

Let $O_i(\tau) = \|\kappa_{v,i}\mathbf{Z}_i[N(\exp\{\mathbf{Z}_i^\top\tilde{\boldsymbol{\beta}}_n(\tau)\})-N(\exp\{\mathbf{Z}_i^\top\boldsymbol{\beta}_0(\tau)\})]\|$. From (3.15), we have

$$\begin{aligned}\sup_{\tau\in(0,\tau_U]}\|\mathrm{II}_n(\tau)\| &\le \sup_i n^{\frac{1}{4}}\Big|\frac{\hat{\kappa}_{v,i}-\kappa_{v,i}}{\kappa_{v,i}}\Big|n^{-\frac{3}{4}}\sup_{\tau\in(0,\tau_U]}\sum_{i=1}^{n}O_i \\ &= \sup_i n^{\frac{1}{4}}\Big|\frac{\hat{\kappa}_{v,i}-\kappa_{v,i}}{\kappa_{v,i}}\Big|n^{-\frac{3}{4}}\sup_{\tau\in(0,\tau_U]}\sum_{i=1}^{n}(O_i(\tau)-E(O_i(\tau)) \\ &\quad+\sup_i n^{\frac{1}{4}}\Big|\frac{\hat{\kappa}_{v,i}-\kappa_{v,i}}{\kappa_{v,i}}\Big|n^{\frac{1}{4}}\sup_{\tau\in(0,\tau_U]}E(O(\tau))\end{aligned} \tag{3.16}$$

From conditions (C7) and (C8), we have $\sup_i n^{\frac{1}{4}}|\hat{\kappa}_{v,i}-\kappa_{v,i}| = o_p(1)$ (Newey, 1994). By the Law of the Iterated Logarithm for empirical processes on Vapnik-Ćrvonenkis (VC) (Alexander et al., 1989), we have

$$\sup_{\tau\in(0,\tau_U]}\sum_{i=1}^{n}(O_i(\tau)-E(O_i(\tau))) = O(n^{1/2}(\log\log n))$$

Then $\sup_i n^{\frac{1}{4}}\Big|\frac{\hat{\kappa}_{v,i}-\kappa_{v,i}}{\kappa_{v,i}}\Big|n^{-\frac{3}{4}}\sup_{\tau\in(0,\tau_U]}\sum_{i=1}^{n}(O_i(\tau)-E(O_i(\tau)) = o_p(1)$.

Next, we show $\sup_{\tau\in(0,\tau_U]}E(O(\tau)) = o_p(n^{-1/4})$. Note that $E(O(\tau)) = E\|I(D_1 > D_0)\mathbf{Z}[N(\exp\{\mathbf{Z}^\top\tilde{\boldsymbol{\beta}}_n(\tau)\})-N(\exp\{\mathbf{Z}^\top\tilde{\boldsymbol{\beta}}_0(\tau)\})]\|$ and the boundness of $\mathbf{Z}$, we only need to show that $\sup_{\tau\in(0,\tau_U]}E(O^{(2)}(\tau)) = E\|I(D_1 > D_0)[N(\exp\{\mathbf{Z}^\top\tilde{\boldsymbol{\beta}}_n(\tau)\})-N(\exp\{\mathbf{Z}^\top\tilde{\boldsymbol{\beta}}_0(\tau)\})]\| =$

$o_p(n^{-1/4})$. Then

$$
\begin{aligned}
&\sup_{\tau\in(0,\tau_U]} E(O^{(2)}(\tau)) \\
=&\sup_{\tau\in(0,\tau_U]} E\|I(D_1 > D_0)[N(\exp\{\mathbf{Z}^\top\tilde{\boldsymbol{\beta}}_n(\tau)\}) - N(\exp\{\mathbf{Z}^\top\boldsymbol{\beta}_0(\tau)\})]\| \\
\leq&\sup_{\tau\in(0,\tau_U]} E\|I(D_1 > D_0)\{N(\exp\{\mathbf{Z}^\top\tilde{\boldsymbol{\beta}}_0(\tau)\} + |\exp\{\mathbf{Z}^\top\tilde{\boldsymbol{\beta}}_n(\tau)\} - \exp\{\mathbf{Z}^\top\boldsymbol{\beta}_0(\tau)|) \\
&\quad - N(\exp\{\mathbf{Z}^\top\tilde{\boldsymbol{\beta}}_0(\tau)\} - |\exp\{\mathbf{Z}^\top\tilde{\boldsymbol{\beta}}_n(\tau)\} - \exp\{\mathbf{Z}^\top\boldsymbol{\beta}_0(\tau)|)\}\| \\
\leq&2\sup_t |\tilde{f}(t|\mathbf{Z})||\exp\{\mathbf{Z}^\top\tilde{\boldsymbol{\beta}}_n(\tau)\} - \exp\{\mathbf{Z}^\top\boldsymbol{\beta}_0(\tau)\}|
\end{aligned}
$$

.

Let $\mathcal{A}(d_0) = \{\boldsymbol{\mu}_c(\mathbf{b}) : \mathbf{b} \in \mathcal{B}_{d_0}\}$. Using similar arguments in the proof of Theorem 1 in Peng and Huang (2008), we can show that there exists an inverse function of $\boldsymbol{\mu}_c$, denoted by $\mathbf{h}$, from $\mathcal{A}(d_0)$ to $\mathcal{B}(d_0)$, such that $\mathbf{h}(\boldsymbol{\mu}_c(\mathbf{b})) = \mathbf{b}$ for any $\mathbf{b} \in \mathcal{B}_{d_0}$. Note that $\mathbf{Z}^\top \exp\{\mathbf{Z}^\top\boldsymbol{\beta}(\tau)\}\mathbf{B}^{-1}(\boldsymbol{\beta}(\tau)) = \frac{d\exp(\mathbf{Z}^\top h\{\boldsymbol{\mu}_c\{\boldsymbol{\beta}(\tau)\}\}}{d\boldsymbol{\mu}_c\{\boldsymbol{\beta}(\tau)\}} = \frac{d\exp\{\mathbf{Z}^\top\boldsymbol{\beta}(\tau)\}}{d\boldsymbol{\mu}_c\{\boldsymbol{\beta}(\tau)\}}$. From additional condition (C3), we have that $\sup_{\mathbf{Z},\tau\in(0,\tau_U]} \|\exp\{\mathbf{Z}^\top h(\boldsymbol{\alpha}(\tau))\}\mathbf{B}^{-1}(h\{\boldsymbol{\alpha}(\tau)\})\|$ is bounded, and $\tilde{f}(t|\mathbf{Z})$ is uniformly bounded in $t$ and $\mathbf{Z}$. Under the boundness of $\mathbf{Z}$, there exists a constant $C_7$, such that

$$
2\sup_t |\tilde{f}(t|\mathbf{Z})| \sup \|\mathbf{Z}^\top\| \sup_{\mathbf{Z},\tau\in(0,\tau_U]} \|\exp\{\mathbf{Z}^\top\boldsymbol{\beta}(\tau)\}\mathbf{B}^{-1}(\boldsymbol{\beta}(\tau))\| \leq C_7.
$$

.

Then

$$
\begin{aligned}
&2\sup_t |\tilde{f}(t|\mathbf{Z})|\,|\exp\{\mathbf{Z}^\top\tilde{\boldsymbol{\beta}}_n(\tau)\}-\exp\{\mathbf{Z}^\top\boldsymbol{\beta}_0(\tau)\}| \\
\le& 2\sup_t |\tilde{f}(t|\mathbf{Z})|\sup\|\mathbf{Z}^\top\|\sup_{\mathbf{Z},\tau\in(0,\tau_U]}\|\exp\{\mathbf{Z}^\top\check{\boldsymbol{\beta}}_n(\tau)\})\}\mathbf{B}^{-1}\{\check{\boldsymbol{\beta}}_n(\tau)\}\}\| \\
&\cdot\sup_{\tau\in(0,\tau_U]}\|\boldsymbol{\mu}_c\{\tilde{\boldsymbol{\beta}}_n(\tau)\}-\boldsymbol{\mu}_c\{\boldsymbol{\beta}_0(\tau)\}\| \\
\le& C_7\sup_{\tau\in(0,\tau_U]}\|\boldsymbol{\mu}_c\{\tilde{\boldsymbol{\beta}}_n(\tau)\}-\boldsymbol{\mu}_c\{\boldsymbol{\beta}_0(\tau)\}\|.
\end{aligned}
$$

Since $n^{1/4}\sup_{\tau\in(0,\tau_U]}\|\boldsymbol{\mu}_c(\tilde{\boldsymbol{\beta}}_n(\tau))-\boldsymbol{\mu}_c(\boldsymbol{\beta}_0(\tau))\|\to_p 0$, then $\sup_{\tau\in(0,\tau_U]}E(O^{(2)}(\tau))=o(n^{-1/4})$. It follows that $\sup_{\tau\in(0,\tau_U]}\|\mathrm{II}_n(\tau)\|\to_p 0$. Therefore, we complete the proof of Lemma 3.2. □

**Lemma 3.3.** *Under conditions (C1)-(C8),*

$$
\sup_{\tau\in(0,\tau_U]}\|n^{1/2}\mathbf{S}_n(\boldsymbol{\beta}_0,\tau)-n^{-1/2}\sum_{i=1}^n\boldsymbol{\psi}_i(\tau)\|=o_p(1)
$$

*where* $\boldsymbol{\psi}_i(\tau)=\mathbf{m}(\mathbf{U}_i,\tau)\left(1-\frac{D_i(1-V_i)}{1-\pi_i}-\frac{(1-D_i)V_i}{\pi_i}\right)+\mathbf{H}(\mathbf{X}_i,\tau)(V_i-\pi(\mathbf{X}_i))$.

*Proof.* By condition (C6), using $\hat{\kappa}_{v,i}$ in place of $\tilde{\kappa}_{v,i}$ in $\mathbf{S}_n(\boldsymbol{\beta},\tau)$ only lead to a difference of $o_{(0,\tau_U]}(1)$. Thus, after some algebra manipulations, we have

$$
\begin{aligned}
n^{1/2}\mathbf{S}_n(\boldsymbol{\beta}_0(\tau))=&\frac{1}{\sqrt{n}}\sum_{i=1}^n\mathbf{m}(\mathbf{U}_i,\tau)\cdot\left(1-\frac{D_i(1-\hat{v}_i)}{1-\pi_i}-\frac{(1-D_i)\hat{v}_i}{\pi_i}\right)\\
&+\frac{1}{\sqrt{n}}\sum_{i=1}^n\mathbf{m}(\mathbf{U}_i,\tau)\cdot\left(\frac{(1-D_i)\cdot v_i}{\pi_i^2}-\frac{D_i\cdot(1-v_i)}{(1-\pi_i)^2}\right)\cdot(\hat{\pi}_i-\pi_i)\\
&+\mathbf{R}_{n,1}+\mathbf{R}_{n,2}
\end{aligned}
\tag{3.17}
$$

where $\mathbf{R}_{n,1}=\frac{1}{n}\sum_{i=1}^n\mathbf{m}(\mathbf{U}_i,\tau)\cdot\left(\frac{(1-D_i)}{\pi_i\hat{\pi}_i}-\frac{D_i}{(1-\pi_i)(1-\hat{\pi}_i)}\right)\cdot n^{1/4}(\hat{\pi}_i-\pi_i)\cdot n^{1/4}(\hat{v}_i-v_i)$ and $\mathbf{R}_{n,2}=\frac{1}{n}\sum_{i=1}^n\mathbf{m}(\mathbf{U}_i,\tau)\cdot\left(\frac{(1-D_i)\cdot v_i}{\pi_i^2\hat{\pi}_i}-\frac{D_i\cdot(1-v_i)}{(1-\pi_i)^2(1-\hat{\pi}_i)}\right)\cdot n^{1/2}(\hat{\pi}_i-\pi_i)^2$.

Applying Lemma B.3 in Newey (1994) with condition (C8), we have $\sup_i \|\hat{\pi}_i - \pi_i\| = o(n^{-1/4})$ and $\sup_i \|\hat{v}_i - v_i\| = o(n^{-1/4})$. Thus, we have $\|\mathbf{R}_{n,1}\| \leq n^{1/4} \sup \|\hat{\pi} - \pi\| n^{1/4} \sup \|\hat{v} - v\| \frac{1}{n} \sum_{i=1}^n \|\mathbf{m}(\mathbf{U}_i, \tau) \cdot \left( \frac{(1-D_i)}{\pi_i \hat{\pi}_i} - \frac{D_i}{(1-\pi_i)(1-\hat{\pi}_i)} \right) \| = o_{(0,\tau_U]}(1)$. Similarly $\mathbf{R}_{n,2} = o_{(0,\tau_U]}(1)$.

Next, we show that

$$\begin{aligned} &\frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbf{m}(\mathbf{U}_i, \tau) \cdot \left( \frac{(1-D_i) \cdot v_i}{\pi_i^2} - \frac{D_i \cdot (1-v_i)}{(1-\pi_i)^2} \right) \cdot (\hat{\pi}_i - \pi_i) \\ &\quad = \frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbf{H}(\mathbf{X}_i, \tau) \cdot (V_i - \pi) + o_{(0,\tau_U]}(1), \end{aligned} \tag{3.18}$$

and

$$\begin{aligned} &\frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbf{m}(\mathbf{U}_i, \tau) \cdot \left( 1 - \frac{D_i(1-\hat{v}_i)}{1-\pi_i} - \frac{(1-D_i)\hat{v}_i}{\pi_i} \right) \\ &\quad = \frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbf{m}(\mathbf{U}_i, \tau) \cdot \left( 1 - \frac{D_i(1-V_i)}{1-\pi_i} - \frac{(1-D_i)V_i}{\pi_i} \right) + o_{(0,\tau_U]}(1). \end{aligned} \tag{3.19}$$

Note that directly applying Theorem 4.2 in Newey (1994) can lead to the pointwise counterparts of (3.18) and (3.19), where $o_{(0,\tau_U]}(1)$ is replaced by $o_p(1)$. The assumptions required by Theorem 4.2 in Newey (1994) are ensured by the fact that $V$ is a binary variable, and conditions (C1), (C7) and (C8).

To establish the results in (3.18) and (3.19), we can follow the same lines in the proof of Newey (1994)'s Theorem 4.2, while applying some strengthened arguments at various steps. First, consider (3.18), and without loss of generality, confine our attention to the second component of both sides of the equation, (i.e, corresponding to the constant component in $\mathbf{Z}$). Following the proofs of Newey (1994)'s Theorem 4.2, we see that the application of his Lemma 5.4 can lead to a stronger conclusion with $o_p(1)$ replaced by $o_{(0,\tau_U]}(1)$ provided $|a^{(2)}(\mathbf{U}, \tau)|$ is uniformly bounded in $\mathbf{u} \in \mathcal{U}$ and $\tau \in (0, \tau_U]$. In order to reach a stronger uniform version of the conclusion from

the subsequent application of his Lemma 5.2, we need to show that

$$n^{-1/2}\sum_{i=1}^{n}\{\varepsilon_i^{\sigma}(\tau)-E(\varepsilon_i^{\sigma}(\tau))\}=o_{(0,\tau_U]}(1), \tag{3.20}$$

where $\varepsilon_i^{\sigma}(\tau)=\zeta_i^{\sigma}(\tau)-\zeta_i(\tau)$ with $\zeta_i^{\sigma}(\tau)=\int H^{(2)}(\mathbf{x},\tau)\{V_i-\pi(\mathbf{x})\}\mathcal{K}_{\sigma}^{*}(\mathbf{x}-\mathbf{X}_i)d\mathbf{x}$ and $\zeta_i^{\sigma}(\tau)=H^{(2)}(\mathbf{x},\tau)(V_i-\pi_i)$. By $\mathcal{K}_{\sigma}^{*}(\cdot)$ having a bounded support and given the uniform boundeness of $|a^{(2)}(\mathbf{U},\tau)|$ and condition (C5), we have $|\zeta_i^{\sigma}(\tau)|$ is bounded above by a positive constant free of $\sigma$. By the Dominated Convergence Theorem, for each $\tau\in(0,\tau_U]$, we have $\zeta_i^{\sigma}(\tau)\to_p\zeta_i(\tau)$ as $\sigma\to 0$ and consequently, $E(\{\varepsilon_i^{\sigma_{1,n}}(\tau)\}^2)\to 0$ as $n\to\infty$. By condition (C2), we can show that $E(\{\varepsilon_i^{\sigma_{1,n}}(\tau)\}^2)$ is equicontinuous in $\tau$. Then it follows from Arzelà-Ascoli Theorem that $\sup_{\tau\in(0,\tau_U]}E(\{\varepsilon_i^{\sigma_{1,n}}(\tau)\}^2)\to 0$, which implies that $\sup_{\tau\in(0,\tau_U]}(n^{-1}\sum_{i=1}^{n}\mathrm{Var}\{\varepsilon_i^{\sigma_{1,n}}(\tau)\})$ can be bounded above by a sequence converging to 0. Following similar arguments of Lai and Ying (1988) for their Lemma 1, we can show that $\sup_{\tau\in(0,\tau_U]}E(\{\varepsilon_i^{\sigma_{1,n}}(\tau)\}^2)\to 0$ implies the result in (3.20) after checking that the entropy assumption (2.4) in Lai and Ying (1988) is satisfied for the function class $\{\varepsilon_i^{\sigma}(\tau):\tau\in(0,\tau_U]\}$. Upon proving (3.20), we can achieve a modified version of Newey (1994)'s proof for his Theorem 4.2, which lead to the stronger uniform result in (3.18).

Similarly, an application of Newey (1994)'s Theorem 4.2 to $\frac{1}{\sqrt{n}}\sum_{k=1}^{n}\mathbf{m}_{ij}(\mathbf{U}_k,\tau)\cdot\left(1-\frac{D_k(1-\hat{v}_{ij,k})}{1-\pi_k}-\frac{(1-D_k)\hat{v}_{ij,k}}{\pi_k}\right)$ enlightens the following result:

$$
\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{k=1}^{n}\mathbf{m}_{ij}(\mathbf{U}_k,\tau)\cdot\left(1-\frac{D_k(1-\hat{v}_{ij,k})}{1-\pi_k}-\frac{(1-D_k)\hat{v}_{ij,k}}{\pi_k}\right)\\
=&\frac{1}{\sqrt{n}}\sum_{k=1}^{n}\mathbf{m}_{ij}(\mathbf{U}_k,\tau)\cdot\left(1-\frac{D_k(1-v_{ij,k})}{1-\pi_k}-\frac{(1-D_k)v_{ij,k}}{\pi_k}\right)\\
&+\frac{1}{\sqrt{n}}\sum_{k=1}^{n}\mathbf{m}_{ij}(\mathbf{U}_k,\tau)\cdot\left(1+\frac{D_k}{1-\pi_k}-\frac{(1-D_k)}{\pi_k}\right)(V_k-v_{ij}(\mathbf{U}_k))+o_{(0,\tau_U]}(1)\\
=&\frac{1}{\sqrt{n}}\sum_{k=1}^{n}\mathbf{m}_{ij}(\mathbf{U}_k,\tau)\cdot\left(1-\frac{D_k(1-V_k)}{1-\pi_k}-\frac{(1-D_k)V_k}{\pi_k}\right)+o_{(0,\tau_U]}(1).
\end{aligned}
\tag{3.21}
$$

We can prove the results in (3.21) using the same strategy for showing (3.18). Summing equation (3.21) for all $i, j$ on both sides, we obtain (3.19).

By (3.17),(3.18) and (3.19), and $\mathbf{R}_{n,k} = o_{(0,\tau_U]}(1)(k = 1, 2)$, it follows that

$$
n^{1/2}\mathbf{S}_n(\boldsymbol{\beta}_0(\tau)) = n^{-1/2}\sum_{i=1}^{n}\boldsymbol{\psi}_i(\tau) + o_{(0,\tau_U]}(1).
$$

This completes the proof of Lemma 3.3. □

#### 3.5.2.2 Proof of theorems

**Proof of Theorem 3.1:** Define $\boldsymbol{\gamma}_{n,j} = n^{-1}\sum_{i=1}^{n}(\hat{k}_{v,i} - k_{v,i})\mathbf{Z}_i[N_i(\exp\{\mathbf{Z}_i^\top\hat{\boldsymbol{\beta}}(\tau_j)\} - \int_0^{\tau_j} I(W_i \geq \exp\{\mathbf{Z}_i^\top\hat{\boldsymbol{\beta}}(u)\})dH(u)], j = 1, \ldots L(n)$. Under conditions (C7) and (C8), Lemma B.3 of (Newey, 1994) implies $\sup_{\mathbf{u}\in\mathcal{U}}|\hat{v}(\mathbf{u}) - v(u)| = o_p(n^{-1/4})$ and $\sup_{\mathbf{x}\in\mathcal{X}}|\hat{\pi}(\mathbf{x}) - \pi(\mathbf{x})| = o_p(n^{-1/4})$. It follows from condition (C6) that

$$
\sup_i|\tilde{\kappa}_{v,i} - \kappa_{v,i}| = o_p(n^{-1/4}). \tag{3.22}
$$

Coupled with the boundedness of $\mathbf{Z}$, we have $\sup_{1\leq j\leq L(n)}\|\boldsymbol{\gamma}_{n,j}\| = o(1), a.s.$

Given that $\hat{\boldsymbol{\beta}}(\tau)$ is defined as the generalized solution of (3.2), we have

$$n^{-1}\sum_{i=1}^{n}\hat{\kappa}_{v,i}\mathbf{Z}_i N_i(\exp\{\mathbf{Z}_i^T\hat{\boldsymbol{\beta}}(\tau_j)\}) = n^{-1}\sum_{i=1}^{n}\int_0^{\tau_j}\hat{\kappa}_{v,i}\mathbf{Z}_i I(W_i \geq \exp\{\mathbf{Z}_i^\top\hat{\boldsymbol{\beta}}(u)\})dH(u)+\boldsymbol{\xi}_{n,j}$$

where $\max_{j=1,\cdots,M}\|\boldsymbol{\xi}_{n,j}\| \leq \sup_i\|\hat{k}_{v,i}\|\sup_i\|\mathbf{Z}_i\|/n$.

Condition C3(b) implies the above equation holds if and only if $\boldsymbol{b}=\boldsymbol{b}'$. Therefore, there exists an inverse function of $\boldsymbol{\mu}_c$, denoted as $\boldsymbol{\kappa}_c$ from $\mathcal{A}(d_0)$ to $\mathcal{B}(d_0)$, such that $\boldsymbol{\kappa}_c(\boldsymbol{\mu}_c(\boldsymbol{b}))=\boldsymbol{b}$ for any $\boldsymbol{b}\in\mathcal{B}(d_0)$.

According to definition of generalized solution, for $j=1,\cdots,L$, we have

$$n^{-1}\sum_{i=1}^{n}\hat{k}_{v,i}\mathbf{Z}_i N_i(\exp\{\mathbf{Z}_i^T\hat{\boldsymbol{\beta}}(\tau_j)\}) = n^{-1}\sum_{i=1}^{n}\int_0^{\tau_j}\hat{k}_{v,i}\mathbf{Z}_i I(W_i \geq \exp\{\mathbf{Z}_i^\top\hat{\boldsymbol{\beta}}(u)\})dH(u)+\boldsymbol{\xi}_{n,j}$$

where $\max_{j=1,\cdots,M}\|\boldsymbol{\xi}_{n,j}\| \leq \sup_i\|\hat{k}_{v,i}\|\sup_i\|\mathbf{Z}_i\|/n$. After simple algebra manipulations, it could be shown

$$\begin{aligned}\boldsymbol{\mu}_c(\hat{\boldsymbol{\beta}}(\tau_j)) - \boldsymbol{\mu}_c(\boldsymbol{\beta}_0(\tau_j)) =& -\boldsymbol{\nu}_n(\hat{\boldsymbol{\beta}}(\tau_j)) + \int_0^{\tau_j}\tilde{\boldsymbol{v}}_n(\hat{\boldsymbol{\beta}}(u))dH(u)\\ &+\sum_{k=1}^{j}\int_{\tau_{k-1}}^{\tau_k}[\tilde{\boldsymbol{\mu}}_c(\hat{\boldsymbol{\beta}}(\tau_j)) - \tilde{\boldsymbol{\mu}}_c(\hat{\boldsymbol{\beta}}_0(u))]dH(u) + \boldsymbol{\xi}_{n,j} - \boldsymbol{\gamma}_{n,j},\end{aligned} \tag{3.23}$$

and $\sup_j\|\boldsymbol{\xi}_{n,j}-\boldsymbol{\gamma}_{n,j}\| = o(1)$, a.s.

Note that equation (3.23) closely resembles to the equation (A.1) in Peng and Huang (2008). The key distinction lies in the inclusion $I(D_1 > D_0)$ (or $P(D_1 > D_0|\mathbf{U})$) within the expectation (or empirical averages) involved in $\boldsymbol{\mu}_c(\cdot)$, $\tilde{\boldsymbol{\mu}}_c(\cdot)$, and $\boldsymbol{\nu}_n(\cdot)$, and $\tilde{\boldsymbol{\nu}}_n(\cdot)$. Therefore, based on equation (3.23), we can apply the same line of arguments in the proof of Theorem 1 of Peng and Huang (2008) to show that for $0<v<\tau_U$,

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

**Proof of Theorem 3.2:** By Lemma 3.1 and Lemma 3.1, we have

$$\begin{aligned}\sup_{\tau\in(0,\tau_U]} \|n^{-1/2}\sum_{i=1}^{n}\hat{k}_{v,i}\mathbf{Z}_i[N_i(\exp\{\mathbf{Z}_i^\top\hat{\boldsymbol{\beta}}(\tau)\}) - N_i(\exp\{\mathbf{Z}_i^\top\boldsymbol{\beta}_0(\tau)\})] \\ - n^{1/2}[\boldsymbol{\mu}_c\{\hat{\boldsymbol{\beta}}(\tau)\} - \boldsymbol{\mu}_c\{\boldsymbol{\beta}_0(\tau)\}]\| \to_p 0.\end{aligned} \tag{3.24}$$

Similarly, we could have

$$\begin{aligned}\sup_{\tau\in(0,\tau_U]} \|n^{-1/2}\sum_{i=1}^{n}\hat{k}_{v,i}\mathbf{Z}_i[I(W_i \geq \exp\{\mathbf{Z}_i^\top\hat{\boldsymbol{\beta}}(\tau)\}) - I(W_i \geq \exp\{\mathbf{Z}_i^\top\boldsymbol{\beta}_0(\tau)\})] \\ - n^{1/2}[\tilde{\boldsymbol{\mu}}_c\{\hat{\boldsymbol{\beta}}(\tau)\} - \tilde{\boldsymbol{\mu}}_c\{\boldsymbol{\beta}_0(\tau)\}]\| \to_p 0.\end{aligned} \tag{3.25}$$

By the definition of $\hat{\boldsymbol{\beta}}(\cdot)$,

$$\begin{aligned}&\sup_{\tau\in[\tau_j,\tau_{j+1}]} n^{1/2}\|\mathbf{S}_n(\hat{\boldsymbol{\beta}},\tau) - \mathbf{S}_n(\hat{\boldsymbol{\beta}},\tau_j)\| \\ &\leq n^1/2C_1\sup_{\mathbf{u}\in\mathcal{U}}|\hat{\kappa}_v|\{H(\tau_{j+1}) - H(\tau_j)\} \\ &\leq n^{1/2}C_1 a_n/(1-\tau_U).\end{aligned} \tag{3.26}$$

Given that $n^{1/2}\|S_L\| \to 0$, we have $n^{1/2}\mathbf{S}_n(\hat{\boldsymbol{\beta}},\tau) = o_{(0,\tau_U]}(1)$.

By equations (3.24) and (3.25), and the result that $\boldsymbol{\mu}_c\{\hat{\boldsymbol{\beta}}(\tau)\}$ converges uniformly to $\boldsymbol{\mu}_c\{\boldsymbol{\beta}_0(\tau)\}$ for $\tau \in (0,\tau_U]$, we have

$$\begin{aligned}&- n^{1/2}\mathbf{S}_n(\boldsymbol{\beta}_0,\tau) \\ =&n^{1/2}[\boldsymbol{\mu}_c\{\hat{\boldsymbol{\beta}}(\tau)\} - \boldsymbol{\mu}_c\{\boldsymbol{\beta}_0(\tau)\}] \\ &- \int_0^\tau n^{1/2}[\tilde{\boldsymbol{\mu}}_c\{\boldsymbol{\beta}(u)\} - \tilde{\boldsymbol{\mu}}_c\{\boldsymbol{\beta}_0(u)\}]dH(u) + o_{(0,\tau_U]}(1) \\ =& n^{1/2}[\boldsymbol{\mu}_c\{\hat{\boldsymbol{\beta}}(\tau)\} - \boldsymbol{\mu}_c\{\boldsymbol{\beta}_0(\tau)\}] \\ &- \int_0^\tau [\mathbf{J}\{\boldsymbol{\beta}_0(u)\}\mathbf{B}\{\boldsymbol{\beta}_0(u)\}^{-1} + o_{(0,\tau_U]}(1)] \\ &\times n^{1/2}[\boldsymbol{\mu}_c\{\hat{\boldsymbol{\beta}}(u)\} - \boldsymbol{\mu}_c\{\boldsymbol{\beta}_0(u)\}]dH(u) + o_{(0,\tau_U]}(1).\end{aligned} \tag{3.27}$$

The above equation can be viewed as a stochastic differential equation for $n^{1/2}[\boldsymbol{\mu}_c\{\hat{\boldsymbol{\beta}}(\tau)\}-\boldsymbol{\mu}_c\{\boldsymbol{\beta}_0(\tau)\}]$. From the production integration theory (Gill et al., 1990), we have

$$n^{1/2}[\boldsymbol{\mu}_c\{\hat{\boldsymbol{\beta}}(\tau)\}-\boldsymbol{\mu}_c\{\boldsymbol{\beta}_0(\tau)\}]=\boldsymbol{\phi}\{n^{1/2}\mathbf{S}_n(\boldsymbol{\beta}_0,\tau)\}+o_{(0,\tau_U]}(1) \tag{3.28}$$

where $\boldsymbol{\phi}$ is a map from $\mathcal{F}$ to $\mathcal{F}$ such that for $\boldsymbol{g}\in\mathcal{F}$,

$$\boldsymbol{\phi}(\boldsymbol{g})(\tau)=\int_v^\tau \mathcal{I}(s,\tau)d\boldsymbol{g}(s)$$

with $\mathcal{I}(s,\tau)=\prod_{u\in(s,t]}[\mathcal{I}_p+\mathbf{J}\{\boldsymbol{\beta}_0(u)\}\mathbf{B}\{\boldsymbol{\beta}_0(u)\}^{-1}dH(u)]$ and $\mathcal{F}=\{\boldsymbol{g}:(0,\tau_U],\rightarrow \mathcal{R}^p,\boldsymbol{g}$ is left-continuous with right limit, $\boldsymbol{g}(v)=\mathbf{0}\}$

From Lemma 3.3, $-n^{1/2}\mathbf{S}_n(\boldsymbol{\beta}_0,\tau)$ converge weakly to a tight Gaussian process, $\boldsymbol{G}(\tau)$, with mean 0 and covariance matrix $\boldsymbol{\Sigma}(s,t)=\mathrm{E}[\boldsymbol{\psi}_j(s)\boldsymbol{\psi}_j(t)^\top]$, where $\boldsymbol{\psi}_j(s)$ is defined in Lemma B.3. Besides, $\boldsymbol{\phi}$ is a linear operator. $\boldsymbol{\phi}\{\boldsymbol{G}(\tau)\}$ for $\tau\in[v,\tau_U]$ is Gaussian process, and $B\{\boldsymbol{\beta}_0(u)\}^{-1}$ is bounded uniformly for $\tau\in[v,\tau_U]$. Following the arguments in Peng and Huang (2008), it could be shown that $n^{1/2}\{\hat{\boldsymbol{\beta}}(\tau)-\boldsymbol{\beta}_0(\tau)\}$ converges weakly to $B\{\boldsymbol{\beta}_0(\tau)\}^{-1}\boldsymbol{\phi}\{\boldsymbol{G}(\tau)\}$ for $\tau\in[v,\tau_u]$ by using the Taylor expansion of $\boldsymbol{\kappa}_c[\boldsymbol{\mu}_c\{\hat{\boldsymbol{\beta}}(\tau)\}]-\boldsymbol{\kappa}_c[\boldsymbol{\mu}_c\{\boldsymbol{\beta}_0(\tau)\}]$ and continuous mapping theorem. Since $B\{\boldsymbol{\beta}_0(\tau)\}^{-1}\boldsymbol{\phi}\{\boldsymbol{G}(\tau)\}$ is a Gaussian process, we complete the proof of Theorem 3.2.

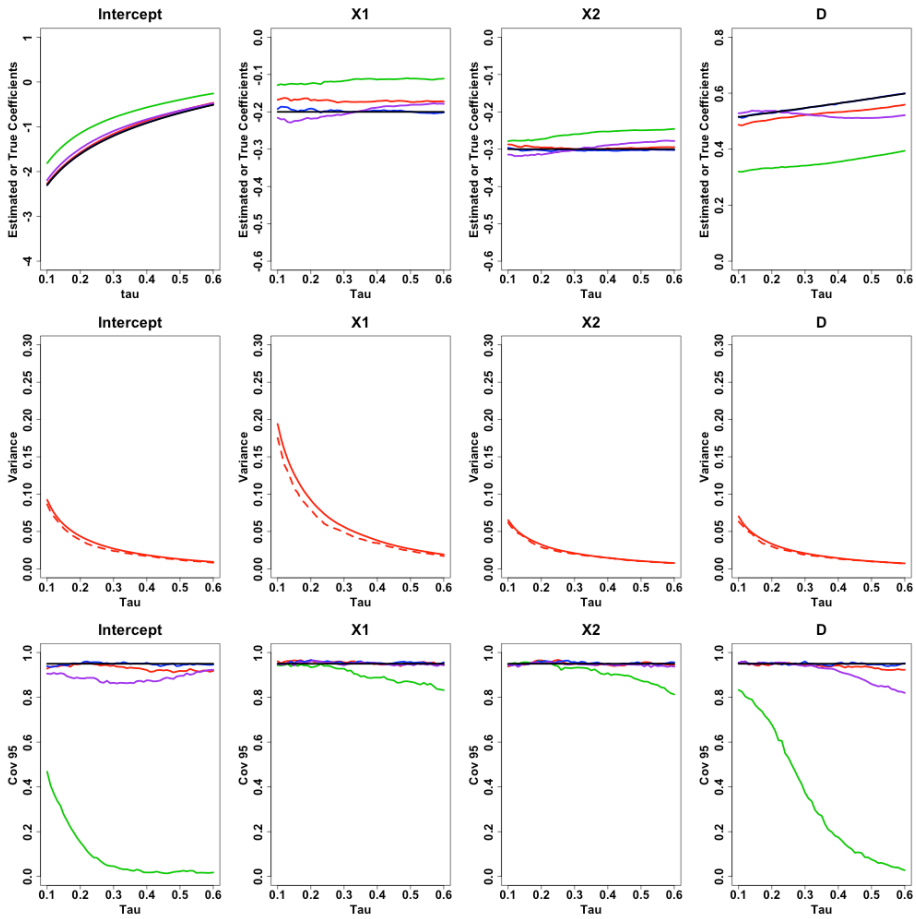


Figure 3.2: The simulation results in scenario (B) with sample size n=1000. In the first and the third rows, black lines correspond to true or nominal values; red lines correspond to the proposed method; blue lines correspond to the benchmark method; green lines correspond to the naive as-treated method; purple lines correspond to a modified proposed method with weights estimated by logistic regression. In the second row, red solid lines represent the average variance estimates from the proposed method, and red dashed lines represent the empirical variances of the proposed estimator.

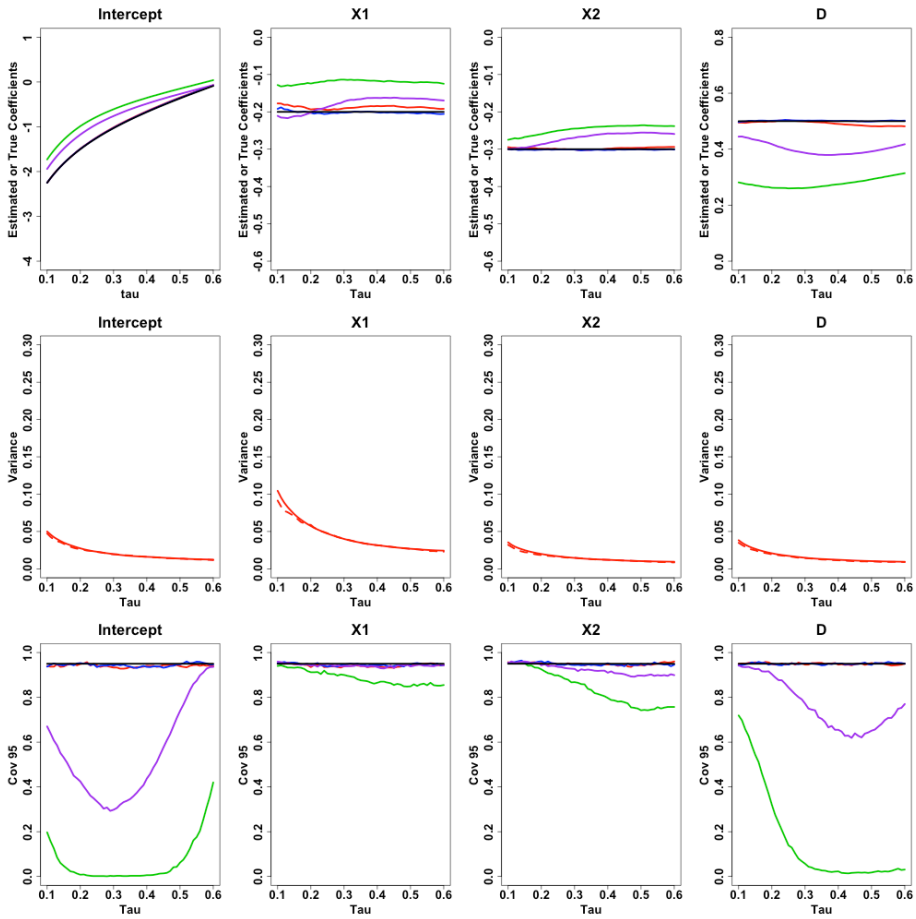


Figure 3.3: The simulation results in scenario (A) with sample size n=2000. In the first and the third rows, black lines correspond to true or nominal values; red lines correspond to the proposed method; blue lines correspond to the benchmark method; green lines correspond to the naive as-treated method; purple lines correspond to a modified proposed method with weights estimated by logistic regression. In the second row, red solid lines represent the average variance estimates from the proposed method, and red dashed lines represent the empirical variances of the proposed estimator.

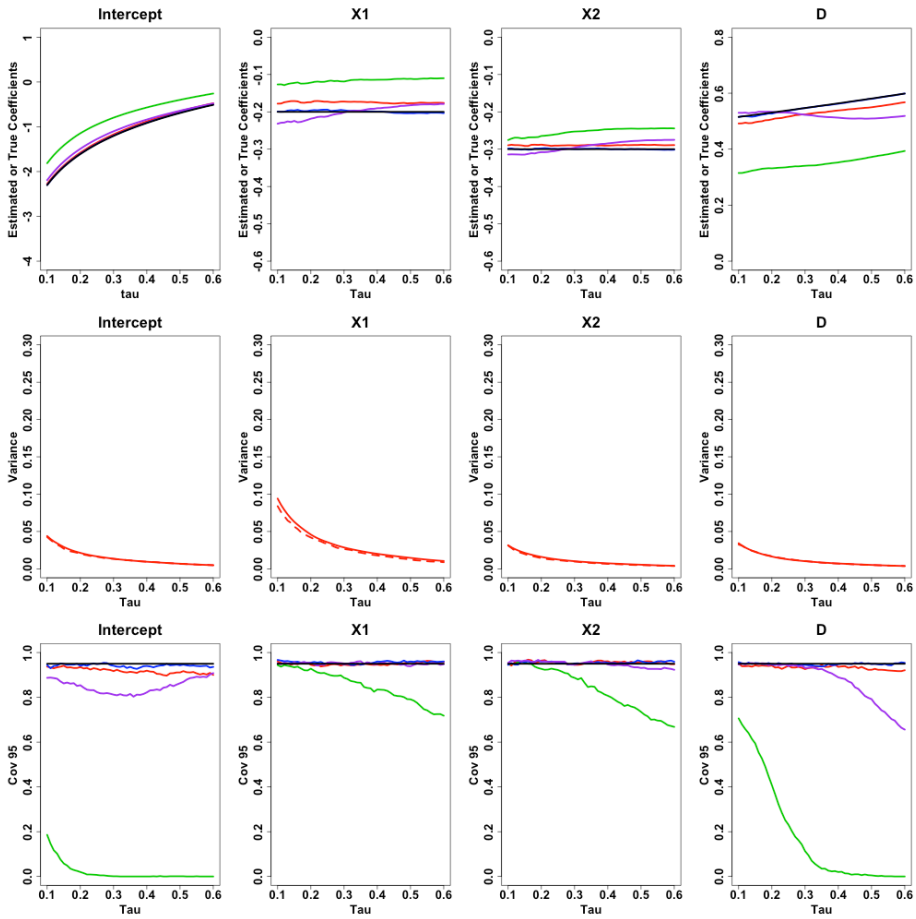


Figure 3.4: The simulation results in scenario (B) with sample size n=2000. In the first and the third rows, black lines correspond to true or nominal values; red lines correspond to the proposed method; blue lines correspond to the benchmark method; green lines correspond to the naive as-treated method; purple lines correspond to a modified proposed method with weights estimated by logistic regression. In the second row, red solid lines represent the average variance estimates from the proposed method, and red dashed lines represent the empirical variances of the proposed estimator.

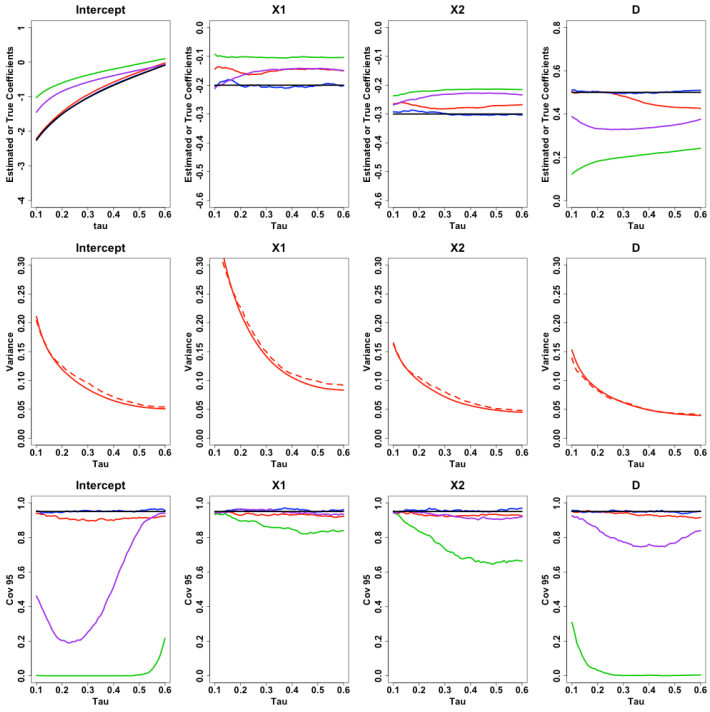


Figure 3.5: The simulation results in scenario (C) with sample size n=1000. In the first and the third rows, black lines correspond to true or nominal values; red lines correspond to the proposed method; blue lines correspond to the benchmark method; green lines correspond to the naive as-treated method; purple lines correspond to a modified proposed method with weights estimated by logistic regression. In the second row, red solid lines represent the average variance estimates from the proposed method, and red dashed lines represent the empirical variances of the proposed estimator.

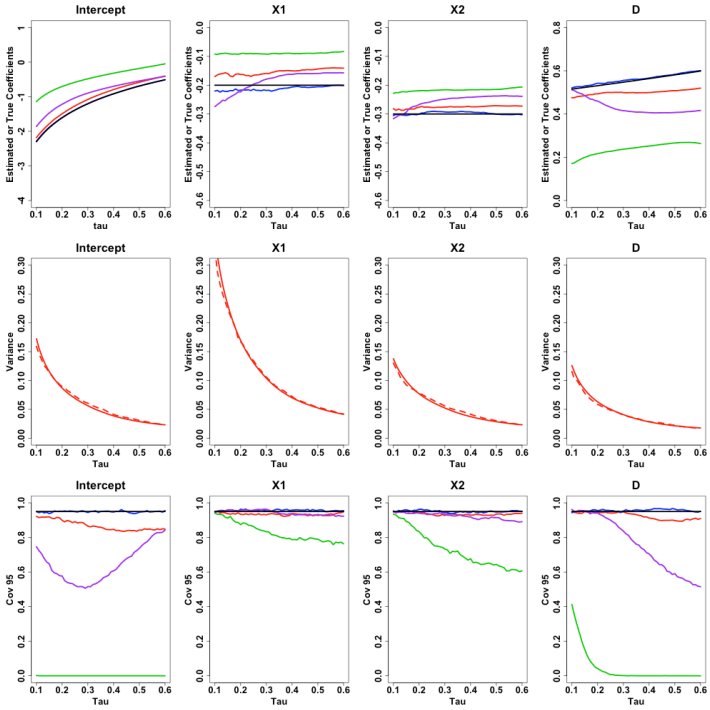


Figure 3.6: The simulation results in scenario (D) with sample size n=1000. In the first and the third rows, black lines correspond to true or nominal values; red lines correspond to the proposed method; blue lines correspond to the benchmark method; green lines correspond to the naive as-treated method; purple lines correspond to a modified proposed method with weights estimated by logistic regression. In the second row, red solid lines represent the average variance estimates from the proposed method, and red dashed lines represent the empirical variances of the proposed estimator.

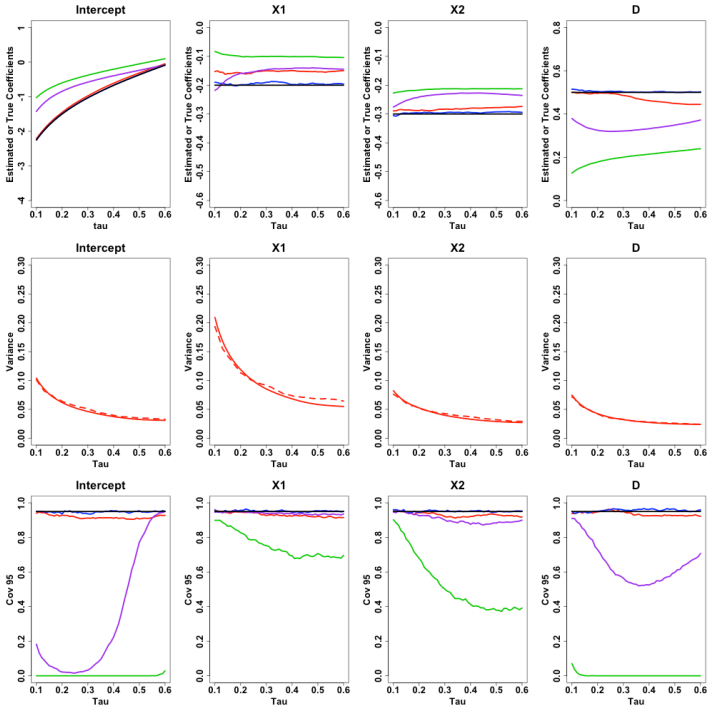


Figure 3.7: The simulation results in scenario (C) with sample size n=2000. In the first and the third rows, black lines correspond to true or nominal values; red lines correspond to the proposed method; blue lines correspond to the benchmark method; green lines correspond to the naive as-treated method; purple lines correspond to a modified proposed method with weights estimated by logistic regression. In the second row, red solid lines represent the average variance estimates from the proposed method, and red dashed lines represent the empirical variances of the proposed estimator.

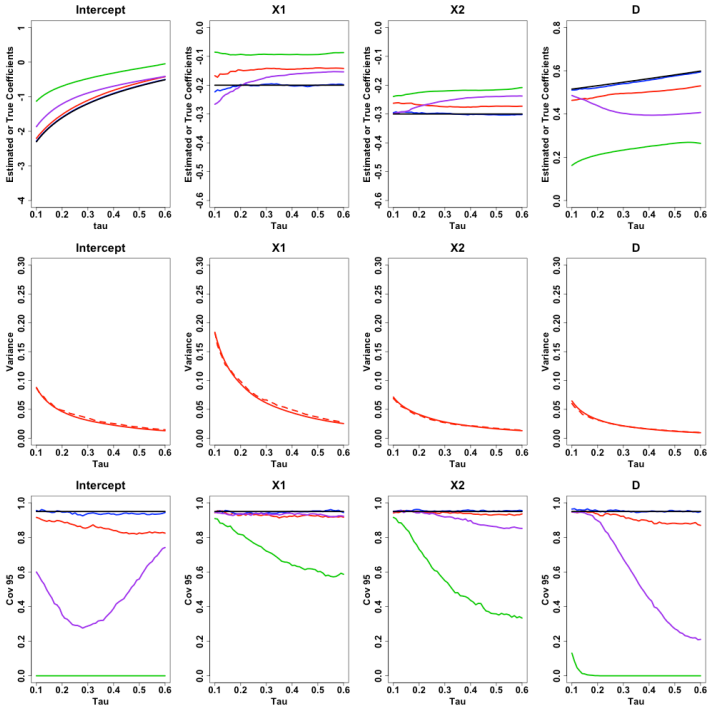


Figure 3.8: The simulation results in scenario (D) with sample size n=2000. In the first and the third rows, black lines correspond to true or nominal values; red lines correspond to the proposed method; blue lines correspond to the benchmark method; green lines correspond to the naive as-treated method; purple lines correspond to a modified proposed method with weights estimated by logistic regression. In the second row, red solid lines represent the average variance estimates from the proposed method, and red dashed lines represent the empirical variances of the proposed estimator.

Table 3.1: The descriptive statistics of patients in CIBMTR between rituximab and control groups

| Variable | | Entire Sample (N=948) | Ritux Group (N=164) | Control Group (N=784) | p-value |
|---|---|---|---|---|---|
| Age> 55 | No | 543 (57.3%) | 74 (45.1%) | 469 (59.8%) | $< 0.001$[1] |
| | Yes | 405 (42.7%) | 90 (54.9%) | 315 (40.2%) | |
| Number of chemo regimen | 1 | 128 (13.5%) | 12 (7.3%) | 116 (14.8%) | $< 0.001$[1] |
| | 2 | 413 (43.6%) | 58 (35.4%) | 355 (45.3%) | |
| | 3 | 290 (30.6%) | 61 (37.2%) | 229 (29.2%) | |
| | 4 | 96 (10.1%) | 31 (18.9%) | 65 (8.3%) | |
| | 5 | 21 (2.2%) | 2 (1.2%) | 19 (2.4%) | |
| Status | PIF sensitive | 172 (18.1%) | 31 (18.9%) | 141 (18%) | 0.006[1] |
| | PIF resistant | 51 (5.4%) | 10 (6.1%) | 41 (5.2%) | |
| | CR1 | 158 (16.7%) | 36 (22%) | 122 (15.6%) | |
| | REL sensitive | 291 (30.7%) | 43 (26.2%) | 248 (31.6%) | |
| | REL resistant | 65 (6.9%) | 12 (7.3%) | 53 (6.8%) | |
| | CR2+ | 154 (16.2%) | 32 (19.5%) | 122 (15.6%) | |
| | Other | 57 (6%) | 0 (0%) | 57 (7.3%) | |
| Karnofsky performance score | $< 90\%$ | 354 (37.3%) | 63 (38.4%) | 291 (37.1%) | 0.823[1] |
| | 90-100% | 594 (62.7%) | 101 (61.6%) | 493 (62.9%) | |
| Event | Progression or Death | 608 (64.1%) | 82 (50%) | 526 (67.1%) | 0.006[2] |
| | Censoring | 340 (35.9%) | 82 (50%) | 258 (32.9%) | |

[1] P-value is calculated from Chi-square test.
[2] P-value is calculated from Log-rank test which compares the PFS between two groups.

Table 3.2: The descriptive statistics of patients in CIBMTR before and after FDA approval date

| Variable | | Entire Sample (N=948) | Before Approval (N=314) | After Approval (N=634) | p-value |
|---|---|---|---|---|---|
| Treatment | Ritux | 164 (17.3%) | 0 (0%) | 164 (25.9%) | $< 0.001$[1] |
| | Control | 784 (82.7%) | 314 (100%) | 470 (74.1%) ) | |
| Age> 55 | No | 543 (57.3%) | 189 (60.2%) | 354 (55.8%) | 0.228[1] |
| | Yes | 405 (42.7%) | 125 (39.8%) | 280 (44.2%) | |
| Number of chemo regimen | 1 | 128 (13.5%) | 53 (16.9%) | 75 (11.8%) | 0.013[1] |
| | 2 | 413 (43.6%) | 125 (39.8%) | 288 (45.4%) | |
| | 3 | 290 (30.6%) | 98 (31.2%) | 192 (30.3%) | |
| | 4 | 96 (10.1%) | 26 (8.3%) | 70 (11%) | |
| | 5 | 21 (2.2%) | 12 (3.8%) | 9 (1.4%) | |
| Status | PIF sensitive | 172 (18.1%) | 57 (18.2%) | 115 (18.1%) | $< 0.001$[1] |
| | PIF resistant | 51 (5.4%) | 13 (4.1%) | 38 (6%) | |
| | CR1 | 158 (16.7%) | 37 (11.8%) | 121 (19.1%) | |
| | REL sensitive | 291 (30.7%) | 112 (35.7%) | 179 (28.2%) | |
| | REL resistant | 65 (6.9%) | 20 (6.4%) | 45 (7.1%) | |
| | CR2+ | 154 (16.2%) | 42 (13.4%) | 112 (17.7%) | |
| | Other | 57 (6%) | 33 (10.5%) | 24 (3.8%) | |
| Karnofsky performance score | $< 90\%$ | 354 (37.3%) | 113 (36%) | 241 (38%) | 0.592[1] |
| | 90-100% | 594 (62.7%) | 201 (64%) | 393 (62%) | |
| Event | Progression or Death | 608 (64.1%) | 239 (76.1%) | 369 (58.2%) | $< 0.001$[2] |
| | Censoring | 340 (35.9%) | 75 (23.9%) | 265 (41.8%) | |

[1] P-value is calculated from Chi-square test.
[2] P-value is calculated from Log-rank test which compares the PFS between two groups.

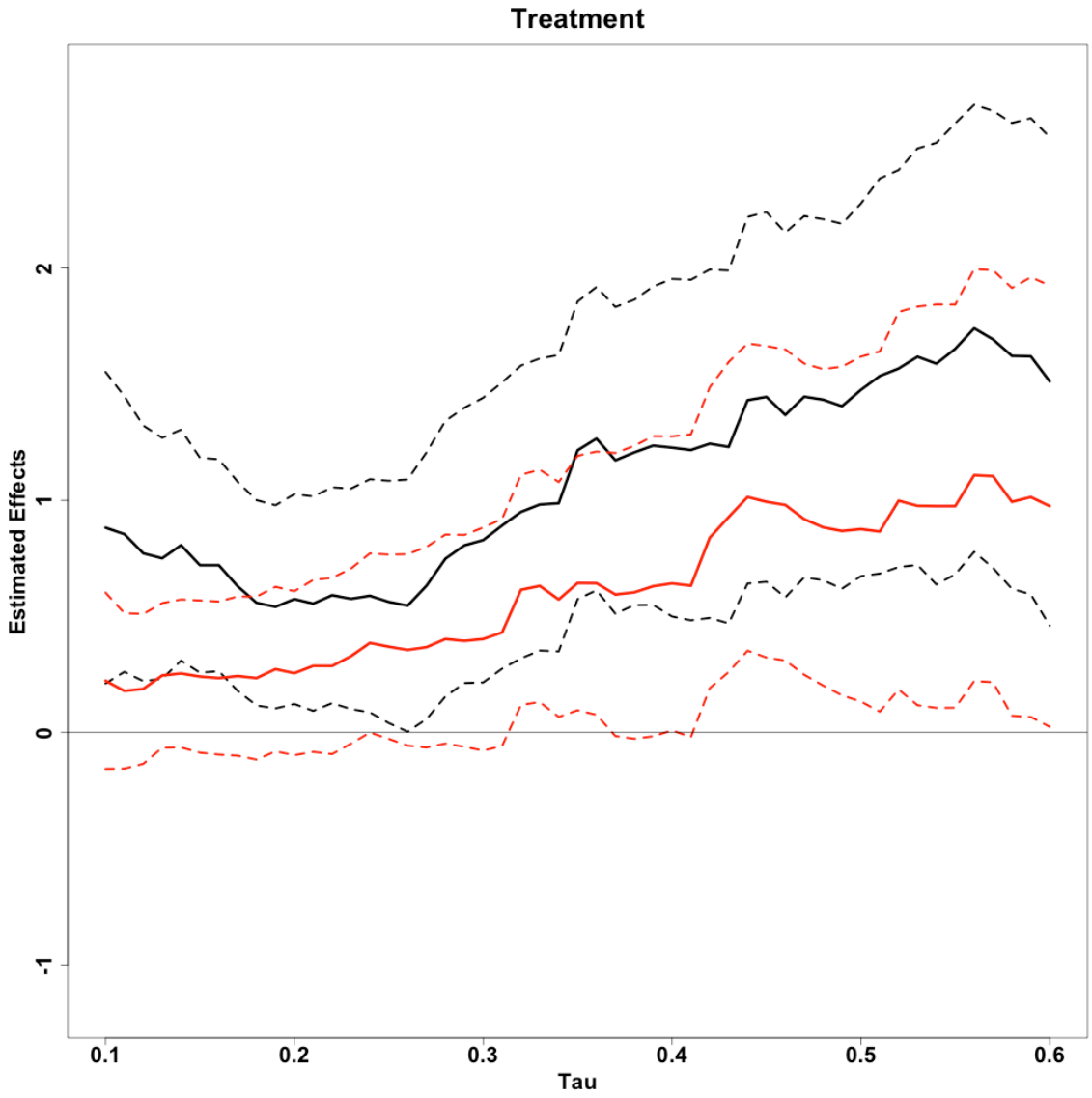


Figure 3.9: The results on treatment coefficients from analyzing the bone marrow transplant dataset based on the proposed IV method, the as-treated censored quantile regression analysis, and the modified proposed IV method with weights estimated from the logistical regression. Black solid and dashed lines respectively represent the coefficient estimates and 95% confidence intervals based on the proposed IV method. Red solid and dashed lines respectively represent the coefficient estimates and 95% confidence intervals based on the as-treated censored quantile regression analysis. Green solid and dashed lines respectively represent the coefficient estimates and 95% confidence intervals based on the modified proposed IV method with weights estimated from the logistical regression.

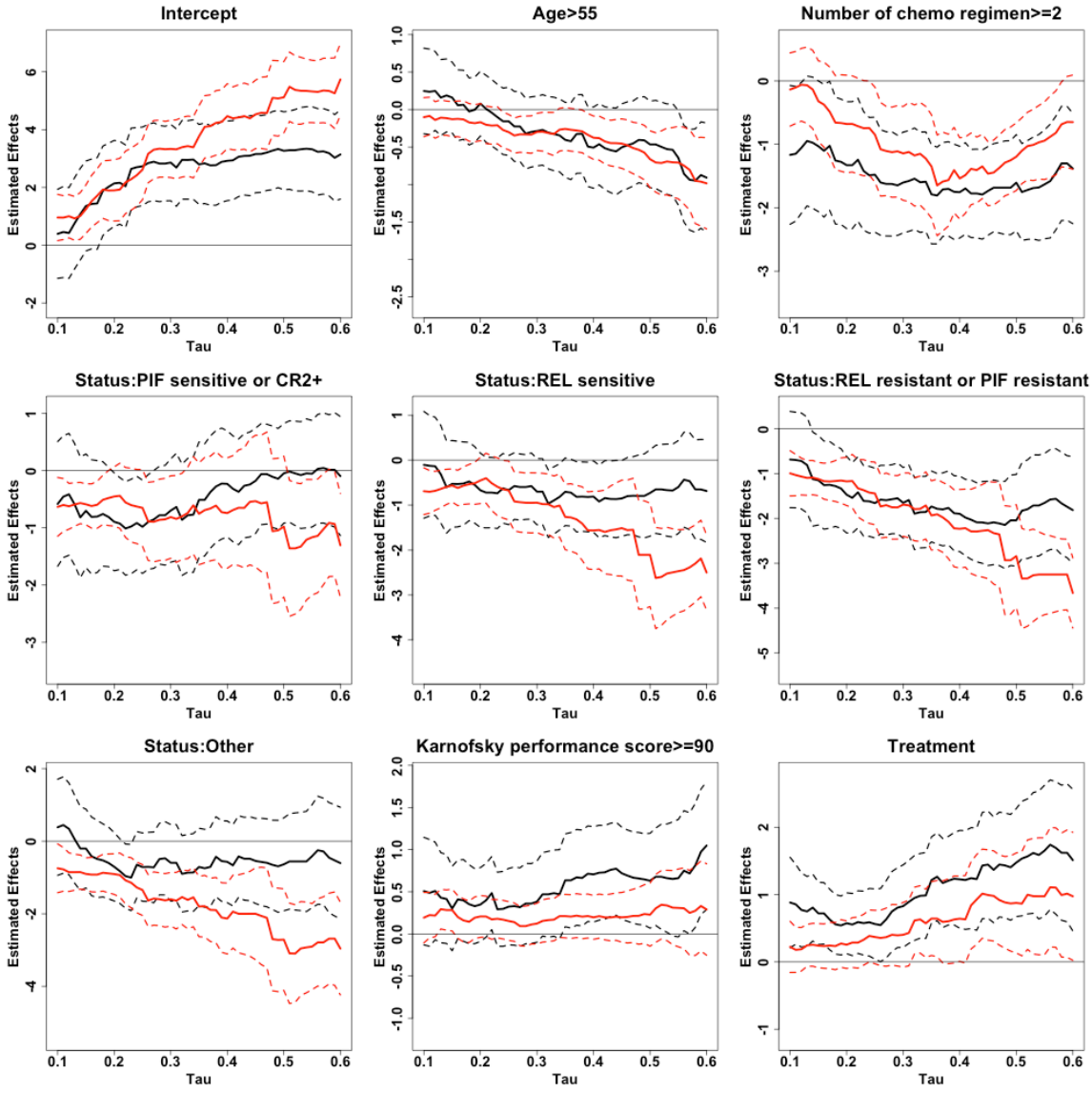


Figure 3.10: The results from analyzing CIBMTR data based on the proposed IV method, the as-treated censored quantile regression analysis, and the modified proposed IV method with weights estimated from the logistical regression. Black solid and dashed lines respectively represent the coefficient estimates and 95% confidence intervals based on the proposed IV method. Red solid and dashed lines respectively represent the coefficient estimates and 95% confidence intervals based on the as-treated censored quantile regression analysis. Green solid and dashed lines respectively represent the coefficient estimates and 95% confidence intervals based on the modified proposed IV method with weights estimated from the logistical regression.

# Chapter 4

# Estimation of Population Causal Quantile Effects with Instrumental Variables and Censored Data

## 4.1 Potential Outcomes Framework and Assumptions

Let $\mathbf{D} \in \mathbb{R}^{q_1}$ and $\mathbf{V} \in \mathbb{R}^{q_2}$ denote the treatment variables and instrumental variables, where $q_2 \geq q_1$. Let $T^{\boldsymbol{d}}$ denote the potential survival time given $\mathbf{D} = \boldsymbol{d}$, and $X$ represent the $p-$dimensional vector of covariates. If $q_2 > q_1$, we can linearly project $\mathbf{D}$ on $\mathbf{V}$ and $\mathbf{X}$ to construct a new instrument with $q_1$ dimension. Thus, we can assume $q_1 = q_2 = q$ without loss of generality. Define $Q_{\log T^{\boldsymbol{d}}}(\tau|\mathbf{X}) = \inf\{t : \Pr(T^{\boldsymbol{d}} \leq t|\mathbf{X}) \geq \tau\}$. Besides, $\perp\!\!\!\perp$ represents the statistical independence. We adopt the following standard IV assumptions.

(A1) Given $\mathbf{X} = \mathbf{x}$, for each $d$, $\log T^{\boldsymbol{d}} = q(\boldsymbol{d}, \mathbf{x}, U_{\boldsymbol{d}})$, where $q(\boldsymbol{d}, \mathbf{x}, \tau)$ is strictly increasing in $\tau$ and $U_{\boldsymbol{d}} \sim Uniform(0, 1)$

(A2) Independence of IV: $U_{\boldsymbol{d}} \perp\!\!\!\perp \mathbf{V}|\mathbf{X}$.

(A3) Selection : There exists an unknown function $g$ and random variable $\mathbf{R}$, such that $g(\mathbf{R}, \mathbf{X}, \mathbf{V}) = \mathbf{D}$

(A4) Rank Similarity: $U_{\boldsymbol{d}}|\mathbf{X}, \mathbf{V}, \mathbf{R}$ are identical distributions for varying $\boldsymbol{d}$.

The above potential outcome framework and assumptions have been commonly adopted in the IV literature, such as Chernozhukov and Hansen (2005, 2006, 2008). Assumption (A1) provides the Skorohod Representation of the potential outcome $\log T^{\boldsymbol{d}}$ (Chernozhukov and Hansen, 2006). Assumption (A2) implies that the potential outcomes are independent of $\mathbf{V}$ given $\mathbf{X}$. The assumption (A3) is a representation of the treatment selection mechanism, which is a weaker assumption compared to the monotoncity assumption in complier causal effect approach and avoid functional form assumptions between treatment variables and IVs. Assumption (A4) is the key assumption in identifying the population quantile causal effect, which restricts the

evolution of the distribution of the ranks across treatment states (Chernozhukov and Hansen, 2005).

## 4.2 The Proposed Model

### 4.2.1 Censored population quantile causal effect (CPQCE) model

In this work, we propose the following censored population quantile causal effect (CPQCE) model:

$$Q_{\log T^{\boldsymbol{d}}}(\tau|\mathbf{X}) = q(\boldsymbol{d}, \mathbf{X}, \tau),\ \tau \in (0, \tau_U], \tag{4.1}$$

where $q(\boldsymbol{d}, \mathbf{X}, \tau) = \boldsymbol{d}^\top \boldsymbol{\beta}_0(\tau) + \mathbf{X}^\top \boldsymbol{\gamma}_0(\tau)$, $Q_{\log T^{\boldsymbol{d}}}(\tau|\mathbf{X}) = \inf\{t : \Pr(T^{\boldsymbol{d}} \leq t|\mathbf{X}) \geq \tau\}$, and $\tau_U$ is a constant between $(0, 1)$. Besides, define $q(\mathbf{D}, \mathbf{X}, \tau) = \mathbf{D}^\top \boldsymbol{\beta}_0(\tau) + \mathbf{X}^\top \boldsymbol{\gamma}_0(\tau)$. After simple algebraic manipulation, we have $\boldsymbol{\beta}_0(\tau) = \partial Q_{\log T^{\boldsymbol{d}}}(\tau|\mathbf{X})/\partial \boldsymbol{d}$, which represents effects of the treatment covariates among $\tau-$th quantile of the potential survival time (in the logarithm scale) given the covariates in $\mathbf{X}$. Thus, $\boldsymbol{\beta}_0(\tau)$ will be referred as population $\tau-$th quantile causal effect.

To estimate $\boldsymbol{\beta}_0(\tau)$, model (4.1) can not be directly estimated because it includes potential survival outcomes, which are not all observable in practice. To address the estimation of $\boldsymbol{\beta}_0(\tau)$, we find that CPQCE model (4.1) can imply the following equation under assumption A1-A4,

$$\Pr(\log T < q(\mathbf{D}, \mathbf{X}, \tau)|\mathbf{X}, \mathbf{V}) = \Pr(\log T \leq q(\mathbf{D}, \mathbf{X}, \tau)|\mathbf{X}, \mathbf{V}) = \tau, \tau \in (0, \tau_U], \tag{4.2}$$

where $T = T^d$ when $D = d$. The justification of equation (4.2) has been shown in Proposition 4.1 in Appendix E. Unlike model (4.1), the equation (4.2) is linked with the observed survival time $T$, and thus can be used to construct an estimating

equation of $\boldsymbol{\beta}_0(\tau)$ without potential outcome $T^d$.

### 4.2.2 Estimation procedure with randomly censored Data

The right censoring data, where observed survival time $T$ is often right censored by $C$, is very common in biomedical studies. In this work, we consider the standard right censoring case that we only observe $W = \min\{T, C\}$ and $\delta = I(T \leq C)$. We adopt the following censoring assumption:

(A5) $U_{C|\mathbf{D},\mathbf{X},\mathbf{V}} \perp\!\!\!\perp (U_{\mathbf{D}}|\mathbf{D}, \mathbf{X}, \mathbf{V})|\mathbf{X}, \mathbf{V}$.

Here, $U_{C|\mathbf{D},\mathbf{X},\mathbf{V}} = \sup\{u : \log C \geq q(\mathbf{D}, \mathbf{X}, u)\}$ and $U_{\mathbf{D}}|\mathbf{D}, \mathbf{X}, \mathbf{V} = U_{\boldsymbol{d}}|\mathbf{D} = \boldsymbol{d}, \mathbf{X}, \mathbf{V}$ when $\mathbf{D} = \boldsymbol{d}$. The observed data consist of $n$ independent replicates of $\mathbf{O}$, denoted by $\mathbf{O}_i = \{W_i, \delta_i, \mathbf{D}_i, \mathbf{V}_i, \mathbf{X}_i\}_{i=1}^n$.

To simplify the notations, let $\mathbf{Z} = (\mathbf{D}^\top, \mathbf{X}^\top)^\top$ and $\mathbf{U} = (\mathbf{X}^\top, \mathbf{V}^\top)^\top$. Suppose $N(t) = I(W \leq t, \delta = 1)$ and $H(x) = -\log(1-x)$. From Proposition 4.2, we have the following estimating equation from equation (4.2):

$$n^{1/2}\boldsymbol{Q}_n(\boldsymbol{\beta}, \boldsymbol{\gamma}, \tau) = 0, \tag{4.3}$$

where $\boldsymbol{Q}_n(\boldsymbol{\beta}, \boldsymbol{\gamma}, \tau) = n^{-1}\sum_{i=1}^n \mathbf{U}_i\Big\{N_i(\exp\{\mathbf{D}_i^\top\boldsymbol{\beta}(\tau)+\mathbf{X}_i^\top\boldsymbol{\gamma}(\tau)\})-\int_0^\tau I(W_i \geq \exp\{\mathbf{D}_i^\top\boldsymbol{\beta}(u)+\mathbf{X}_i^\top\boldsymbol{\gamma}(u)\})dH(u)\Big\}$. Since equation (4.3) is not monotone, it may not be reliably solved by the standard numerical optimization algorithms. It is remarkable that the $\mathbf{X}-$part of $\boldsymbol{Q}_n(\boldsymbol{\beta}, \boldsymbol{\gamma}, \tau)$, $n^{-1}\sum_{i=1}^n \mathbf{X}_i\Big\{N_i(\exp\{\mathbf{D}_i^\top\boldsymbol{\beta}(\tau) + \mathbf{X}_i^\top\boldsymbol{\gamma}(\tau)\}) - \int_0^\tau I(W_i \geq \exp\{\mathbf{D}_i^\top\boldsymbol{\beta}(u) + \mathbf{X}_i^\top\boldsymbol{\gamma}(u)\})dH(u)\Big\}$ is a monotone function with $\boldsymbol{\beta}(\tau)$ fixed. Following the work of Chernozhukov and Hansen (2006, 2008), we propose a two-step estimator procedure to estimate $\boldsymbol{\beta}(\tau)$ and $\boldsymbol{\gamma}(\tau)$. Let the estimators of $\boldsymbol{\beta}(\tau)$ and $\boldsymbol{\gamma}(\tau)$, denoted by $\hat{\boldsymbol{\beta}}(\tau)$ and $\hat{\boldsymbol{\gamma}}(\tau)$, be piece-wise constant functions that only jump on a grid,

$S_{L(n)} = \{0 = \tau_0 < \tau_1 \cdots < \tau_{L(n)} = \tau_U < 1\}$. Let $N(\exp\{\mathbf{D}^\top \hat{\boldsymbol{\beta}}(0) + \mathbf{X}^\top \hat{\boldsymbol{\gamma}}(0)\}) \equiv 0$ and

$$\mathbf{S}_n(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \tau) = n^{-1} \sum_{i=1}^{n} \mathbf{X}_i \Big\{ N_i(\exp\{\mathbf{D}_i^\top \boldsymbol{\beta} + \mathbf{X}_i^\top \boldsymbol{\gamma}\}) - \int_0^\tau I(W_i \geq \exp\{\mathbf{D}_i^\top \hat{\boldsymbol{\beta}}(u) + \mathbf{X}_i^\top \hat{\boldsymbol{\gamma}}(u)\}) dH(u) \Big\}.$$

The two-step estimation procedure for $\hat{\boldsymbol{\beta}}(\tau)$ and $\hat{\boldsymbol{\gamma}}(\tau)$ can be implemented by applying grid-search and solving $L_1$ convex minimization problem iteratively from $\tau_0$ to $\tau_{L(n)}$. Specifically, suppose $\hat{\boldsymbol{\beta}}(\tau_m)$ and $\hat{\boldsymbol{\gamma}}(\tau_m)$ have been solved for $m < j$, for any $j = 1, \cdots, L(n)$. Let $\hat{\boldsymbol{\gamma}}(\boldsymbol{\beta}, \tau_j)$ denote the solution of $\mathbf{S}_n(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \tau_j) = 0$ given $\boldsymbol{\beta} \in [-M, M]^p$ and $(\hat{\boldsymbol{\beta}}(\tau_j)^\top, \hat{\boldsymbol{\gamma}}(\tau_j)^\top)^\top$. Then $\hat{\boldsymbol{\beta}}(\tau_j)$ is defined as the solution to

$$\mathbf{U}_n(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}}, \boldsymbol{\beta}, \tau_j) = \underline{0}$$

where $\mathbf{U}_n(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}}, \boldsymbol{\beta}, \tau_j) = n^{-1} \sum_{i=1}^n \mathbf{U}_i \Big\{ N_i(\exp\{\mathbf{D}_i^\top \boldsymbol{\beta}(\tau) + \mathbf{X}_i^\top \hat{\boldsymbol{\gamma}}(\boldsymbol{\beta}, \tau_j)\}) - \int_0^\tau I(W_i \geq \exp\{\mathbf{D}_i^\top \hat{\boldsymbol{\beta}}(u) + \mathbf{X}_i^\top \hat{\boldsymbol{\gamma}}(u)\}) dH(u) \Big\}$. Finally, define $\hat{\boldsymbol{\gamma}}(\tau_j) = \hat{\boldsymbol{\gamma}}(\hat{\boldsymbol{\beta}}(\tau_j), \tau_j)$. The detailed estimation procedure for $\hat{\boldsymbol{\beta}}(\tau_j)$ and $\hat{\boldsymbol{\gamma}}(\tau_j)$ is described as follows:

Step 1 Let $N(\exp\{\mathbf{D}^\top \hat{\boldsymbol{\beta}}(0) + \mathbf{X}^\top \hat{\boldsymbol{\gamma}}(0)\}) \equiv 0$

Step 2 : For $j = 1, \ldots, L(n)$

Step 2.1 Define a grid of values $\{\boldsymbol{\beta}_k, k = 1, \cdots K\}$ in $[-M, M]^p$ and $G_{\boldsymbol{\beta}} = \max_{s \in [-M,M]^p} \min_{k=1,\cdots K}$ $\|\boldsymbol{\beta}_k\|$

Step 2.2 For each $\boldsymbol{\beta}_k$, solve $\hat{\boldsymbol{\gamma}}(\boldsymbol{\beta}_k, \tau_j)$ by minimizing the following $L_1$-type convex

objective function:

$$
\begin{aligned}
l_j(\boldsymbol{\gamma}) = & \sum_{i=1}^{n} \left| \delta_i \log W_i - \delta_i (\mathbf{D}_i^\top \boldsymbol{\beta}_k + \mathbf{X}_i^\top \boldsymbol{\gamma}) \right| \\
& + \left| R^* - \boldsymbol{\gamma}^\top \Big\{ \sum_{l=1}^{n} (-\delta_l \mathbf{X}_l) \Big\} \right| \\
& + \left| R^* - \boldsymbol{\gamma}^\top \sum_{r=1}^{n} \left( 2\mathbf{X}_r \times \sum_{m=0}^{j-1} I\Big[ W_r \geq \exp\Big\{ \mathbf{D}_r^\top \hat{\boldsymbol{\beta}}(\tau_m) + \mathbf{X}_r^\top \hat{\boldsymbol{\gamma}}(\tau_m) \Big\} \Big] \right. \right. \\
& \left. \left. \times \{ H(\tau_{m+1}) - H(\tau_m) \} \right) \right|
\end{aligned}
\tag{4.4}
$$

Step 2.3 Let $k_0 = \text{argmin}_{k=1,\cdots,K} \|\mathbf{U}_n(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}}, \boldsymbol{\beta}_k, \tau_j)\|$. Denote $\hat{\boldsymbol{\beta}}(\tau_j) = \boldsymbol{\beta}_{k_0}$ and $\hat{\boldsymbol{\gamma}}(\tau_j) = \hat{\boldsymbol{\gamma}}(\boldsymbol{\beta}_{k_0}, \tau_j)$

In practice, the grid search used for obtaining $\hat{\boldsymbol{\beta}}(\tau)$ may lead to large variation of $\hat{\boldsymbol{\beta}}(\tau)$ at some neighborhood of quantiles. To address this undesirable numerical properties, we can apply smoothing techniques, such as moving average method, on $\hat{\boldsymbol{\beta}}(\tau)$ and $\hat{\boldsymbol{\gamma}}(\tau)$ to obtain smoothed estimator $\tilde{\boldsymbol{\beta}}(\tau)$ and $\tilde{\boldsymbol{\gamma}}(\tau)$, for any $v \in (0, \tau_U]$. The Lipschitz continuity of $\boldsymbol{\beta}_0(\tau)$ and $\boldsymbol{\gamma}_0(\tau)$ and the uniformly continuity and weak convergence of $\hat{\boldsymbol{\beta}}(\tau)$ and $\hat{\boldsymbol{\gamma}}(\tau)$ for $tau \in [v, \tau_U]$ ensure the uniformly continuity and weak convergence of $\tilde{\boldsymbol{\beta}}(\tau)$ and $\tilde{\boldsymbol{\gamma}}(\tau)$ for any $v \in (0, \tau_U]$.

### 4.2.3 Asymptotic properties

To establish the asymptotic properties of the proposed estimator $(\hat{\boldsymbol{\beta}}^\top(\tau), \hat{\boldsymbol{\gamma}}^\top(\tau))^\top$, we firstly introduce some new notations.

Let $\|S_{L_{(n)}}\| = \max\{\tau_j - \tau_{j-1} : j = 1, 2 \cdots L_{(n)}\}$. Recall $\mathbf{Z} = (\mathbf{D}^\top, \mathbf{X}^\top)^\top$ and $\mathbf{U} = (\mathbf{V}^\top, \mathbf{X}^\top)^\top$. From assumptions (A1) and (A5), we have $T \perp\!\!\!\perp \mathbf{V}|\mathbf{X}, \mathbf{D}$ and $W \perp\!\!\!\perp \mathbf{V}|\mathbf{X}, \mathbf{D}$. Thus, we denote $F(t|\mathbf{Z}) = F(W \leq t|\mathbf{Z}) = F(W \leq t|\mathbf{X}, \mathbf{D}, \mathbf{V})$, $\bar{F}(t|\mathbf{Z}) = 1 - F(t|\mathbf{Z})$, $\tilde{F}(t|\mathbf{Z}) = F(W \leq t, \delta = 1|\mathbf{Z}) = F(W \leq t, \delta = 1|\mathbf{X}, \mathbf{D}, \mathbf{V})$, $\bar{f}(x|\mathbf{Z}) = -f(x|\mathbf{Z}) = d\bar{F}(x|\mathbf{Z})/dx$, and $\tilde{f}(x|\mathbf{Z}) = d\tilde{F}(x|\mathbf{Z})/dx$. Define $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\gamma}^\top)^\top$,

$\boldsymbol{\theta}_0(\tau) = (\boldsymbol{\beta}_0^\top(\tau), \boldsymbol{\gamma}_0^\top(\tau))^\top$ and $\hat{\boldsymbol{\theta}}(\tau) = (\hat{\boldsymbol{\beta}}^\top(\tau), \hat{\boldsymbol{\gamma}}^\top(\tau))^\top$. Besides, define $\mathbf{B}(\boldsymbol{\theta}) = E\{\mathbf{U}\mathbf{Z}^\top \tilde{f}(\exp(\mathbf{Z}^\top\boldsymbol{\theta})|\mathbf{Z})\exp(\mathbf{Z}^\top\boldsymbol{\theta})\}$, and $\mathbf{J}(\boldsymbol{\theta}) = E\{\mathbf{U}\mathbf{Z}^\top \bar{f}(\exp(\mathbf{Z}^\top\boldsymbol{\theta})|\mathbf{Z})\exp(\mathbf{Z}^\top\boldsymbol{\theta})\}$. Let $\mathbf{B}_x(\mathbf{b}, \boldsymbol{c}) = E\{\mathbf{X}^{\otimes 2}\tilde{f}(\exp(\mathbf{D}^\top\mathbf{b} + \mathbf{X}^\top\boldsymbol{c})|\mathbf{Z})\exp(\mathbf{D}^\top\mathbf{b} + \mathbf{X}^\top\boldsymbol{c})\}$, where $\mathbf{u}^{\otimes 2}$ denotes $\mathbf{u}\mathbf{u}^\top$. Define $\mathbf{B}^S(\boldsymbol{\theta}) = \{\mathbf{B}(\boldsymbol{\theta}) + \mathbf{B}^\top(\boldsymbol{\theta})\}/2$, and $\mathcal{B}(d) = \{\boldsymbol{\theta} \in \mathbb{R}^{p+q} : \inf_{\tau\in(0,\tau_U]} \|\mathrm{E}[\mathbf{U}N(\exp\{\mathbf{Z}^\top\boldsymbol{\theta}\})] - \mathrm{E}[\mathbf{U}N(\exp\{\mathbf{Z}^\top\boldsymbol{\theta}_0(\tau)\})]\| \le d\}$ for $d > 0$. We adopt the following regularity conditions:

(C1) $(\mathbf{D}^\top, \mathbf{X}^\top, \mathbf{V}^\top)^\top$ belongs to a compact set.

(C2) (i)For $v \in (0, \tau_U]$, each component of $E\Big\{\mathbf{U}N(\exp\{\mathbf{Z}^\top\boldsymbol{\theta}_0(\tau)\})\Big\}$ is Lipschitz function of $\tau$; (ii)$\tilde{f}(t|\mathbf{Z})$ and $f(t|\mathbf{Z})$ are bounded above uniformly in $t$ and $\mathbf{Z}$.

(C3) For some $d_0 > 0$, (i) $\tilde{f}(\exp(\mathbf{Z}^\top\boldsymbol{\theta})|\mathbf{Z}) > 0$ for all $\boldsymbol{\theta} \in \mathcal{B}(d_0)$; (ii)$E(\mathbf{X}^{\otimes 2}) > 0$; (iii) each component of $E\{\mathbf{U}\mathbf{Z}^\top \bar{f}(\exp(\mathbf{Z}^\top\boldsymbol{\theta})|\mathbf{U})\exp(\mathbf{Z}^\top\boldsymbol{\theta})\} \times E\{\mathbf{U}\mathbf{Z}^\top \tilde{f}(\exp(\mathbf{Z}^\top\boldsymbol{\theta})|\mathbf{U})\exp(\mathbf{Z}^\top\boldsymbol{\theta})\}^{-1}$ is uniformly bounded in all $\boldsymbol{\theta} \in \mathcal{B}(d_0)$; (iv) each component of $\exp(\mathbf{D}^\top\boldsymbol{\beta}_0(\tau) + \mathbf{X}^\top\boldsymbol{\gamma})\mathbf{B}_x(\boldsymbol{\beta}_0(\tau), \boldsymbol{\gamma})^{-1}$ is uniformly bounded in $(\boldsymbol{\beta}_0(\tau), \boldsymbol{\gamma}) \in \mathcal{B}(d_0)$ for $\tau \in (0, \tau_U]$ and $\mathbf{Z}$.

(C4) For some $d_0 > 0$, (i) $\mathbf{B}^S(\boldsymbol{\theta})$ is positive-definite or negative-definite for all $\boldsymbol{\theta} \in \mathcal{B}(d_0)$; (ii) for any $v \in (0, \tau_U]$, the absolute value of each component of $E^{-1}\{\mathbf{U}\mathbf{Z}^\top \tilde{f}(\exp[\mathbf{Z}^\top\boldsymbol{\theta}_0(\tau)]|\mathbf{Z})\exp[\mathbf{Z}^\top\boldsymbol{\theta}_0(\tau)]\}$ is uniformly bounded for $\tau \in [v, \tau_U]$

Condition (C1) implies the boundedness of the covariates $\mathbf{D}$, $\mathbf{X}$ and $\mathbf{V}$. Conditions (C2) and (C3) are the similar conditions that adopted in Peng and Huang (2008), which assume realistic assumptions for on covariates, the underlying coefficient process, and the density functions. Condition (C4) ensures the identifiability of the $\{\boldsymbol{\theta}(\tau) : \tau \in (0, \tau_U]\}$ and its uniform consistency.

We establish the uniform consistency and weak convergence of the proposed estimator under regularity conditions C1-C4 described above. The main asymptotic results are stated in the Theorems 4.1 and 4.2.

**Theorem 4.1** (Uniform consistency)**.** *Under conditions C1-C4, if* $\lim_{n\to\infty}\|S_L\| = 0$*, then for any* $v \in (0, \tau_U]$*,*

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

**Theorem 4.2** (Weak convergence)**.** *Under conditions C1-C4, if* $\lim_{n\to\infty} n^{1/2}\|S_{L_{(n)}}\| = 0$*, then for any* $v \in (0, \tau_U]$*,* $n^{1/2}\{\hat{\boldsymbol{\theta}}(\tau) - \boldsymbol{\theta}_0(\tau)\}$ *converges weakly to a Gaussian Process for* $\tau \in [v, \tau_U]$*.*

The proofs of Theorems 4.1 and 4.2 follow the similar arguments in Peng and Huang (2008). One of the complication is to justify the existence of a one-to-one map between $\boldsymbol{\theta}$ and $E\{\mathbf{U}N\{\mathbf{Z}^\top\boldsymbol{\theta}\}\}$ in a neighborhood of $\{\boldsymbol{\theta}_0(\tau), \tau \in (0, \tau_U]\}$. The derivative of $E\{\mathbf{U}N\{\mathbf{Z}^\top\boldsymbol{\theta}\}\}$ is a function of $\mathbf{U}\mathbf{Z}^\top$, which may not be positive-definiteness in generally. To overcome this difficulty, we require the positive or negative definiteness of $\mathbf{B}^s(\boldsymbol{\theta})$ to ensure the existence of this one-to-one map. Another complication is to bound $\|\hat{\boldsymbol{\theta}}(\tau) - \boldsymbol{\theta}_0(\tau)\|$. The proposed estimator $\hat{\boldsymbol{\theta}}(\tau) = (\hat{\boldsymbol{\beta}}(\tau)^\top, \hat{\boldsymbol{\gamma}}(\tau)^\top)^\top$ is obtained by a two-step procedure, which estimates $\boldsymbol{\gamma}(\tau)$ with $\boldsymbol{\beta}(\tau)$ fixed firstly, and estimate $\boldsymbol{\beta}(\tau)$ based on the estimated $\boldsymbol{\gamma}(\tau)$. Thus, it is difficult to bound $\|\hat{\boldsymbol{\theta}}(\tau) - \boldsymbol{\theta}_0(\tau)\|$ by directly applying the procedure in Peng and Huang (2008). To solve this problem, we use an estimator with $\hat{\boldsymbol{\beta}}(\tau) = \boldsymbol{\beta}_0(\tau)$, $\tilde{\boldsymbol{\theta}}(\tau) = (\boldsymbol{\beta}_0(\tau)^\top, \hat{\boldsymbol{\gamma}}(\boldsymbol{\beta}_0(\tau), \tau)^\top)^\top$, as a intermediate step to bound $\|\hat{\boldsymbol{\theta}}(\tau) - \boldsymbol{\theta}_0(\tau)\|$ with an upper bound $\|\hat{\boldsymbol{\theta}}(\tau) - \tilde{\boldsymbol{\theta}}_0(\tau)\| + \|\tilde{\boldsymbol{\theta}}(\tau) - \boldsymbol{\theta}_0(\tau)\|$. In this upper bound, $\|\tilde{\boldsymbol{\theta}}(\tau) - \boldsymbol{\theta}_0(\tau)\|$ can be bounded by directly applying the similar procedure in Peng and Huang (2008) since $\tilde{\boldsymbol{\beta}}(\tau) = \boldsymbol{\beta}_0(\tau)$. Another part of this upper bound, $\|\hat{\boldsymbol{\theta}}(\tau) - \tilde{\boldsymbol{\theta}}_0(\tau)\|$ can be bounded by delicately using the result $\|\mathbf{U}_n(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{\beta}}(\tau), \tau)\| \le \|\mathbf{U}_n(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}}, \boldsymbol{\beta}_0(\tau), \tau)\|$, which is implied by the definition of $\hat{\boldsymbol{\theta}}(\tau)$. The detailed proofs are provided in Appendix F.

## 4.3 Inference

To make inference on $\boldsymbol{\theta}_0(\tau)$, the standard bootstrapping procedures can be used, such as the classical resampling with replacement or resampling with pertubed estimating equation proposed in Jin et al. (2001). However, these methods need long computation time for resampling and repeating grid-search in each resampled dataset.

In this work, we propose a sample-based inference approach for covariance estimation by following the lines of Sun et al. (2016). This sample-based approach inference does not involve resampling, and thus can save considerable computation time. The key idea of this procedure is to find consistent estimators for $\mathbf{B}(\boldsymbol{\theta}_0(\tau))$ and $\mathbf{J}(\boldsymbol{\theta}_0(\tau))$, and then plug them into the closed form derived for the asymptotic covariance matrix of $n^{1/2}\{\hat{\boldsymbol{\theta}}(\tau) - \boldsymbol{\theta}_0(\tau)\}$ to obtain an consistent estimator.

To describe this inference procedure, let $\mathbf{L}_n(\mathbf{b}, \boldsymbol{c}) = n^{-1/2}\sum_{i=1}^n (\mathbf{X}_i^\top, \mathbf{V}_i^\top)^\top N_i(\exp(\mathbf{D}^\top \mathbf{b} + \mathbf{X}_i^\top \boldsymbol{c}))$, $\tilde{\mathbf{L}}_n(\mathbf{b}, \boldsymbol{c}) = n^{-1/2}\sum_{i=1}^n (\mathbf{X}_i^\top, \mathbf{V}_i^\top)^\top Y_i(\exp(\mathbf{D}^\top \mathbf{b} + \mathbf{X}_i^\top \boldsymbol{c}))$, $\boldsymbol{\iota}_j(u) = (\mathbf{X}_j^\top, \mathbf{V}_j^\top)^\top N_i(\exp(\mathbf{D}^\top \hat{\boldsymbol{\beta}}(u) + \mathbf{X}_i^\top \hat{\boldsymbol{\gamma}}(u)))$ and $\boldsymbol{\Omega}(u) = n^{-1}\sum_{j=1}^n \{\boldsymbol{\iota}_j(u)\}^{\otimes 2}$. The procedure for estimating $\mathbf{B}(\boldsymbol{\theta}_0(u))$ and $\mathbf{J}_0(\boldsymbol{\theta}(u))$ can be described as below.

1. Find a symmetric and nonsingular $(p+q)\times(p+q)$ matrix $\mathbf{E}_n(u) = \{\mathbf{e}_{n,1}(u), \cdots, \mathbf{e}_{n,p+q}(u)\}$, such that $\boldsymbol{\Omega}(u) = \{\mathbf{E}_n(u)\}^2$

2. Find the solution of

$$\mathbf{L}_n(\mathbf{b}, \boldsymbol{c}) = \mathbf{L}_n(\hat{\boldsymbol{\beta}}(u), \hat{\boldsymbol{\gamma}}(u)) + \mathbf{e}_{n,j}(u),$$

   and the solution is denoted by $(\hat{\mathbf{b}}_{n,j}(u), \hat{\boldsymbol{c}}_{n,j}(u))$. The following steps are used to solve this equation.

   2.1. Define a grid of values $\{\boldsymbol{\beta}_k, k = 1, \cdots, K\}$ around $\hat{\boldsymbol{\beta}}(u)$.

2.2. For each $\boldsymbol{\beta}_k$, obtain $\hat{\boldsymbol{c}}(\boldsymbol{\beta}_k)$ by solving

$$n^{-1/2}\sum_{i=1}^{n}\mathbf{X}_i N_i(\exp(\mathbf{D}^\top\boldsymbol{\beta}_k+\mathbf{X}_i^\top\boldsymbol{c}))=n^{-1/2}\sum_{i=1}^{n}\mathbf{X}_i N_i(\exp(\mathbf{D}^\top\hat{\boldsymbol{\beta}}(u)+\mathbf{X}_i^\top\hat{\boldsymbol{\gamma}}(u)))+\tilde{\mathbf{e}}_{n,j}(u),$$

where $\tilde{\mathbf{e}}_{n,j}(u)$ is the first p elements of $\mathbf{e}_{n,j}(u)$

2.3. Let $k_0=\text{argmin}_{k=1,\cdots,K}\|W_n(\hat{\boldsymbol{\beta}}(u),\hat{\boldsymbol{\gamma}}(u),\boldsymbol{\beta}_k,\hat{\boldsymbol{c}}(\boldsymbol{\beta}_k))\|$, where

$$\begin{aligned}&W_n(\hat{\boldsymbol{\beta}}(u),\hat{\boldsymbol{\gamma}}(u),\boldsymbol{\beta}_k,\hat{\boldsymbol{c}}(\boldsymbol{\beta}_k))\\ =&n^{-1/2}\sum_{i=1}^{n}(\mathbf{X}_i^\top,\mathbf{V}_i^\top)^\top N_i(\exp(\mathbf{D}^\top\boldsymbol{\beta}_k+\mathbf{X}_i^\top\hat{\boldsymbol{c}}(\boldsymbol{\beta}_k)))\\ &-n^{-1/2}\sum_{i=1}^{n}(\mathbf{X}_i^\top,\mathbf{V}_i^\top)^\top N_i(\exp(\mathbf{D}^\top\hat{\boldsymbol{\beta}}(u)+\mathbf{X}_i^\top\hat{\boldsymbol{\gamma}}(u)))-\mathbf{e}_{n,j}(u),\end{aligned}$$

Denote $(\boldsymbol{\beta}_{k_0},\hat{\boldsymbol{c}}_{\boldsymbol{\beta}_{k_0}})$ as $(\hat{\mathbf{b}}_{n,j}(u),\hat{\boldsymbol{c}}_{n,j}(u))$.

3. Calculate the $\mathbf{D}_n(u)\equiv\{(\hat{\mathbf{b}}_{n,1}(u)-\hat{\boldsymbol{\beta}}(u),\hat{\boldsymbol{c}}_{n,1}(u)-\hat{\boldsymbol{\gamma}}(u)),\cdots,(\hat{\mathbf{b}}_{n,p+q}(u)-\hat{\boldsymbol{\beta}}(u),\hat{\boldsymbol{c}}_{n,p+q}(u)-\hat{\boldsymbol{\gamma}}(u))\}$ and $\tilde{\mathbf{E}}_n(u)\equiv\{\tilde{L}_n(\hat{\mathbf{b}}_{n,1}(u),\hat{\boldsymbol{c}}_{n,1}(u))-\tilde{L}_n(\hat{\boldsymbol{\beta}}(u),\hat{\boldsymbol{\gamma}}(u)),\cdots,\tilde{L}_n(\hat{\mathbf{b}}_{n,p+q}(u),\hat{\boldsymbol{c}}_{n,p+q}(u))-\tilde{L}_n(\hat{\boldsymbol{\beta}}(u),\hat{\boldsymbol{\gamma}}(u))\}$

4. Compute $n^{-1/2}\mathbf{E}_n(u)\mathbf{D}_n(u)^{-1}$ and Compute $n^{-1/2}\tilde{\mathbf{E}}_n(u)D_n(u)^{-1}$, which provide consistent estimator for $\mathbf{B}_0(\boldsymbol{\theta}_0(u))$ and $\mathbf{J}_0(\boldsymbol{\theta}_0(u))$

Steps 2.1 to 2.3 are used to solve the non-monotone equation, $\mathbf{L}_n(\mathbf{b},\boldsymbol{c})=\mathbf{L}_n(\hat{\boldsymbol{\beta}}(u),\hat{\boldsymbol{\gamma}}(u))+\mathbf{e}_{n,j}(u)$. The motivation in these step is to utilize the monotoncity of the $X$-part of $\mathbf{L}_n(\mathbf{b},\boldsymbol{c})$ given fixed $\mathbf{b}$, $n^{-1/2}\sum_{i=1}^{n}\mathbf{X}_i N_i(\exp(\mathbf{D}^\top\boldsymbol{\beta}_k+\mathbf{X}_i^\top\boldsymbol{c}))$, to overcome numerical optimization issues from non-monotoncity of $L_n(\mathbf{b},\boldsymbol{c})$, which is similar to estimating procedure described in Section 4.2.2. In this inference procedure, we only need to do grid search $p+q$ times. In practice, the $p+q$ is generally much smaller than the numbers of the resampled datasets in resampling approach, and thus can save considerable times compared to bootstrapping procedure.

Let $\hat{\mathbf{B}}(u)$ and $\hat{\mathbf{J}}(u)$ denote the estimators of $\mathbf{B}(\boldsymbol{\theta}_0(u))$ and $\mathbf{J}(\boldsymbol{\theta}_0(u))$. Like Sun et al. (2016), the sample-based estimator for the covariance matrix of $n^{1/2}\{\hat{\boldsymbol{\theta}}(\tau) - \boldsymbol{\theta}_0(\tau)\}$ can be expressed by

$$n^{-1/2}\sum_{i=1}^{n}\hat{\boldsymbol{\xi}}_i(u)\hat{\boldsymbol{\xi}}_i^{\top}(v)$$

where $\hat{\boldsymbol{\xi}}_i^{\top}(u) = \hat{\mathbf{B}}^{-1}(u)\hat{\boldsymbol{\Phi}}(\hat{\boldsymbol{\Psi}}_i(u))$. Here

$$\boldsymbol{\Psi}_i(u) = \mathbf{U}_i\Big\{N_i(\exp\{\mathbf{Z}_i^{\top}\boldsymbol{\theta}_0(u)\}) - \int_0^u I(W_i \geq \exp\{\mathbf{Z}_i^{\top}\boldsymbol{\theta}_0(s)\})dH(s)\Big\}, i = 1, 2, \cdots, n,$$

and $\hat{\boldsymbol{\Phi}}(\cdot)$ is plug-in estimate for the operator $\boldsymbol{\Phi}(\cdot)$ defined in the proof of Theorem 4.2 in Appendix F. The justification of the proposed covariance estimator is in Appendix G.

## 4.4 Numerical Studies

### 4.4.1 Monte-Carlo simulations

We conduct extensive Monte-Carlo simulations to evaluate the proposed estimator for CPQCE model in the finite samples. To satisfy assumptions A1-A4, the simulated data $\{W, \delta, D, X, V\}$ is generated as follows.

1. Generate $(\epsilon, v_1)$ from the multivariate normal distribution $N_2(\mathbf{0}, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma} = \begin{pmatrix} 0.25 & 0.1 \\ 0.1 & 0.25 \end{pmatrix}$.
2. Generate $V$ from $N(0, 1)$, and $X$ from $N(0, 0.5^2)$ independently.
3. Determine $D$ by $D = \Phi(V + v_1 + 2X)$, where $\Phi$ is the cumulative distribution function of the standard normal distribution.
3. Generate event time $T$ for two scenarios (A) and (B).

(A) Generate $T = \exp(1 + D + 0.5X + (0.5 + D)\epsilon)$.

(B) Generate $T = \exp(1 + D + 0.5X + 0.5\epsilon)$.

4. Generating $U_C$ from uniform distribution to yield around 30% censoring in each scenario. Here censoring time $C$ is determined by $C = 1 + 0.25q(min(U_C, 1)) + \{1+0.5q(min(U_C, 1))\}D+0.5X$ in scenario (A) and $C = 1+0.25q(min(U_C, 1))+ 0.25q(min(U_C, 1))D + 0.5X$ in scenario (B).

5. Determine $W$ and $\delta$ by $W = \min(T, C)$, and $\delta = 1(T \leq C)$.

In this simulation, we consider two scenarios, (A) and (B), for generating survival time $T$. We can show that in scenario (A), $Q_{\log T^d}(\tau|X) = 1 + 0.25q(\tau) + \{1 + 0.5q(\tau)\}d + 0.5X$, and in scenario (B), $Q_{\log T^d}(\tau|X) = 1 + 0.25q(\tau) + d + 0.5X$,where $q(\tau)$ denotes the $\tau$-th quantile of $Normal(0, 1)$. The population quantile causal effect in scenario (A) is quantile-varying but constant in scenario (B).

In our simulations, two sample sizes $n = 1000$ and $n = 2000$ are considered for each scenario. In each setting, we generate 1000 datasets. In the proposed method, the coefficient of $D$ is grid searched in $[0, 3]$ with step size 0.01 in both settings. Except for the proposed method, we consider the naive as-treated method which directly applies Peng and Huang (2008) to the whole dataset without considering the unmeasured confounders. Figure 4.1 and Figure 4.2 present the results from both methods under scenario (A) and (B) with $n = 1000$. The results include the average coefficient estimates and the coverage probabilities of 95% confidence intervals from both methods, and the empirical variances and the average sample-based variance estimators in the proposed method for $\tau \in [0.1, 0.8]$.

From Figure 4.1, the average coefficient estimators for all covariates, and the coverage probabilities of 95% confidence intervals from the proposed method are fairly close to the true value or nominal level. Besides, the difference between average sample-based variance estimators and the empirical variances in the proposed model

are quite small. All these results suggest the satisfied performance of the proposed method for the scenario of quantile-varying PQCE. In contrast, the average coefficient estimates and their coverage probabilities of 95% confidence intervals in the as-treated method are far away from the true value or the nominal level. It indicates the as-treated method would lead to a biased estimator. Similar to the results in Figure 4.1, Figure 4.2 suggests the propose method works well for the scenario of quantile-constant PQCE. The as-treated method would still lead to a biased estimator even with quantile-constant PQCE.

Figure 4.3 and Figure 4.4 show the simulation results from the proposed method and the as-treated method for scenario (A) and (B) with $n = 2000$. The observations in these two figures are similar to the Figure 4.1 and Figure 4.2. The biases and differences between average of estimated sample-based variance and the empirical variance in the proposed method decrease as the sample size increases from 1000 to 2000.

### 4.4.2 Application to bone marrow transplant Dataset

Diffuse large B-cell lymphoma (DLBCL) is a common type of fast-growing non-Hodgkin lymphoma. In this work, we consider a dataset from the Center for International Blood and Marrow Transplant Research (CIBMTR), which includes 986 DLBCL patients between 18 to 76 years old with autologous hematopoietic stem cell transplantation (TX) between 1996 and 2003. There are 38 patients with missing data. After excluding these 38 patient, 164 patients are in the pre-transplant rituximab treatment group, while others are in the control group. The primary objective in this study is to evaluate the effect of the pre-transplant rituximab treatment on the progression free survival (PFS), which is defined as time to the composite endpoint of progressive lymphoma post-transplant, lymphoma recurrence, and death from any cause, in DLBCL patients. 340 (35.9%) subjects are censored in this dataset. The dei-

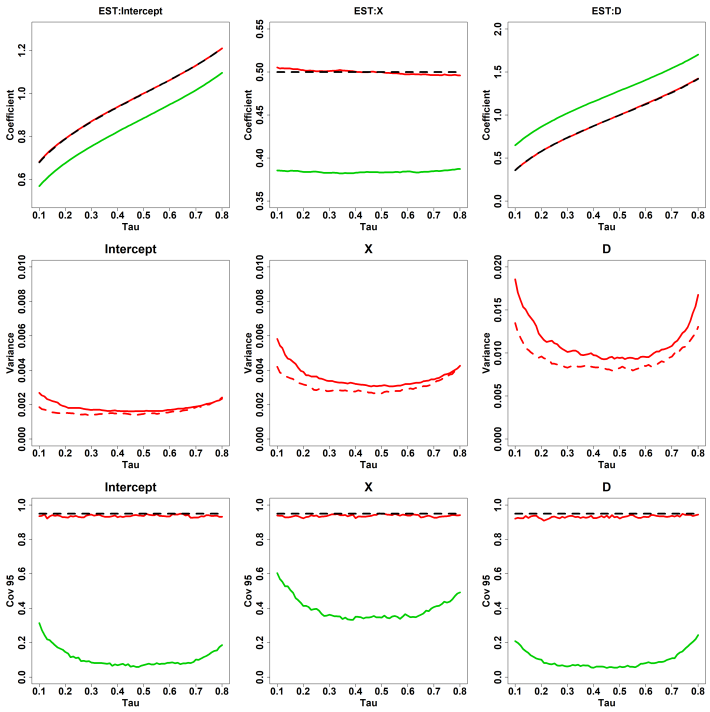


Figure 4.1: The simulation results in scenario (A) with sample size n=1000. In the first and the third rows, red lines correspond to the proposed method; green lines correspond to the naive as-treated method; dashed black lines correspond to true or nominal values. In the second row, red solid lines represent the average variance estimates from the sample-based inference procedure in the proposed method, and red dashed lines represent the empirical variances of the proposed estimator.

dentified SAS dataset can be found at the CIBMTR webisite: `https://www.cibmtr.org/ReferenceCenter/PubList/PubDsDownload/Pages/default.aspx`.

In this retrospective study, the observed variables, except the treatment choice, include age, number of chemotherapy regimens, disease status, and Karnofsky performance score. Their summary statistics are shown in Table 4.1. From Table 4.1, the patients in the rituximab group tended to be older and had more chemotherapy regimens than patients in the control group. It suggests the existence of some confounders that affect both of the pre-transplant treatment choices and the PFS in DLBCL patients. Some of these factors, such as the molecular subtype of lymphoma, are not measured in this dataset (Zheng et al., 2017). The standard method without considering unobserved confounders may lead to a biased estimator for the efficacy

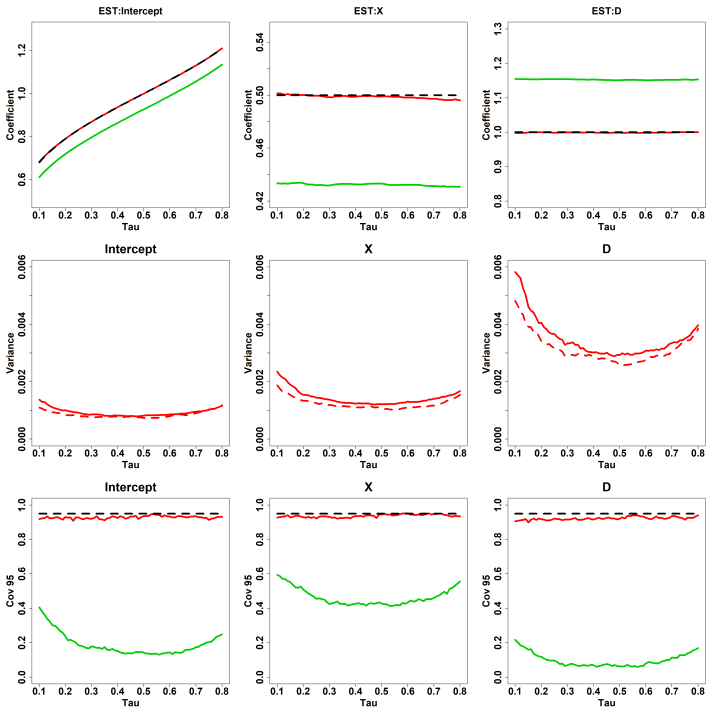


Figure 4.2: The simulation results in scenario (B) with sample size n=1000. In the first and the third rows, red lines correspond to the proposed method; green lines correspond to the naive as-treated method; dashed black lines correspond to true or nominal values. In the second row, red solid lines represent the average variance estimates from the sample-based inference procedure in the proposed method, and red dashed lines represent the empirical variances of the proposed estimator.

of rituximab in DLBCL patients.

We apply the proposed CPQCE model to account for the impact of unmeasured confounders in the estimation of causal treatment effect of the pre-transplant rituximab in DLBCL patients. The IV adopted in this analysis is whether the start date of the treatment is after the FDA approval date of rituximab. The use of this IV is motivated by the fact that no major changes in clinical practice or technical improvements for the autologous hematopoietic stem cell transplantation, except for the FDA approval of rituximab, happened during this study period. It is reasonable to assume this IV can only affect the PFS through the treatment choice, and is independent to unmeasured confounders. The association between IV and the treatment choice can be validated by the fact that only patients with start year after FDA ap-

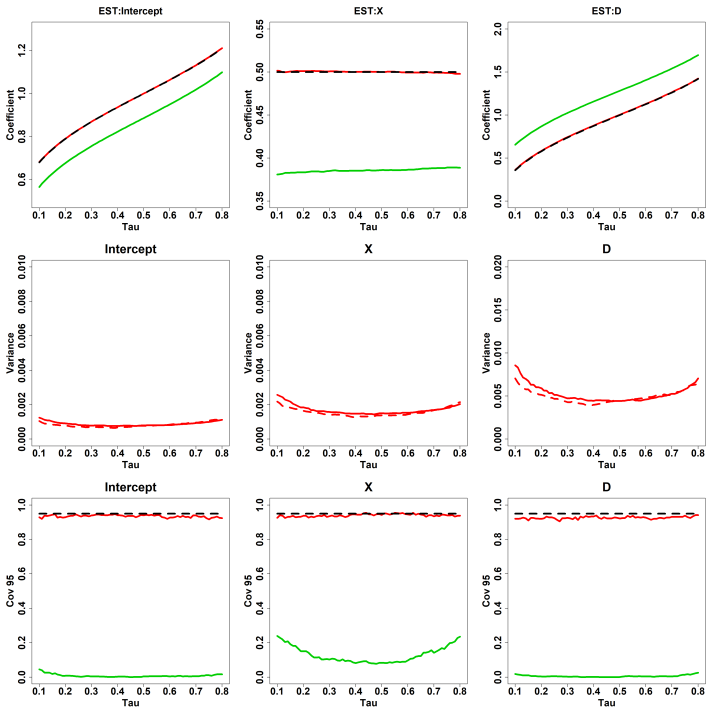


Figure 4.3: The simulation results in scenario (A) with sample size n=2000. In the first and the third rows, red lines correspond to the proposed method; green lines correspond to the naive as-treated method; dashed black lines correspond to true or nominal values. In the second row, red solid lines represent the average variance estimates from the sample-based inference procedure in the proposed method, and red dashed lines represent the empirical variances of the proposed estimator.

proval date can choose pre-treatment rituximab as their treatment choice. Because the treatment choice may also affected by unobserved factors, such the molecular subtype of lymphoma, the selection assumption is reasonable for the treatment choice. To best of knowledge, there is no formal test of rank similarity assumption for censored data. To check the rank similarity assumption, we follow the arguments in Frandsen and Lefgren (2018) and Dong and Shen (2018) to test $H_0 : \boldsymbol{\omega} = \mathbf{0}$ under $\hat{F}_{D_i}(T_i) = \alpha + D_i\beta + \mathbf{X}_i^\top\boldsymbol{\gamma} + D_i\mathbf{X}_i^\top\boldsymbol{\omega} + \epsilon$ for the uncensored subjects in CIBMTR dataset. Here $D_i$, $\mathbf{X}_i$ and $\hat{F}_{D_i}(\cdot)$ denote the treatment choice and observed covariates, and the estimated cumulative distribution function of the potential outcome $T^{D_i}$ conditional on $\mathbf{X}_i$ from the proposed CPQCE model. The p-value of this test is 0.369, which suggests the rank similarity assumption may be a reasonable assumption for

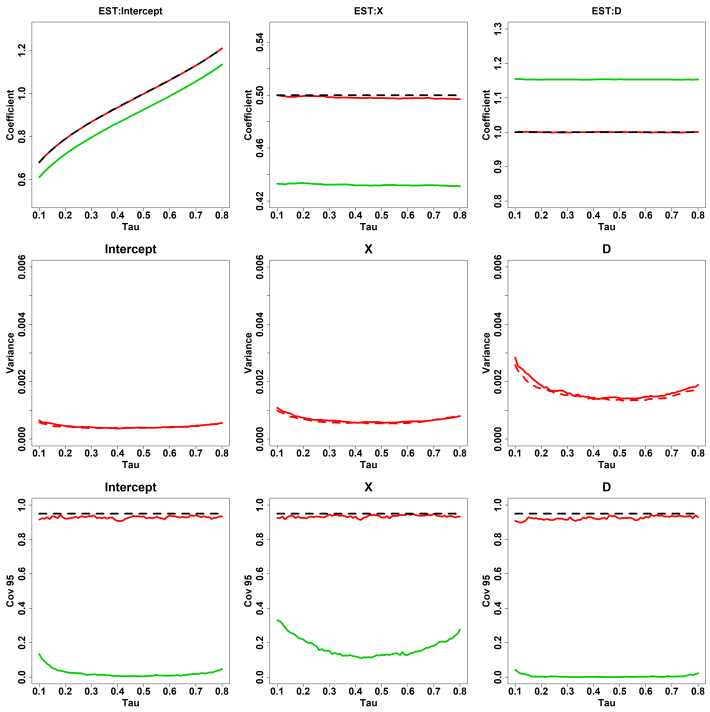


Figure 4.4: The simulation results in scenario (B) with sample size n=2000. In the first and the third rows, red lines correspond to the proposed method; green lines correspond to the naive as-treated method; dashed black lines correspond to true or nominal values. In the second row, red solid lines represent the average variance estimates from the sample-based inference procedure in the proposed method, and red dashed lines represent the empirical variances of the proposed estimator.

this CIBMTR dataset.

We also apply the proposed CPQCE model, the censored complier quantile casual effect (CCQCE) model with the same binary IV, and the as-treated analysis based on the method of Peng and Huang (2008) by adjusting for all the observed covariates in this CIBMTR dataset. In the proposed CPQCE model, we do grid search for $\beta(\tau)$, the coefficient of the treatment choice, between $[-1, 3]$ with step size 0.005. The local polynomial regression with degree 2 and $\alpha = 0.1$ has been applied in here to smooth the estimator of $\beta(\tau)$. The results on the treatment coefficients from all methods are presented in Figure 4.5. The results of other covariates are shown in Figure 4.6. From Figure 4.5, the estimated $PQCE(0.5)$, which equals to 2.47, indicates a significant treatment effect on median PFS survival time. This estimator indicates for

the patients with covariates set at their reference levels (i.e. younger than 55 years, having 0 or 1 chemotherapy regimens, with CR1 disease status and karnofsky score $< 90$), receiving rituximab may prolong their median PFS 11.83 (exp(2.47)) times compared to receiving standard intervention.

Figure 4.5 shows the estimated benefit of rituximab on prolonging PFS in CPQCE model is strongest among all methods. The $PQCE(\tau)$ of rituximab from CPQCE model is significantly above 0 across all $\tau$'s between 0.1 and 0.6, and signficantly above 2 across all $\tau$'s between 0.25 to 0.4. The $CQCE(\tau)$ of rituximab from CCQCE model is significantly above 0 across all $\tau$'s between 0.1 and 0.6, but significantly below 2 for $\tau \in [0.2, 0.4]$. The as-treated method provides the weakest benefit of rituximab on PFS, which only signficantly prolongs PFS until $\tau > 0.3$. The finding that the as-treated method underestimates the benefit of that rituximab on PFS is consistent with the conclusion in Zheng et al. (2017). Another interesting finding is that the benefit of rituximab in populations is stronger than the benefit in treated population. These findings can be explained by the difference characteristics of patients between treated group and whole population. Table 4.1 shows the patients with high-risk, such as older patients with more chemo regimens, are more likely in the treated group. Since the high-risk patients generally imply shorter PFS, the as-treated method can underestimate the benefit of rituximab, particularly on low-quantiles of PFS, without considering the selection bias induced by unmeasured confounders. The differences in characteristics between treated group and whole population may also lead to difference of effect of rituximab on PFS between these groups. From Table 4.1, the treated population has large percentage of high risk patients than the whole population. Since the high-risk patients always imply shorter PFS, the difference between treated population and whole population is large at low quantiles of PFS, and becomes small at high quantiles of PFS. Besides, $CQCE(\tau)$ denotes the quantile causal effect in the treated population because this study is one-compliance.

Therefore, the difference between $PQCE(\tau)$ and $CQCE(\tau)$ is large at low-quantiles, specially for $\tau \in [0.2, 0.4]$, and becomes small at high quantiles, such as $\tau > 0.5$. Our finding about rituximab's effect is consistent with that from Zheng et al. (2017), that the standard analysis without address the endogenous treatment selection would underestimate the effect of the rituximab.

Table 4.1: Summary statistics of patients in the CIBMTR dataset, overall and stratified by the treatment group

| Variable | | Entire Sample (N=948) | Ritux Group (N=164) | Control Group (N=784) | p-value |
|---|---|---|---|---|---|
| Age> 55 | No | 543 (57.3%) | 74 (45.1%) | 469 (59.8%) | $< 0.001$[1] |
| | Yes | 405 (42.7%) | 90 (54.9%) | 315 (40.2%) | |
| Number of chemo regimen | 1 | 128 (13.5%) | 12 (7.3%) | 116 (14.8%) | $< 0.001$[1] |
| | 2 | 413 (43.6%) | 58 (35.4%) | 355 (45.3%) | |
| | 3 | 290 (30.6%) | 61 (37.2%) | 229 (29.2%) | |
| | 4 | 96 (10.1%) | 31 (18.9%) | 65 (8.3%) | |
| | 5 | 21 (2.2%) | 2 (1.2%) | 19 (2.4%) | |
| Status | PIF sensitive | 172 (18.1%) | 31 (18.9%) | 141 (18%) | 0.006[1] |
| | PIF resistant | 51 (5.4%) | 10 (6.1%) | 41 (5.2%) | |
| | CR1 | 158 (16.7%) | 36 (22%) | 122 (15.6%) | |
| | REL sensitive | 291 (30.7%) | 43 (26.2%) | 248 (31.6%) | |
| | REL resistant | 65 (6.9%) | 12 (7.3%) | 53 (6.8%) | |
| | CR2+ | 154 (16.2%) | 32 (19.5%) | 122 (15.6%) | |
| | Other | 57 (6%) | 0 (0%) | 57 (7.3%) | |
| Karnofsky score | $< 90\%$ | 354 (37.3%) | 63 (38.4%) | 291 (37.1%) | 0.823[1] |
| | 90-100% | 594 (62.7%) | 101 (61.6%) | 493 (62.9%) | |
| Event | Progression or Death | 608 (64.1%) | 82 (50%) | 526 (67.1%) | 0.006[2] |
| | Censoring | 340 (35.9%) | 82 (50%) | 258 (32.9%) | |

[1] P-value is calculated from Chi-square test.
[2] P-value is calculated from Log-rank test which compares the PFS between two groups.

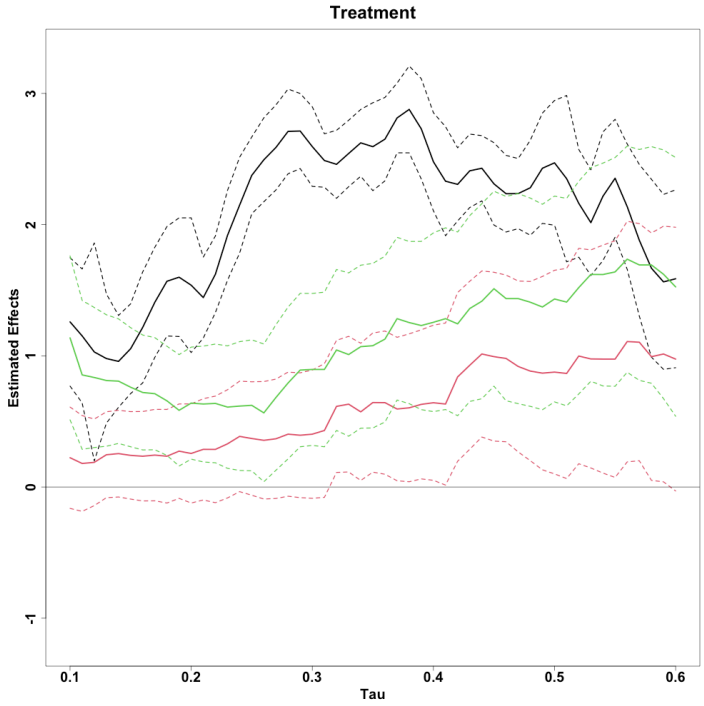


Figure 4.5: The results on treatment coefficients from analyzing the bone marrow transplant dataset based on the proposed CPQCE method, the as-treated censored quantile regression analysis, CCQCE method. Black solid and dashed lines respectively represent the coefficient estimates and 95% confidence intervals based on the proposed CPQCE method. Red solid and dashed lines respectively represent the coefficient estimates and 95% confidence intervals based on the as-treated censored quantile regression analysis. Green solid and dashed lines respectively represent the coefficient estimates and 95% confidence intervals based on the CCQCE method

## 4.5 Remarks

In this work, we propose a censored population quantile causal effect model to quantify the population quantile causal effect (PQCE) for standard right censoring data. PQCE is the quantile counterpart of population average causal effect (PACE), which has been commonly studies among IV literatures. Like the standard quantile treatment effect, PQCE requires weaker conditions for identifiability than PACE under censoring data (Klein and Moeschberger, 2006; Peng and Fine, 2007), and has advantages in capturing heterogeneous causal effects among potential outcome distribution. We propose a simple and justified two-stage estimation procedure to solve the nu-

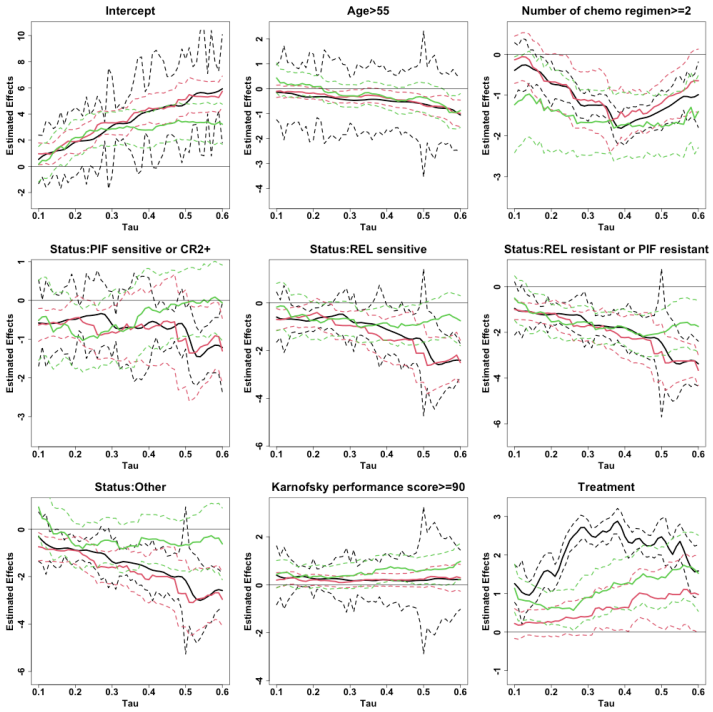


Figure 4.6: The results on all coefficients from analyzing the bone marrow transplant dataset based on the proposed CPQCE method, the as-treated censored quantile regression analysis, CCQCE method. Black solid and dashed lines respectively represent the coefficient estimates and 95% confidence intervals based on the proposed CPQCE method. Red solid and dashed lines respectively represent the coefficient estimates and 95% confidence intervals based on the as-treated censored quantile regression analysis. Green solid and dashed lines respectively represent the coefficient estimates and 95% confidence intervals based on the CCQCE method

merical issues from the non-monotonicity of the estimating equation. Furthermore, a sample-based covariance estimator has been proposed to reduce the burden of computation in inference procedure. The proposed two-stage estimation procedure can easily be implemented by existing software for quantile regression with grid search.

The main assumption in this work is the rank similarity assumption, which restricts the evolution of the distribution of the ranks across treatment states to identify the PQCE (Chernozhukov and Hansen, 2005). The rank similarity assumption can be a more plausible assumption if a rich set of covariates exists(Chernozhukov and Hansen, 2006). Several approaches (Frandsen and Lefgren, 2018; Dong and Shen, 2018) have been proposed to test this assumption. These approaches work only for

uncensored data, and can not be directly applied to censored data. Based on our knowledge, there is no existing literature about testing this assumption for censored data. It may be desirable to develop methods to test rank similarity for censored data in future studies.

The proposed two-step estimation procedure solves the non-monotonous the estimating equation by minimizing $L_1$ convex function after fixing $\boldsymbol{\beta}(\tau)$ on a fine grid. It can be solved fast when $\boldsymbol{\beta}(\tau)$ is a scalar. The computation burden may grow fast when the dimension of $\boldsymbol{\beta}(\tau)$ increases. When the dimension $\boldsymbol{\beta}(\tau)$ is high, some other numerical optimization approaches may be considered, such as Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm or a mixed integer linear programming (Pouliot, 2019), to reduce the computation burden.

## 4.6 Appendix

### 4.6.1 Appendix E: Propositions 4.1 and 4.2 and their proofs

#### 4.6.1.1 Proposition 4.1 and its proof

**Proposition 4.1.** *Under assumptions A1-A4,* $P(\log T < q(\mathbf{D}, \mathbf{X}, \tau)|\mathbf{X}, \mathbf{V}) = P(\log T \leq q(\mathbf{D}, \mathbf{X}, \tau)|\mathbf{X}, \mathbf{V}) = \tau$ *for* $\tau \in (0, 1)$ *a.s.*

*Proof.* For presentation simplicity, we will omit $\mathbf{X}$ in the following proof. From Assumption A3, we have $g(\mathbf{R}, \mathbf{V}) = \mathbf{D}$. Thus we can obtain

$$\begin{aligned} P(\log T \leq q(\mathbf{D}, \tau)|\mathbf{R}, \mathbf{V}) =& P(\log T^{\mathbf{D}} \leq q(\mathbf{D}, \tau)|\mathbf{R}, \mathbf{V}) \\ =& P(q(\mathbf{D}, U_{\mathbf{D}}) \leq q(\mathbf{D}, \tau)|\mathbf{R}, \mathbf{V}) \\ =& P(U_{\mathbf{D}} \leq \tau|\mathbf{R}, \mathbf{V}). \end{aligned} \tag{4.5}$$

The last equation holds because $q(\boldsymbol{d}, \mathbf{x}, \tau)$ is strictly increasing in $\tau$ from assumption A1.

From Assumption A4, there exists a specified $\boldsymbol{d}_0 \in \mathbf{D}$, such that $P(U_{\mathbf{D}} \leq \tau|\mathbf{R}, \mathbf{V}) = P(U_{\boldsymbol{d}_0} \leq \tau|\mathbf{R}, \mathbf{V})$. By (4.5), we have

$$\begin{aligned} P(\log T \leq q(\mathbf{D}, \tau)|\mathbf{V}) &= \int P(\log T \leq q(\mathbf{D}, \tau)|\mathbf{R}, \mathbf{V}) dP(\mathbf{R}|\mathbf{V}) \\ &= \int P(U_{\boldsymbol{d}_0} \leq \tau|\mathbf{R}, \mathbf{V}) dP(\mathbf{R}|\mathbf{V}) \\ &= P(U_{\boldsymbol{d}_0} \leq \tau|\mathbf{V}) \\ &= P(U_{\boldsymbol{d}_0} \leq \tau) \\ &= \tau. \end{aligned}$$

The last second equation holds because of $U_{\boldsymbol{d}} \perp\!\!\!\perp \mathbf{V}|\mathbf{X}$ in assumption A2.

Similarly, we can show $P(\log T < q(\mathbf{D}, \tau)|\mathbf{V}) = \tau$. □

#### 4.6.1.2 Proposition 4.2 and its proof

**Proposition 4.2.** *Under assumptions A1-A5, model* (4.1) *and equation* (4.2) *has the following estimating equation*

$$n^{1/2}\boldsymbol{Q}_n(\boldsymbol{\beta}, \boldsymbol{\gamma}, \tau) = 0, \tag{4.6}$$

*where* $\boldsymbol{Q}_n(\boldsymbol{\beta}, \boldsymbol{\gamma}, \tau) = n^{-1}\sum_{i=1}^n \mathbf{U}_i\Big\{N_i(\exp\{\mathbf{D}_i^\top\boldsymbol{\beta}(\tau)+\mathbf{X}_i^\top\boldsymbol{\gamma}(\tau)\})-\int_0^\tau I(W_i \geq \exp\{\mathbf{D}_i^\top\boldsymbol{\beta}(u)+\mathbf{X}_i^\top\boldsymbol{\gamma}(u)\})dH(u)\Big\}$.

*Proof.* Based on equation (4.2), we have the following equation after simple algebra manipulation:

$$\begin{aligned} &\int_0^\tau \Pr\{\log C \geq q(\mathbf{D}, \mathbf{X}, u)|\mathbf{U}\} d(\Pr\{\log T \leq q(\mathbf{D}, \mathbf{X}, u)|\mathbf{U}\}) \\ =&\int_0^\tau \Pr\{\log C \geq q(\mathbf{D}, \mathbf{X}, u)|\mathbf{U}\}\Pr\{\log T \geq q(\mathbf{D}, \mathbf{X}, u)|\mathbf{U}\}\frac{du}{1-u}. \end{aligned} \tag{4.7}$$

For the LHS of equation (4.7), we have

$$\begin{aligned}
&\int_0^\tau \Pr\{\log C \geq q(\mathbf{D},\mathbf{X},u)|\mathbf{U}\} d(\Pr\{\log T \leq q(\mathbf{D},\mathbf{X},u)|\mathbf{U}\}) \\
=&\int_0^\tau \mathrm{E}_{D|\mathbf{U}}\{\Pr(U_{C|\mathbf{U},\mathbf{D}} \geq u)\}\mathrm{E}_{D|\mathbf{U}}[d\{Pr(U_{\mathbf{D}} \leq u|\mathbf{D},\mathbf{U})\}] \\
=&\int_0^\tau \mathrm{E}_{D|\mathbf{U}}[\Pr(U_{C|\mathbf{U},\mathbf{D}} \geq u) d\{Pr(U_{\mathbf{D}} \leq u|\mathbf{D},\mathbf{U})\}] \\
=&\mathrm{E}_{D|\mathbf{U}}[\int_0^\tau \Pr(U_{C|\mathbf{U},\mathbf{D}} \geq u) d\{Pr(U_{\mathbf{D}} \leq u|\mathbf{D},\mathbf{U})\}] \\
=&\mathrm{E}_{D|\mathbf{U}}(\Pr[T \leq \exp\{q(\mathbf{D},\mathbf{X},\tau)\}, C \geq T|\mathbf{D},\mathbf{U}]) \\
=&\mathrm{E}(I[T \leq \exp\{q(\mathbf{D},\mathbf{X},\tau)\}, C \geq T]|\mathbf{U}).
\end{aligned} \tag{4.8}$$

The second equation holds because $U_{C|\mathbf{D},\mathbf{U}} \perp\!\!\!\perp (U_{\mathbf{D}}|\mathbf{D},\mathbf{U})|\mathbf{U}$.

For the RHS of equation (4.7), we have

$$\begin{aligned}
&\int_0^\tau \Pr\{\log C \geq q(\mathbf{D},\mathbf{X},u)|\mathbf{U}\}\Pr\{\log T \geq q(\mathbf{D},\mathbf{X},u)|\mathbf{U}\}\frac{du}{1-u} \\
=&\int_0^\tau \mathrm{E}_{D|\mathbf{U}}\{\Pr(U_{C|\mathbf{U},\mathbf{D}} \geq u)\}\mathrm{E}_{D|\mathbf{U}}\{Pr(U_{\mathbf{D}} \geq u|\mathbf{D},\mathbf{U})\}\frac{du}{1-u} \\
=&\int_0^\tau \mathrm{E}_{D|\mathbf{U}}[\Pr(U_{C|\mathbf{U},\mathbf{D}} \geq u) Pr(U_{\mathbf{D}} \geq u|\mathbf{D},\mathbf{U})]\frac{du}{1-u} \\
=&\mathrm{E}_{D|\mathbf{U}}\Big[\int_0^\tau \Pr(U_{C|\mathbf{U},\mathbf{D}} \geq u) Pr(U_{\mathbf{D}} \geq u|\mathbf{D},\mathbf{U})\frac{du}{1-u}\Big] \\
=&\mathrm{E}_{D|\mathbf{U}}\Big[\int_0^\tau Pr[W \geq \exp\{q(\mathbf{D},\mathbf{X},u)\}|\mathbf{D},\mathbf{U}]\frac{du}{1-u}\Big] \\
=&\mathrm{E}_{D|\mathbf{U}}\Big(\mathrm{E}\Big[\int_0^\tau I[W \geq \exp\{q(\mathbf{D},\mathbf{X},u)\}]\frac{du}{1-u}|\mathbf{D},\mathbf{U}\Big]\Big) \\
=&\mathrm{E}\Big[\int_0^\tau I[W \geq \exp\{q(\mathbf{D},\mathbf{X},u)\}]\frac{du}{1-u}|\mathbf{U}\Big].
\end{aligned} \tag{4.9}$$

The second equation holds because $U_{C|\mathbf{D},\mathbf{U}} \perp\!\!\!\perp (U_{\mathbf{D}}|\mathbf{D},\mathbf{U})|\mathbf{U}$.

Note that $N(t) = I(W \leq t, \delta = 1)$. Plugging equations (4.8) and (4.9) into (4.7), we can have

$$\mathrm{E}\Big(N_i[\exp\{q(\mathbf{D}_i,\mathbf{X}_i,u)\}] - \int_0^\tau I[W_i \geq \exp\{q(\mathbf{D}_i,\mathbf{X}_i,u)\}]\frac{du}{1-u}|\mathbf{U}_i\Big) = 0,$$

Then, we can get

$$n^{1/2}\boldsymbol{Q}_n(\boldsymbol{\beta},\boldsymbol{\gamma},\tau)=0,$$

where $\boldsymbol{Q}_n(\boldsymbol{\beta},\boldsymbol{\gamma},\tau)=n^{-1}\sum_{i=1}^{n}\mathbf{U}_i\Big\{N_i(\exp\{\mathbf{D}_i^\top\boldsymbol{\beta}(\tau)+\mathbf{X}_i^\top\boldsymbol{\gamma}(\tau)\})-\int_0^\tau I(W_i\geq\exp\{\mathbf{D}_i^\top\boldsymbol{\beta}(u)+\mathbf{X}_i^\top\boldsymbol{\gamma}(u)\})dH(u)\Big\}$. □

## 4.6.2 Appendix F: Proofs of Theorems 4.1 and 4.2

### 4.6.2.1 Proof of Theorem 4.1

*Proof.* Define $\boldsymbol{\mu}(\boldsymbol{\theta})=E[\mathbf{U}N\{\exp(\mathbf{Z}^\top\boldsymbol{\theta})\}]$ and $\mathbf{v}_n(\boldsymbol{\theta})=n^{-1}\sum_{i=1}^{n}\mathbf{U}_iN\{\exp(\mathbf{Z}_i^\top\boldsymbol{\theta})\}-\boldsymbol{\mu}\{\boldsymbol{\theta}\}$. Define $\tilde{\boldsymbol{\mu}}(\boldsymbol{\theta})=E[\mathbf{U}I\{W\geq\exp(\mathbf{Z}^\top\boldsymbol{\theta})\}]$ and $\tilde{\mathbf{v}}_n(\boldsymbol{\theta})=n^{-1}\sum_{i=1}^{n}\mathbf{U}_iI\{W_i\geq\mathbf{Z}_i^\top\boldsymbol{\theta}\}-\boldsymbol{\mu}\{\boldsymbol{\theta}\}$. Moreover, denote $\boldsymbol{\mu}_{\mathbf{x}}(\boldsymbol{\theta})=E[\mathbf{X}N\{\exp(\mathbf{Z}^\top\boldsymbol{\theta})\}]$ and $\mathbf{v}_{\mathbf{x},n}(\boldsymbol{\theta})=n^{-1}\sum_{i=1}^{n}\mathbf{X}_iN\{\exp(\mathbf{Z}_i^\top\boldsymbol{\theta})\}-\boldsymbol{\mu}_{\mathbf{x}}\{\boldsymbol{\theta}\}$. Define $\tilde{\boldsymbol{\mu}}_{\mathbf{x}}(\boldsymbol{\theta})=E[\mathbf{X}I\{W\geq\exp(\mathbf{Z}^\top\boldsymbol{\theta})\}]$ and $\tilde{\mathbf{v}}_{\mathbf{x},n}(\boldsymbol{\theta})=n^{-1}\sum_{i=1}^{n}\mathbf{X}_iI\{W_i\geq\exp(\mathbf{Z}_i^\top\boldsymbol{\theta})\}-\boldsymbol{\mu}\{\boldsymbol{\theta}\}$. Without loss of generality, we assume that $\tau_1<\tau_2<\cdots<\tau_L-1$ are equally spaced between 0 and $\tau_U$. Let $a_n=\|S_L\|$, then $L=\tau_U/a_n$. Besides, let $b_n=a_n/(1-\tau_U)$, and thus we have$0<H(\tau_j)-H(\tau_{j-1})\leq b_n$ for $j=1,\cdots,L$

Besides, let $\hat{\boldsymbol{\gamma}}(\boldsymbol{\beta}_0(\tau),\tau)$ denote the generalized solution of

$$\mathbf{S}_n(\hat{\boldsymbol{\beta}},\hat{\boldsymbol{\gamma}},\boldsymbol{\beta}_0(\tau),\boldsymbol{\gamma},\tau)=n^{-1}\sum_{i=1}^{n}\mathbf{X}_i\Big\{N_i(\exp\{\mathbf{D}_i^\top\boldsymbol{\beta}_0(\tau)+\mathbf{X}_i^\top\boldsymbol{\gamma}\})-\int_0^\tau I(W_i\geq\exp\{\mathbf{D}_i^\top\hat{\boldsymbol{\beta}}(u)+\mathbf{X}_i^\top\hat{\boldsymbol{\gamma}}(u)\})dH(u)\Big\}.$$

Let $\mathcal{A}(d)=\{\boldsymbol{\mu}(\boldsymbol{\theta}):\boldsymbol{\theta}\in\mathcal{B}(d)\}$. We can show $\boldsymbol{\mu}$ is a one-to-one map from $\mathcal{B}(d_0)$

to $\mathcal{A}(d_0)$. Suppose $\boldsymbol{\theta}'$ and $\boldsymbol{\theta} \in \mathbb{B}(d_0)$ such that $\boldsymbol{\mu}(\boldsymbol{\theta}') = \boldsymbol{\mu}(\boldsymbol{\theta})$. Then

$$\begin{aligned}
0 &= (\boldsymbol{\theta}' - \boldsymbol{\theta}')\{\boldsymbol{\mu}(\boldsymbol{\theta}') - \boldsymbol{\mu}(\boldsymbol{\theta})\} \\
&= \mathrm{E}\big((\mathbf{U}^\top\boldsymbol{\theta}' - \mathbf{U}^\top\boldsymbol{\theta})[\tilde{F}\{\exp(\mathbf{Z}^\top\boldsymbol{\theta}')|\mathbf{Z}\} - \tilde{F}\{\exp(\mathbf{Z}^\top\boldsymbol{\theta})|\mathbf{Z}\}]\big) \\
&= \mathrm{E}\big(\tilde{f}\{\exp(\mathbf{Z}^\top\tilde{\boldsymbol{\theta}})|\mathbf{Z}\}\exp(\mathbf{Z}^\top\tilde{\boldsymbol{\theta}})(\boldsymbol{\theta}' - \boldsymbol{\theta})^\top\mathbf{U}\mathbf{Z}^\top(\boldsymbol{\theta}' - \boldsymbol{\theta})\big) \\
&= (\boldsymbol{\theta}' - \boldsymbol{\theta})^\top\mathbf{B}(\tilde{\boldsymbol{\theta}})(\boldsymbol{\theta}' - \boldsymbol{\theta}) \\
&= (\boldsymbol{\theta}' - \boldsymbol{\theta})^\top\mathbf{B}^S(\tilde{\boldsymbol{\theta}})(\boldsymbol{\theta}' - \boldsymbol{\theta}),
\end{aligned}$$

where $\tilde{\boldsymbol{\theta}}$ is between $\boldsymbol{\theta}$ and $\boldsymbol{\theta}'$. By condition C3(i) and C4(i), the foregoing equation holds if and only if $\boldsymbol{\theta}' = \boldsymbol{\theta}$ with probability 1. Thus, there exists an inverse function of $\boldsymbol{\mu}$ denoted by $\boldsymbol{\kappa}$ from $\mathcal{A}(d_0)$ to $\mathcal{B}(d_0)$, such that $\boldsymbol{\kappa}(\boldsymbol{\mu}(\boldsymbol{\theta})) = \boldsymbol{\theta}$

Besides, let $\mathcal{A}_{\boldsymbol{\beta}}(d) = \{\boldsymbol{\mu}_x(\boldsymbol{\beta}, \boldsymbol{\gamma}) : (\boldsymbol{\beta}^\top, \boldsymbol{\gamma}^\top)^\top \in \mathcal{B}(d)\}$ and $\mathcal{B}_{\boldsymbol{\beta}}(d) = \{\boldsymbol{\gamma} : (\boldsymbol{\beta}^\top, \boldsymbol{\gamma}^\top)^\top \in \mathcal{B}(d)\}$. Given $\boldsymbol{\beta}$, we can show $\boldsymbol{\mu}_x$ is a one-to-one map from $\mathcal{B}_{\boldsymbol{\beta}}(d_0)$ to $\mathcal{A}_{\boldsymbol{\beta}}(d_0)$. Suppose $\boldsymbol{\gamma}'$ and $\boldsymbol{\gamma} \in \mathcal{B}_{\boldsymbol{\beta}}(d_0)$, and $\boldsymbol{\mu}_x(\boldsymbol{\beta}, \boldsymbol{\gamma}') = \boldsymbol{\mu}_x(\boldsymbol{\beta}, \boldsymbol{\gamma})$. Then

$$\begin{aligned}
0 &= (\boldsymbol{\gamma}' - \boldsymbol{\gamma})^\top\{\boldsymbol{\mu}_x(\boldsymbol{\beta}, \boldsymbol{\gamma}') - \boldsymbol{\mu}_x(\boldsymbol{\beta}, \boldsymbol{\gamma})\} \\
&= \mathrm{E}\big((\mathbf{X}^\top\boldsymbol{\gamma}' - \mathbf{X}^\top\boldsymbol{\gamma})[\tilde{F}\{\exp(\mathbf{D}^\top\boldsymbol{\beta} + \mathbf{X}^\top\boldsymbol{\gamma}')|\mathbf{Z}\} - \tilde{F}\{\exp(\mathbf{D}^\top\boldsymbol{\beta} + \mathbf{X}^\top\boldsymbol{\gamma})|\mathbf{Z}\}]\big)
\end{aligned}$$

By condition C3(i), the foregoing equation holds if and only if $\mathbf{X}^\top\boldsymbol{\gamma}' = \mathbf{X}^\top\boldsymbol{\gamma}$ with probability 1. Because of the positive definiteness of $E(\mathbf{X}^{\otimes 2})$, we have $\boldsymbol{\gamma}' = \boldsymbol{\gamma}$. Therefore, there exists an inverse function of $\boldsymbol{\mu}_{\mathbf{x}}(\boldsymbol{\beta}, \cdot)$ denoted by $\boldsymbol{\kappa}_{\boldsymbol{\beta}}$ from $\mathcal{A}_{\boldsymbol{\beta}}(d_0)$ to $\mathcal{B}_{\boldsymbol{\beta}}(d_0)$, such that $\boldsymbol{\kappa}_{\boldsymbol{\beta}}(\boldsymbol{\mu}_{\mathbf{x}}(\boldsymbol{\beta}, \boldsymbol{\gamma})) = \boldsymbol{\gamma}$.

For $j = 1, \cdots, L$, we have

$$\begin{aligned}
& n^{-1}\sum_{i=1}^{n}\mathbf{X}_i N_i(\exp\{\mathbf{D}_i^\top\boldsymbol{\beta}_0(\tau_j) + \mathbf{X}_i^\top\hat{\boldsymbol{\gamma}}(\boldsymbol{\beta}_0(\tau_j), \tau_j)\}) \\
=& n^{-1}\sum_{i=1}^{n}\int_0^{\tau_j}\mathbf{X}_i I(W_i \geq \exp\{\mathbf{D}_i^\top\hat{\boldsymbol{\beta}}(u) + \mathbf{X}_i^\top\hat{\boldsymbol{\gamma}}(u)\})dH(u) + \boldsymbol{\xi}_{n,j}
\end{aligned}$$

where $\sup_j \|\boldsymbol{\xi}_{n,j}\| \leq \sup_i \|\mathbf{X}_i\|/n$.

Denote $\tilde{\boldsymbol{\theta}}(\tau) = (\boldsymbol{\beta}_0^\top(\tau), \hat{\boldsymbol{\gamma}}(\boldsymbol{\beta}_0(\tau), \tau))$. Since $\boldsymbol{\mu}_{\mathbf{x}}(\boldsymbol{\theta}_0(\tau_j)) = \int_0^{\tau_j} \tilde{\boldsymbol{\mu}}_{\mathbf{x}}\{\boldsymbol{\theta}_0(u)\}dH(u)$, we have

$$\begin{aligned}
&\boldsymbol{\mu}_{\mathbf{x}}\{\tilde{\boldsymbol{\theta}}(\tau_j)\} - \boldsymbol{\mu}_{\mathbf{x}}\{\boldsymbol{\theta}_0(\tau_j)\} \\
&= -\mathbf{v}_{\mathbf{x},n}\{\tilde{\boldsymbol{\theta}}(\tau_j)\} + \int_0^{\tau_j} \tilde{\mathbf{v}}_{\mathbf{x},n}\{\hat{\boldsymbol{\theta}}(u)\}dH(u) + \sum_{k=1}^{j}\int_{\tau_{k-1}}^{\tau_k} [\tilde{\boldsymbol{\mu}}_{\mathbf{x}}\{\hat{\boldsymbol{\theta}}(u)\} - \tilde{\boldsymbol{\mu}}_{\mathbf{x}}\{\boldsymbol{\theta}_0(u)\}]dH(u) + \boldsymbol{\xi}_{n,j}
\end{aligned} \tag{4.10}$$

Following the similar arguments in the Proof of Theorem 1 in Peng and Huang (2008) and Glivenko-Cantelli theorem, we have $\sup_{\boldsymbol{\theta}\in\mathbb{R}^{p+q}} \|\mathbf{v}_{\mathbf{x},n}(\boldsymbol{\theta})\}\| \to_{a.s.} 0$, $\sup_{\boldsymbol{\theta}\in\mathbb{R}^{p+q}} \|\mathbf{v}_n(\boldsymbol{\theta})\}\| \to_{a.s.} 0$, $\sup_{\boldsymbol{\theta}\in\mathbb{R}^{p+q}} \|\tilde{\mathbf{v}}_n(\boldsymbol{\theta})\}\| \to_{a.s.} 0$ and $\sup_{\boldsymbol{\theta}\in\mathbb{R}^{p+q}} \|\tilde{\mathbf{v}}_{\mathbf{x},n}(\boldsymbol{\theta})\}\| \to_{a.s.} 0$. Therefore, for any given $C_1 > 0$, $\|-\mathbf{v}_{\mathbf{x},n}\{\tilde{\boldsymbol{\theta}}(\tau_j)\} + \int_0^{\tau_j} \tilde{\mathbf{v}}_{\mathbf{x},n}\{\hat{\boldsymbol{\theta}}(u)\}dH(u)\| \le C_1$ and $\|-\mathbf{v}_n\{\tilde{\boldsymbol{\theta}}(\tau_j)\} + \int_0^{\tau_j} \tilde{\mathbf{v}}_n\{\hat{\boldsymbol{\theta}}(u)\}dH(u)\| \le C_1$ with probability 1 as $n$ is sufficiently large. By condition C1, there exists $C_2 > 0$, such that $\sup_i \|X_i\| < C_2$. Condition C2(i) implies there exists constant $C_3 > 0$, such that $\|\boldsymbol{\mu}\{\theta_0(\tau')\} - \boldsymbol{\mu}\{\theta_0(\tau)\}\| \le C_3|\tau' - \tau|$ for any $\tau', \tau \in (0, \tau_U]$. From the condition C3(iii), we have $\|\{\mathbf{B}(\boldsymbol{\theta})\}^{-1}J(\boldsymbol{\theta})\mathbf{y}\| \le C_4\|\mathbf{y}\|$ for any $\mathbf{y} \in \mathbb{R}^{p+q}$ and $\boldsymbol{\theta} \in \mathcal{D}_{d_0}$. By the condition C1, C2(ii) and C3(iv), for $(\boldsymbol{\beta}_0(\tau_j), \boldsymbol{\gamma}_1), (\boldsymbol{\beta}_0(\tau_j), \boldsymbol{\gamma}_2) \in \mathcal{B}_{d_0}$, there exists a constant $C_5$, such that $\sup_{j,\mathbf{Z},\boldsymbol{\gamma}_1,\boldsymbol{\gamma}_2} \|\{\mathbf{U}\tilde{f}(\exp(\mathbf{D}^\top\boldsymbol{\beta}_0(\tau_j)+\mathbf{X}^\top\boldsymbol{\gamma}_1)|\mathbf{Z})\mathbf{X}^\top \exp(\mathbf{D}^\top\boldsymbol{\beta}_0(\tau_j)+\mathbf{X}^\top\boldsymbol{\theta}_2)\mathbf{B}_x[\boldsymbol{\beta}_0(\tau_j), \boldsymbol{\theta}_3]^{-1}\} \le C_5$. Besides, let $C_6 = C_5 + 2$

Define $\epsilon_0 = C_3a_n$, $\epsilon_1 = C_6(C_1 + \epsilon_0C_4b_n + C_2/n + C_3a_n)$, and $\epsilon_l = C_6(C_1 + \sum_{i=1}^{l-1}(\epsilon_i)C_4b_n + C_2/n + C_3a_n)$ for $l = 2, \cdots, L-1$. We can obtain that $\epsilon_l = (1 + C_6C_4b_n)^{l-1}(C_1 + \epsilon_0C_4b_n + C_2/n + C_3a_n)$. Given $\lim_{n\to\infty} a_n = 0$ and $L = \tau_U/a_n$, we have $\lim_{n\to\infty}(1 + C_6C_4b_n)^{l-1} = \exp\{C_6C_4\tau_U/(1-\tau_U)\}$. Because $\epsilon_l$ is increasing with $l$, it is easy to see that for $N \ge$ some $N_0$, $C_0$ can be chosen sufficiently small so that $\epsilon_l \le 2\exp\{C_6C_4\tau_U/(1-\tau_U)\}C_1 < d_0$, for $l = 0, \cdots L-1$.

Next, we show $\epsilon_l$ is an upper bound for $\sup_{\tau\in[\tau_l,\tau_{l+1})} \|\boldsymbol{\mu}\{\hat{\boldsymbol{\theta}}(\tau)\} - \boldsymbol{\mu}\{\boldsymbol{\theta}_0(\tau)\}\|$. From the definition of $\hat{\boldsymbol{\theta}}(\tau)$, it is easy to see that $\sup_{\tau\in[\tau_0,\tau_1)} \|\boldsymbol{\mu}\{\hat{\boldsymbol{\theta}}(\tau)\} - \boldsymbol{\mu}\{\boldsymbol{\theta}_0(\tau)\}\| =$

$\sup_{\tau\in[\tau_0,\tau_1)} \|\boldsymbol{\mu}\{\boldsymbol{\theta}_0(\tau)\}\| \leq C_3 a_n = \epsilon_0$. Following the similar argument in Peng and Huang (2008), for $\tau \in [\tau_0, \tau_1)$,

$$\|\boldsymbol{\mu}_x\{\hat{\boldsymbol{\theta}}(\tau)\} - \boldsymbol{\mu}_x\{\boldsymbol{\theta}_0(\tau)\}\| \leq \|\boldsymbol{\mu}\{\hat{\boldsymbol{\theta}}(\tau)\} - \boldsymbol{\mu}\{\boldsymbol{\theta}_0(\tau)\}\| \leq C_4\epsilon_0$$

Thus, from (4.10),

$$\begin{aligned}
&\sup_{\tau\in[\tau_1,\tau_2)} \|\boldsymbol{\mu}_x\{\tilde{\boldsymbol{\theta}}(\tau)\} - \boldsymbol{\mu}_x\{\boldsymbol{\theta}_0(\tau)\}\| \\
\leq &\|\boldsymbol{\mu}_x\{\tilde{\boldsymbol{\theta}}(\tau_1)\} - \boldsymbol{\mu}_x\{\boldsymbol{\theta}_0(\tau_1)\}\| + \sup_{\tau\in[\tau_1,\tau_2)} \|\boldsymbol{\mu}_x\{\boldsymbol{\theta}(\tau)\} - \boldsymbol{\mu}_x\{\boldsymbol{\theta}_0(\tau_1)\}\| \\
\leq &C_1 + \epsilon_0 C_4 b_n + C_2/n + C_3 a_n
\end{aligned}$$

By the definition of $\hat{\boldsymbol{\beta}}(\tau_1)$, we know that

$$\|\mathbf{U}_n(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{\beta}}(\tau_1), \tau_1)\| \leq \|\mathbf{U}_n(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}}, \boldsymbol{\beta}_0(\tau_1), \tau_1)\|, \tag{4.11}$$

where $\mathbf{U}_n(\hat{\boldsymbol{\theta}}, \boldsymbol{\beta}, \tau_1) = n^{-1}\sum_{i=1}^n \mathbf{U}_i\Big\{N_i(\exp\{\mathbf{D}_i^\top\boldsymbol{\beta}+\mathbf{X}_i^\top\hat{\boldsymbol{\gamma}}(\boldsymbol{\beta},\tau_1)\})-\int_0^{\tau_1} I(W_i \geq \exp\{\mathbf{Z}_i^\top\hat{\boldsymbol{\theta}}(u)\})dH(u)\Big\}$
Coupled with $\boldsymbol{\mu}(\boldsymbol{\theta}_0(\tau_1)) = \int_{\tau_0}^{\tau_1} \tilde{\boldsymbol{\mu}}(\boldsymbol{\theta}_0(u))dH(u)$, we have

$$\begin{aligned}
&\mathbf{U}_n(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\beta}}(\tau_1), \tau_1) \\
=&\boldsymbol{\mu}\{\hat{\boldsymbol{\theta}}(\tau_1)\} - \boldsymbol{\mu}\{\boldsymbol{\theta}_0(\tau_1)\} - \int_{\tau_0}^{\tau_1} (\tilde{\boldsymbol{\mu}}\{\hat{\boldsymbol{\theta}}(u)\} - \tilde{\boldsymbol{\mu}}\{\boldsymbol{\theta}_0(u)\})dH(u) \\
&+ \mathbf{v}_n\{\hat{\boldsymbol{\theta}}(\tau_1)\} - \int_0^{\tau_1} \tilde{\mathbf{v}}_n\{\hat{\boldsymbol{\theta}}(u)\}dH(u)
\end{aligned} \tag{4.12}$$

and

$$\begin{aligned}
&\mathbf{U}_n(\hat{\boldsymbol{\theta}}, \boldsymbol{\beta}_0(\tau_1), \tau_1) \\
=&\boldsymbol{\mu}\{\tilde{\boldsymbol{\theta}}(\tau_1)\} - \boldsymbol{\mu}\{\boldsymbol{\theta}_0(\tau_1)\} - \int_{\tau_0}^{\tau_1} (\tilde{\boldsymbol{\mu}}\{\hat{\boldsymbol{\theta}}(u)\} - \tilde{\boldsymbol{\mu}}\{\boldsymbol{\theta}_0(u)\})dH(u) \\
&+ \mathbf{v}_n\{\tilde{\boldsymbol{\theta}}(\tau_1)\} - \int_0^{\tau_1} \tilde{\mathbf{v}}_n\{\hat{\boldsymbol{\theta}}(u)\}dH(u).
\end{aligned} \tag{4.13}$$

From (4.11), (4.12) and (4.13), we have

$$\begin{aligned}
&\|\boldsymbol{\mu}(\hat{\boldsymbol{\theta}}(\tau_1)) - \boldsymbol{\mu}(\boldsymbol{\theta}_0(\tau_1))\| \\
\leq &\|\boldsymbol{\mu}\{\tilde{\boldsymbol{\theta}}(\tau_1)\} - \boldsymbol{\mu}\{\boldsymbol{\theta}_0(\tau_1)\}\| + 2\Big\| \int_{\tau_0}^{\tau_1} (\tilde{\boldsymbol{\mu}}\{\hat{\boldsymbol{\theta}}(u)\} - \tilde{\boldsymbol{\mu}}\{\boldsymbol{\theta}_0(u)\}) dH(u)\Big\| \\
&+ \|\mathbf{v}_n\{\hat{\boldsymbol{\theta}}(\tau_1)\} - \int_0^{\tau_1} \tilde{\mathbf{v}}_n\{\hat{\boldsymbol{\theta}}(u)\} dH(u)\| + \|\mathbf{v}_n\{\tilde{\boldsymbol{\theta}}(\tau_1)\} - \int_0^{\tau_1} \tilde{\mathbf{v}}_n\{\hat{\boldsymbol{\theta}}(u)\} dH(u)\|
\end{aligned}$$

Since

$$\begin{aligned}
&\boldsymbol{\mu}\{\tilde{\boldsymbol{\theta}}(\tau_1)\} - \boldsymbol{\mu}\{\boldsymbol{\theta}_0(\tau_1)\} \\
=&\mathrm{E}\{\mathbf{U}\tilde{f}(\exp(\mathbf{Z}^\top\check{\boldsymbol{\theta}})|\mathbf{Z})[\exp(\mathbf{Z}^\top\tilde{\boldsymbol{\theta}}(\tau_1) - \exp(\mathbf{Z}^\top\boldsymbol{\theta}_0(\tau_1)]\} \\
=&\mathrm{E}\{\mathbf{U}\tilde{f}(\exp(\mathbf{Z}^\top\check{\boldsymbol{\theta}})|\mathbf{Z})[\exp(\mathbf{D}^\top\boldsymbol{\beta}_0 + \mathbf{X}^\top\boldsymbol{\kappa}_{\boldsymbol{\beta}_0(\tau_1)}(\boldsymbol{\mu}_x\{\tilde{\boldsymbol{\theta}}(\tau_1)\})) - \exp(\mathbf{D}^\top\boldsymbol{\beta}_0 + \mathbf{X}^\top\boldsymbol{\kappa}_{\boldsymbol{\beta}_0(\tau_1)}(\boldsymbol{\mu}_x\{\boldsymbol{\theta}_0(\tau_1)\}))]\} \\
=&\mathrm{E}\{\mathbf{U}\tilde{f}(\exp(\mathbf{Z}^\top\check{\boldsymbol{\theta}})|\mathbf{Z})\mathbf{X}^\top\exp(\mathbf{D}^\top\boldsymbol{\beta}_0 + \mathbf{X}^\top\boldsymbol{\kappa}_{\boldsymbol{\beta}_0(\tau_1)}\{\check{\mathbf{a}}\})\mathbf{B}_x[\boldsymbol{\kappa}_{\boldsymbol{\beta}_0(\tau_1)}\{\check{\mathbf{a}}\}]^{-1}[\boldsymbol{\mu}_x\{\tilde{\boldsymbol{\theta}}(\tau_1)\} - \boldsymbol{\mu}_x\{\boldsymbol{\theta}_0(\tau_1)\}]\}
\end{aligned}$$

where $\check{\mathbf{a}}$ is on the line segment between $\boldsymbol{\mu}_x\{\tilde{\boldsymbol{\theta}}(\tau_1)\}$ and $\boldsymbol{\mu}_x\{\boldsymbol{\theta}_0(\tau_1)\}$, and $\check{\boldsymbol{\theta}}$ is on the line segment between $\tilde{\boldsymbol{\theta}}(\tau_1)$ and $\boldsymbol{\theta}_0(\tau_1)$. Then $\|\boldsymbol{\mu}(\hat{\boldsymbol{\theta}}(\tau_1)) - \boldsymbol{\mu}(\boldsymbol{\theta}_0(\tau_1))\| \leq C_5(C_1 + \epsilon_0 C_4 b_n + C_2/n + C_3 a_n) + 2\epsilon_0 C_4 b_n + 2C_1$, and thus

$$\begin{aligned}
&\sup_{\tau\in[\tau_1,\tau_2)} \|\boldsymbol{\mu}(\hat{\boldsymbol{\theta}}(\tau)) - \boldsymbol{\mu}(\boldsymbol{\theta}_0(\tau))\| \\
\leq &\|\boldsymbol{\mu}(\hat{\boldsymbol{\theta}}(\tau_1)) - \boldsymbol{\mu}(\boldsymbol{\theta}_0(\tau_1))\| + \sup_{\tau\in[\tau_1,\tau_2)} \|\boldsymbol{\mu}(\boldsymbol{\theta}_0(\tau)) - \boldsymbol{\mu}(\boldsymbol{\theta}_0(\tau_1))\| \\
\leq &C_5(C_1 + \epsilon_0 C_4 b_n + C_2/n + C_3 a_n) + 2\epsilon_0 C_4 b_n + 2C_1 + C_3 a_n \\
\leq &C_6(C_1 + \epsilon_0 C_4 b_n + C_2/n + C_3 a_n).
\end{aligned}$$

Inductively for $l = 2, \cdots, L-1$, we can obtain $\sup_{\tau\in[\tau_l,\tau_{l+1})} \|\boldsymbol{\mu}\{\hat{\boldsymbol{\theta}}(\tau)\} - \boldsymbol{\mu}\{\boldsymbol{\theta}_0(\tau)\| \leq \epsilon_l$ and $\hat{\boldsymbol{\theta}}(\tau_l) \in \mathcal{B}(d_0)$. Since $a_n = o(1)$, $b_n = o(1)$, and $C_1$ can be arbitrary small as $n$ increases, it follows that $\sup_{\tau\in(0,\tau_U]} \|\boldsymbol{\mu}\{\hat{\boldsymbol{\theta}}(\tau)\} - \boldsymbol{\mu}\{\boldsymbol{\theta}_0(\tau)\}\| = o_p(1)$. Coupled with Condition C4 with taylor expansions of $\boldsymbol{\kappa}(\boldsymbol{\alpha}(\tau))$ around $\boldsymbol{\alpha}(\tau) = \boldsymbol{\mu}(\boldsymbol{\theta}_0(\tau))$ for $[v, \tau_U]$, we can apply the same line of arguments in the proof of Theorem 1 in Peng and

Huang (2008) to prove that $\sup_{\tau\in[v,\tau_U]} \|\hat{\boldsymbol{\theta}}(\tau) - \boldsymbol{\theta}_0(\tau)\| = o_p(1)$. □

#### 4.6.2.2 Lemma 4.1 and its proof

Before proving Theorem 4.2, we first prove the following Lemma 4.1.

**Lemma 4.1.** *For any* $\{\boldsymbol{\theta}(\tau), \tau \in (0, \tau_U]\}$, *which satisfies* $\sup_{\tau\in(0,\tau_U]} \|\boldsymbol{\mu}\{\boldsymbol{\theta}(\tau)\} - \boldsymbol{\mu}\{\boldsymbol{\theta}_0(\tau)\}\| \to_p 0$, *we have*

$$\sup_{\tau\in(0,\tau_U]} \Big\| n^{-1/2} \sum_{i=1}^{n} \mathbf{U}_i \big( N_i(\exp\{\mathbf{Z}_i^\top \boldsymbol{\theta}(\tau)\}) - N_i(\exp\{\mathbf{Z}_i^\top \boldsymbol{\theta}_0(\tau)\}) \big) - n^{1/2}[\boldsymbol{\mu}\{\boldsymbol{\theta}(\tau)\} - \boldsymbol{\mu}\{\boldsymbol{\theta}_0(\tau)\}] \Big\| \to_p 0$$

*Proof.* Denote $\mu_1(\boldsymbol{\theta}) = \mathrm{E}[N\{\exp(\mathbf{Z}^\top\boldsymbol{\theta})\}]$ and $\sigma_d^2(\boldsymbol{\theta}) = \mathrm{Var}[N(\exp\{\mathbf{Z}^\top\boldsymbol{\theta}\})-N(\exp\{\mathbf{Z}^\top\boldsymbol{\theta}_0(\tau)\})-\mu_1\{\boldsymbol{\theta}\}+\mu_1\{\boldsymbol{\theta}_0(\tau)\}]$. Following the arguments of Alexander et al. (1984) and Lai and Ying (1988), $\sigma_d^2(\boldsymbol{\theta}) \to_p 0$ is sufficient to prove Lemma 1. We can apply the same line of arguments in Lemma B.1. in Peng and Huang (2008) to prove $\sigma_d^2(\boldsymbol{\theta}) \to_p 0$, □

#### 4.6.2.3 Proof of Theorem 4.2

*Proof.* Let $o_I(a_n)$ denote a term which converges to 0 in probability uniformly in $\tau \in I$ after being divided by $a_n$, and $O_I(a_n)$ denote a term which is bounded in probability uniformly in $\tau \in I$ after being divided by $a_n$. From Theorem 2 and Lemma 1, we can apply the similar arguments in the proof of Theorem 3 in Peng and Huang (2008) and obtain that

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

The above estimating equation can be viewed as a stochastic differential equation for $n^{1/2}[\boldsymbol{\mu}\{\hat{\boldsymbol{\theta}}(\tau)\}-\boldsymbol{\mu}\{\boldsymbol{\theta}_0(\tau)\}]$. By using the production integration theory (Gill et al., 1990), we have

$$n^{1/2}[\boldsymbol{\mu}\{\hat{\boldsymbol{\theta}}(\tau)\}-\boldsymbol{\mu}\{\boldsymbol{\theta}_0(\tau)\}]=\boldsymbol{\Phi}\{-n^{1/2}\boldsymbol{Q}_n(\boldsymbol{\theta}_0,\tau)\}+o_{(0,\tau_U]}(1),$$

where $\boldsymbol{\Phi}$ is a map from $\mathcal{F}$ to $\mathcal{F}$, such that for $\boldsymbol{g}\in\mathcal{F}$,

$$\boldsymbol{\Phi}(\boldsymbol{g})(\tau)=\int_0^{\tau}\mathcal{I}(s,\tau)d\boldsymbol{g}(s)$$

□

with $\mathcal{I}(s,t)=\prod_{u\in(s,t]}[\boldsymbol{I}_{p+q}+J\{\boldsymbol{\theta}_0(u)B\{\boldsymbol{\theta}_0(u)^{-1}dH(u)$ and $\mathcal{F}=\{\boldsymbol{g}:(0,\tau_U]\rightarrow\mathbb{R}^{p+q}$, $\boldsymbol{g}(\cdot)$ is left continuous with right limit, $\boldsymbol{g}(0)=\mathbf{0}\}$

Note that $-n^{1/2}\boldsymbol{Q}_n(\boldsymbol{\theta}_0,\tau)=n^{1/2}\boldsymbol{\Psi}_i(\tau)$, where $\boldsymbol{\Psi}_i(\tau)=\mathbf{U}_i\Big\{N_i(\exp\{\mathbf{Z}_i^{\top}\boldsymbol{\theta}_0(\tau)\})-\int_0^{\tau}I(W_i\geq\exp\{\mathbf{Z}_i^{\top}\boldsymbol{\theta}_0(\tau)\})dH(u)\Big\}$. Given the boundedness of $\mathbf{U}_i$, $\mathbf{Z}_i$ and Lipschitz continuity of $\int_0^{\tau}I(W_i\geq\exp\{\mathbf{Z}_i^{\top}\boldsymbol{\theta}_0(\tau)\})dH(u)$ in $\tau$, we can easily obtain that $\{\boldsymbol{\Psi}_i(\tau):\tau\in(0,\tau_U]\}$ is a Donsker class. By the Donsker Theorem, $-n^{-1/2}\boldsymbol{Q}_n(\boldsymbol{\theta}_0,\tau)$ converges weakly to a tight Gaussian process $\mathbf{G}(\tau)$ with mean $\mathbf{0}$ and covariance matrix $\boldsymbol{\Sigma}(s,t)=E[\boldsymbol{\Psi}_1(\tau)\boldsymbol{\Psi}_1(\tau)^{\top}]$. Since $\boldsymbol{\Phi}$ is a linear operator, $\boldsymbol{\Phi}(\mathbf{G})(\tau)$ for $\tau\in[v,\tau_U]$ is Gaussian process. Note that $\mathbf{B}\{\boldsymbol{\theta}_0(\tau)\}^{-1}$ is uniformly bounded for $\tau\in[v,\tau_U]$ by condition C4. Coupled with continuous mapping theorem and the taylor expansion of $n^{1/2}[\boldsymbol{\mu}\{\hat{\boldsymbol{\theta}}(\tau)\}-\boldsymbol{\mu}\{\boldsymbol{\theta}_0(\tau)\}]$, we can obtain that $n^{1/2}\{\hat{\boldsymbol{\theta}}(\tau)-\boldsymbol{\theta}_0(\tau)\}$ converges weakly to $\mathbf{B}\{\boldsymbol{\theta}_0(\tau)\}^{-1}\boldsymbol{\Phi}(\mathbf{G})(\tau)$ for $\tau\in[v,\tau_U]$, which is a mean zero Gaussian process

### 4.6.3 Appendix G: Justification of the Proposed Covariance Estimator

The justification of the proposed covariance estimator can be shown by following the similar arguments in Sun et al. (2016). The key step is to proof $\sup_{u\in[v,\tau_U]} \|\mathbf{B}(\boldsymbol{\beta}_0(u),\boldsymbol{\gamma}_0(u))-n^{-1/2}\mathbf{E}_n(u)\mathbf{D}_n(u)^{-1}\| \to_p 0$ and $\sup_{u\in[v,\tau_U]} \|\mathbf{J}(\boldsymbol{\beta}_0(u),\boldsymbol{\gamma}_0(u))-n^{-1/2}\tilde{\mathbf{E}}_n(u)\mathbf{D}_n(u)^{-1}\| \to_p 0$.

Following the similar arguments in Sun et al. (2016), we can easily show $\sup_{(0,\tau_U]} \|\boldsymbol{\Omega}_n(u)\| = O_p(1)$ and $\sup_{(0,\tau_U]} \|\mathbf{E}_n(u)\| = O_p(1)$. Here $\|\cdot\|$ with a matrix argument means the entrywise Euclidean norm. Thus, we have

$$\lim_{n\to\infty} \sup_{u\in(0,\tau_U]} \|n^{-1/2}\{L_n(\mathbf{b}_{n,j}(u),\boldsymbol{c}_{n,j}(u))-L_n(\hat{\boldsymbol{\beta}}(u),\hat{\boldsymbol{\gamma}}(u))\}\| = \lim_{n\to\infty} \sup_{u\in(0,\tau_U]} \|n^{-1/2}\mathbf{e}_{n,j}(u)\| = 0 \tag{4.14}$$

Recall $\boldsymbol{\mu}(\mathbf{b},\boldsymbol{c}) = \mathrm{E}\{\mathbf{U}N(\exp(\mathbf{D}^\top\mathbf{b}+\mathbf{X}^\top\boldsymbol{c}))\}$. From Glivenko-Cantelli Theorem, we have $\sup_{u\in(0,\tau_U]} \|n^{-1/2}L_n(\mathbf{b}_{n,j}(u),\boldsymbol{c}_{n,j}(u)) - \boldsymbol{\mu}(\mathbf{b}_{n,j}(u),\boldsymbol{c}_{n,j}(u))\| = o_p(1)$ and $\sup_{u\in(0,\tau_U]} \|n^{-1/2}L_n(\hat{\boldsymbol{\beta}}(u),\hat{\boldsymbol{\gamma}}(u)) - \boldsymbol{\mu}(\hat{\boldsymbol{\beta}}(u),\hat{\boldsymbol{\gamma}}(u))\| = o_p(1)$.

Coupled with $\sup_{\tau\in(0,\tau_U]} \|\boldsymbol{\mu}\{\hat{\boldsymbol{\theta}}(\tau)\} - \boldsymbol{\mu}\{\boldsymbol{\theta}_0(\tau)\}\| = o_p(1)$ in the proof of Theorem 4.1, the above equations and Equation (4.14) can imply

$$\sup_{u\in(0,\tau_U]} \|\boldsymbol{\mu}(\mathbf{b}_{n,j}(u),\boldsymbol{c}_{n,j}(u)) - \boldsymbol{\mu}(\boldsymbol{\beta}_0(u),\boldsymbol{\gamma}_0(u))\| \to_p 0$$

From Lemma 4.1 and uniform consistency of $(\hat{\boldsymbol{\beta}},\hat{\boldsymbol{\gamma}})$, we have

$$\begin{aligned}
&L_n(\mathbf{b}_{n,j}(u),\boldsymbol{c}_{n,j}(u)) - L_n(\hat{\boldsymbol{\beta}}(u),\hat{\boldsymbol{\gamma}}(u)) \\
=&\{\mathbf{B}(\boldsymbol{\beta}_0(u),\boldsymbol{\gamma}_0(u)) + \epsilon_{n,j}(u)\}n^{1/2}\{(\mathbf{b}_{n,j}(u)^\top,\boldsymbol{c}_{n,j}(u)^\top)^\top - (\boldsymbol{\beta}_0(u)^\top,\boldsymbol{\gamma}_0(u)^\top)^\top\},
\end{aligned}$$

where $\sup_{u\in[v,\tau_U]} \epsilon_{n,j}(u) \to_p 0$. Therefore, $\sup_{u\in[v,\tau_U]} \|\mathbf{B}(\boldsymbol{\beta}_0(u),\boldsymbol{\gamma}_0(u))-n^{-1/2}\mathbf{E}_n(u)\mathbf{D}_n(u)^{-1}\| \to_p 0$.

Given the $\sup_{\tau\in[v,U]} \|(\hat{\boldsymbol{\beta}}(u)^\top,\hat{\boldsymbol{\gamma}}(u)^\top)^\top - (\boldsymbol{\beta}_0(u)^\top,\boldsymbol{\gamma}_0(u)^\top)^\top\| = o_p(1)$ from Theo-

rem 4.1 and $\sup_{\tau\in[v,U]} \|(\mathbf{b}_{n,j}(u)^\top, \boldsymbol{c}_{n,j}(u)^\top)^\top - (\boldsymbol{\beta}_0(u)^\top, \boldsymbol{\gamma}_0(u)^\top)^\top\| = o_p(1)$ implied by Equation (4.14), the following results can be shown by adopting similar argument in the proof of Theorem 4 in Huang and Peng (2009), we have

$$\sup_{u\in(0,\tau_U]} \|\tilde{L}_n(\mathbf{b}_n(u), \boldsymbol{c}_n(u)) - \tilde{L}_n(\hat{\boldsymbol{\beta}}(u), \hat{\boldsymbol{\gamma}}(u)) - n^{1/2}\{\tilde{\boldsymbol{\mu}}(\mathbf{b}_n(u), \boldsymbol{c}_n(u)) - \tilde{\boldsymbol{\mu}}(\boldsymbol{\beta}_0(u), \boldsymbol{\gamma}_0(u))\}\| = o_p(1) \tag{4.15}$$

where $(\mathbf{b}_n(u)^\top, \boldsymbol{c}_n(u)^\top)^\top$ can be either $(\mathbf{b}_{n,j}(u)^\top, \boldsymbol{c}_{n,j}(u)^\top)^\top$ or $(\hat{\boldsymbol{\beta}}(u)^\top, \hat{\boldsymbol{\gamma}}(u)^\top)^\top$, and we can show that

$$\begin{aligned}
&\tilde{L}_n(\mathbf{b}_{n,j}(u), \boldsymbol{c}_{n,j}(u)) - \tilde{L}_n(\hat{\boldsymbol{\beta}}(u), \hat{\boldsymbol{\gamma}}(u)) \\
=&\{\mathbf{J}(\boldsymbol{\beta}_0(u), \boldsymbol{\gamma}_0(u)) + \epsilon_{n,j}(u)\} n^{1/2}\{(\mathbf{b}_{n,j}(u)^\top, \boldsymbol{c}_{n,j}(u)^\top)^\top - (\boldsymbol{\beta}_0(u)^\top, \boldsymbol{\gamma}_0(u)^\top)^\top\},
\end{aligned}$$

$$\tilde{L}_n(\mathbf{b}_{n,j}(u), \boldsymbol{c}_{n,j}(u)) - \tilde{L}_n(\hat{\boldsymbol{\beta}}(u), \hat{\boldsymbol{\gamma}}(u)) = \{\mathbf{J}(\boldsymbol{\beta}_0(u), \boldsymbol{\gamma}_0(u)) + \epsilon^*_{n,j}(u)\} n^{1/2}\{(\mathbf{b}_{n,j}(u) - \hat{\boldsymbol{\beta}}(u), \boldsymbol{c}_{n,j}(u) - \hat{\boldsymbol{\gamma}}(u))$$

where $\sup_{u\in[v,\tau_U]} \epsilon^*_{n,j}(u) \to_p 0$. Therefore,

$$\sup_{u\in[v,\tau_U]} \|\mathbf{J}(\boldsymbol{\beta}_0(u), \boldsymbol{\gamma}_0(u)) - n^{-1/2}\tilde{\mathbf{E}}_n(u)\mathbf{D}_n(u)^{-1}\| \to_p 0$$

# Chapter 5

# Summary and Future Work

## 5.1 Summary

In this dissertation, we developed new statistical methods for two complex issues in survival analysis: recurrent events terminated by a dependent event and treatment selection bias. In this dissertation, we develop new statistical methods to handle these complications in survival analysis.

In the first project, we propose two sensible adaptations of the generalized accelerated recurrence time (GART) model (Sun et al., 2016) to handle the recurrent events terminated by a dependent event. The modeling strategies align with the rationale underlying the use of the survivors' rate function or the adjusted rate function to account for the presence of the dependent terminal event. We identify and develop estimation and inference procedures, and establish the asymptotic properties of the new estimator. Simulation studies demonstrate good finite-sample performance of the proposed methods. An application to a dataset from the Cystic Fibrosis Foundation Patient Registry (CFFPR) illustrate the practical utility of the new methods.

In the second project, we quantify causal treatment effect by complier quantile causal effect (CQCE), which is a meaningful counterpart of the complier average causal effect (CACE) that has been commonly studied in standard IV literature. For a time-to-event outcome subject to censoring, CQCE is identifiable under weaker conditions than CACE, which generally cannot be estimated with bounded censoring. CQCE also offers greater flexibility in depicting the causal treatment effect than other causal estimands in survival analysis, such as CCPHR and complier location shift effect. We develop a simple and rigorously justified two-stage estimation procedure, and elaborate how it can readily be implemented by existing software. The ease of implementation should facilitates future applications of the proposed method. We applied the proposed method to a dataset from the Center for International Blood and Marrow Transplant Research (CIBMTR) to evaluate the causal effect of rituximab in diffuse large B-cell lymphoma (DLBCL) patients.

In the third project, we propose an IV method to quantify the population quantile causal effect (PQCE) for standard right censoring data. PQCE is the quantile counterpart of population average causal effect (PACE), which has been commonly studies among IV literatures. Like the standard quantile treatment effect, PQCE requires weaker conditions for identifiability than PACE under censoring data (Klein and Moeschberger, 2006; Peng and Fine, 2007), and has advantages in capturing heterogeneous causal effects among potential outcome distribution. We propose a simple and justified two-stage estimation procedure to solve the numerical issues from the non-monotonicity of the estimating equation. Furthermore, a sample-based covariance estimator has been proposed to reduce the burden of computation in inference procedure. The proposed two-stage estimation procedure can easily be implemented by existing software for quantile regression with grid search. An application to a dataset from the Center for International Blood and Marrow Transplant Research (CIBMTR) demonstrates the practical utility of the proposed method.

## 5.2 Future work

In the second and third projects, we proposed new IV methods to estimate the complier quantile causal effect and global quantile causal effect. All proposed methods face challenges when the selected IV is a weak IV that is characterized by a weak association between IVs and treatment variables. when weak IVs are appeared, the proposed estimation and the associated inference can produce unstable and inaccurate results. Another potential problem from a weak IV is high sensitivity to the violation of the independence of IV assumption, as discussed in Baiocchi et al. (2014). That is, when the IV has only a minor association with unmeasured confounders, the resulting estimation bias can be exacerbated when the IV is weak. Therefore, we plan to extend the proposed method to incorporate the weak IV setting.

In the third project, the main assumption in this work is the rank similarity assumption, which restricts the evolution of the distribution of the ranks across treatment states to identify the PQCE (Chernozhukov and Hansen, 2005). Based on best of our knowledge, only a few approaches have been proposed to test this assumption (Frandsen and Lefgren, 2018; Dong and Shen, 2018). No existing testing approach has been proposed for censored data. We plan to develop methods to test rank similarity for censored data in future. Besides, the rank similarity assumption may not hold in practice, we want to investigate how to extend the proposed method without rank similarity assumption.

# Bibliography

Abadie, A. (2003), 'Semiparametric instrumental variable estimation of treatment response models', Journal of econometrics **113**(2), 231–263.

Abadie, A., Angrist, J. and Imbens, G. (2002), 'Instrumental variables estimates of the effect of subsidized training on the quantiles of trainee earnings', Econometrica **70**(1), 91–117.

Alexander, K. S., Talagrand, M. et al. (1989), 'The law of the iterated logarithm for empirical processes on vapnik-cervonenkis classes', Journal of Multivariate Analysis **30**(1), 155–166.

Alexander, K. S. et al. (1984), 'Probability inequalities for empirical processes and a law of the iterated logarithm', The Annals of Probability **12**(4), 1041–1067.

Andersen, P. K. and Gill, R. D. (1982), 'Cox's regression model for counting processes: a large sample study', The Annals of Statistics **10**(4), 1100–1120.

Angrist, J. D., Imbens, G. W. and Rubin, D. B. (1996a), 'Identification of causal effects using instrumental variables', Journal of the American statistical Association **91**(434), 444–455.

Angrist, J. D., Imbens, G. W. and Rubin, D. B. (1996b), 'Identification of causal effects using instrumental variables: Rejoinder', Journal of the American Statistical Association **91**(434), 468–472.

Angrist, J. D. and Krueger, A. B. (2001), 'Instrumental variables and the search for identification: From supply and demand to natural experiments', Journal of Economic perspectives **15**(4), 69–85.

Angrist, J. and Imbens, G. (1995), 'Identification and estimation of local average treatment effects'.

Austin, P. C., Lee, D. S. and Fine, J. P. (2016), 'Introduction to the analysis of survival data in the presence of competing risks', Circulation **133**(6), 601–609.

Baiocchi, M., Cheng, J. and Small, D. S. (2014), 'Instrumental variable methods for causal inference', Statistics in medicine **33**(13), 2297–2340.

Baker, S. G. (1998), 'Analysis of survival data from a randomized trial with all-or-none compliance: estimating the cost-effectiveness of a cancer screening program', Journal of the American Statistical Association **93**(443), 929–934.

Baker, S. G. and Lindeman, K. S. (1994), 'The paired availability design: a proposal for evaluating epidural analgesia during labor', Statistics in Medicine **13**(21), 2269–2278.

Bilias, Y., Gu, M., Ying, Z. et al. (1997), 'Towards a general asymptotic theory for cox model with staggered entry', The Annals of Statistics **25**(2), 662–682.

Blundell, R. and Powell, J. L. (2007), 'Censored regression quantiles with endogenous regressors', Journal of Econometrics **141**(1), 65–83.

Blundell, R. W. and Powell, J. L. (2004), 'Endogeneity in semiparametric binary response models', The Review of Economic Studies **71**(3), 655–679.

Cai, J. and Prentice, R. L. (1995), 'Estimating equations for hazard ratio parameters based on correlated failure time data', Biometrika **82**(1), 151–164.

Chen, S. (2018), 'Sequential estimation of censored quantile regression models', Journal of Econometrics **207**(1), 30–52.

Chen, S. and Khan, S. (2001), 'Semiparametric estimation of a partially linear censored regression model', Econometric Theory **17**(3), 567–590.

Cheng, J., Qin, J. and Zhang, B. (2009), 'Semiparametric estimation and inference for distributional and general treatment effects', Journal of the Royal Statistical Society: Series B (Statistical Methodology) **71**(4), 881–904.

Cheng, J., Small, D. S., Tan, Z. and Ten Have, T. R. (2009), 'Efficient nonparametric estimation of causal effects in randomized trials with noncompliance', Biometrika **96**(1), 19–36.

Chernozhukov, V., Fernández-Val, I. and Kowalski, A. E. (2015), 'Quantile regression with censoring and endogeneity', Journal of Econometrics **186**(1), 201–221.

Chernozhukov, V. and Hansen, C. (2005), 'An iv model of quantile treatment effects', Econometrica **73**(1), 245–261.

Chernozhukov, V. and Hansen, C. (2006), 'Instrumental quantile regression inference for structural and treatment effect models', Journal of Econometrics **132**(2), 491–525.

Chernozhukov, V. and Hansen, C. (2008), 'Instrumental variable quantile regression: A robust inference approach', Journal of Econometrics **142**(1), 379–398.

Chernozhukov, V. and Hong, H. (2002), 'Three-step censored quantile regression and extramarital affairs', Journal of the American Statistical Association **97**(459), 872–882.

Cook, R. J. and Lawless, J. (2007), The statistical analysis of recurrent events, Springer Science & Business Media.

Cook, R. J. and Lawless, J. F. (1997), 'Marginal analysis of recurrent events and a terminating event', Statistics in medicine **16**(8), 911–924.

Cuzick, J., Sasieni, P., Myles, J. and Tyrer, J. (2007), 'Estimating the effect of treatment in a proportional hazards model in the presence of non-compliance and contamination', Journal of the Royal Statistical Society: Series B (Statistical Methodology) **69**(4), 565–588.

Davies, J. C. (2002), 'Pseudomonas aeruginosa in cystic fibrosis: pathogenesis and persistence', Paediatric respiratory reviews **3**(2), 128–134.

Deaton, A. S. (2009), Instruments of development: Randomization in the tropics, and the search for the elusive keys to economic development, Technical report, National Bureau of Economic Research.

Deng, L., Song, Y., Young, K. H., Hu, S., Ding, N., Song, W., Li, X., Shi, Y., Huang, H., Liu, W. et al. (2015), 'Hepatitis b virus-associated diffuse large b-cell lymphoma: unique clinical features, poor outcome, and hepatitis b surface antigen-driven origin', Oncotarget **6**(28), 25061.

Dong, Y. and Shen, S. (2018), 'Testing for rank invariance or similarity in program evaluation', Review of Economics and Statistics **100**(1), 78–85.

Fenske, T. S., Hari, P. N., Carreras, J., Zhang, M.-J., Kamble, R. T., Bolwell, B. J., Cairo, M. S., Champlin, R. E., Chen, Y.-B., Freytes, C. O. et al. (2009), 'Impact of pre-transplant rituximab on survival after autologous hematopoietic stem cell transplantation for diffuse large b cell lymphoma', Biology of Blood and Marrow Transplantation **15**(11), 1455–1464.

Fine, J. P. and Gray, R. J. (1999), 'A proportional hazards model for the sub-distribution of a competing risk', Journal of the American statistical association **94**(446), 496–509.

Frandsen, B. R. and Lefgren, L. J. (2018), 'Testing rank similarity', Review of Economics and Statistics **100**(1), 86–91.

Fygenson, M. and Ritov, Y. (1994), 'Monotone estimating equations for censored data', The Annals of Statistics **22**(2), 732–746.

Ghosh, D. and Lin, D. (2002), 'Marginal regression models for recurrent and terminal events', Statistica Sinica **12**(3), 663–688.

Ghosh, D. and Lin, D. (2003), 'Semiparametric analysis of recurrent events data in the presence of dependent censoring', Biometrics **59**(4), 877–885.

Gill, R. D., Johansen, S. et al. (1990), 'A survey of product-integration with a view toward application in survival analysis', The annals of statistics **18**(4), 1501–1555.

Hansen, B. E. (2008), 'Uniform convergence rates for kernel estimation with dependent data', Econometric Theory pp. 726–748.

Harding, M. and Lamarche, C. (2012), Quantile regression estimation of panel duration models with censored data, in 'Essays in Honor of Jerry Hausman', Emerald Group Publishing Limited, pp. 237–267.

Heckman, J. J. and Robb Jr, R. (1985), 'Alternative methods for evaluating the impact of interventions: An overview', Journal of econometrics **30**(1-2), 239–267.

Heckman, J. J. and Urzua, S. (2010), 'Comparing iv with structural models: What simple iv can and cannot identify', Journal of Econometrics **156**(1), 27–37.

Heltshe, S., Khan, U., Beckett, V., Baines, A., Emerson, J., Sanders, D., Gibson, R., Morgan, W. and Rosenfeld, M. (2018), 'Longitudinal development of initial, chronic and mucoid pseudomonas aeruginosa infection in young children with cystic fibrosis', Journal of Cystic Fibrosis **17**(3), 341–347.

Hernán, M. A. (2010), 'The hazards of hazard ratios', Epidemiology (Cambridge, Mass.) **21**(1), 13.

Hong, H. and Tamer, E. (2003), 'Inference in censored models with endogenous regressors', Econometrica **71**(3), 905–932.

Horowitz, J. L. (2001), The bootstrap, in 'Handbook of econometrics', Vol. 5, Elsevier, pp. 3159–3228.

Hsieh, J.-J., Ding, A. A., Wang, W. and Chi, Y.-L. (2013), 'Quantile regression based on semi-competing risks data'.

Hsieh, J.-J. and Wang, H.-R. (2016), 'Quantile regression based on counting process approach under semi-competing risks data', Annals of the Institute of Statistical Mathematics **70**(2), 395–419.
**URL:** *http://dx.doi.org/10.1007/s10463-016-0593-6*

Huang, C.-Y. and Wang, M.-C. (2004), 'Joint modeling and estimation for recurrent event processes and failure time data', Journal of the American Statistical Association **99**(468), 1153–1165.

Huang, X. and Liu, L. (2007), 'A joint frailty model for survival and gap times between recurrent events', Biometrics **63**(2), 389–397.

Huang, Y. (2010), 'Quantile calculus and censored regression', Annals of statistics **38**(3), 1607.

Huang, Y. and Peng, L. (2009), 'Accelerated recurrence time models', Scandinavian Journal of Statistics **36**(4), 636–648.

Huling, J. D., Yu, M. and O'Malley, A. J. (2019), 'Instrumental variable based estimation under the semiparametric accelerated failure time model', Biometrics **75**(2), 516–527.

Imbens, G. and Angrist, J. (1994), 'Identification and estimation of local average treatment effects', Econometrica **62**, 467–476.

Imbens, G. W. and Lancaster, T. (1994), 'Combining micro and macro data in microeconometric models', The Review of Economic Studies **61**(4), 655–680.

Imbens, G. W. and Rubin, D. B. (1997), 'Estimating outcome distributions for compliers in instrumental variables models', The Review of Economic Studies **64**(4), 555–574.

Imbens, G. W. and Rubin, D. B. (2015), Causal inference in statistics, social, and biomedical sciences, Cambridge University Press.

Jin, Z., Ying, Z. and Wei, L.-J. (2001), 'A simple resampling method by perturbing the minimand', Biometrika **88**(2), 381–390.

Kalbfleisch, J. D. and Prentice, R. L. (2002), The statistical analysis of failure time data, Vol. 360 of Wiley Series in Probability and Statistics, John Wiley & Sons.

Khan, S., Ponomareva, M. and Tamer, E. (2016), 'Identification of panel data models with endogenous censoring', Journal of Econometrics **194**(1), 57–75.

Khan, S. and Tamer, E. (2009), 'Inference on endogenously censored regression models using conditional moment inequalities', Journal of Econometrics **152**(2), 104–119.

Kianian, B., Kim, J. I., Fine, J. P. and Peng, L. (2019), 'Causal proportional hazards estimation with a binary instrumental variable', arXiv preprint arXiv:1901.11050 .

Klein, J. P. and Moeschberger, M. L. (2006), Survival analysis: techniques for censored and truncated data, Springer Science & Business Media.

Koenker, R. (2005), Quantile regression, Vol. 38, Cambridge Univ Pr.

Koenker, R. and Bassett Jr, G. (1978), 'Regression quantiles', Econometrica: journal of the Econometric Society pp. 33–50.

Koller, M. T., Raatz, H., Steyerberg, E. W. and Wolbers, M. (2012), 'Competing risks and the clinical community: irrelevance or ignorance?', Statistics in medicine **31**(11-12), 1089–1097.

Lai, T. L. and Ying, Z. (1988), 'Stochastic integrals of empirical-type processes with applications to censored regression', Journal of multivariate analysis **27**(2), 334–358.

Lau, B., Cole, S. R. and Gange, S. J. (2009), 'Competing risk regression models for epidemiologic data', American journal of epidemiology **170**(2), 244–256.

Li, J., Fine, J. and Brookhart, A. (2015), 'Instrumental variable additive hazards models', Biometrics **71**(1), 122–130.

Li, R. and Peng, L. (2011), 'Quantile regression for left-truncated semicompeting risks data', Biometrics **67**(3), 701–710.

Li, S. and Gray, R. J. (2016), 'Estimating treatment effect in a proportional hazards model in randomized clinical trials with all-or-nothing compliance', Biometrics **72**(3), 742–750.

Lin, D., Sun, W. and Ying, Z. (1999), 'Nonparametric estimation of the gap time distribution for serial events with censored data', Biometrika **86**(1), 59–70.

Lin, D., Wei, L., Yang, I. and Ying, Z. (2000), 'Semiparametric regression for the mean and rate functions of recurrent events', Journal of the Royal Statistical Society: Series B (Statistical Methodology) **62**(4), 711–730.

Lin, D., Wei, L. and Ying, Z. (1998), 'Accelerated failure time models for counting processes', Biometrika **85**(3), 605–618.

Lin, D. Y., Wei, L.-J. and Ying, Z. (1993), 'Checking the cox model with cumulative sums of martingale-based residuals', Biometrika **80**(3), 557–572.

Lin, H., Li, Y., Jiang, L. and Li, G. (2014), 'A semiparametric linear transformation model to estimate causal effects for survival data', Canadian Journal of Statistics **42**(1), 18–35.

Liu, L., Wolfe, R. A. and Huang, X. (2004), 'Shared frailty models for recurrent events and a terminal event', Biometrics **60**(3), 747–756.

Loeys, T. and Goetghebeur, E. (2003), 'A causal proportional hazards estimator for the effect of treatment actually received in a randomized trial with all-or-nothing compliance', Biometrics **59**(1), 100–105.

Loomba, R. and Liang, T. J. (2017), 'Hepatitis b reactivation associated with immune suppressive and biological modifier therapies: current concepts, management strategies, and future directions', Gastroenterology **152**(6), 1297–1309.

Luo, X., Wang, M.-C. and Huang, C.-Y. (2010), 'A comparison of various rate functions of a recurrent event process in the presence of a terminal event', Statistical Methods in Medical Research **19**(2), 167–182.

Lyczak, J. B., Cannon, C. L. and Pier, G. B. (2002), 'Lung infections associated with cystic fibrosis', Clinical microbiology reviews **15**(2), 194–222.

Mack, Y.-p. and Silverman, B. W. (1982), 'Weak and strong uniform consistency of kernel regression estimates', Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete **61**(3), 405–415.

Martínez-Camblor, P., Mackenzie, T., Staiger, D. O., Goodney, P. P. and O'Malley, A. J. (2019), 'Adjusting for bias introduced by instrumental variable estimation in the cox proportional hazards model', Biostatistics **20**(1), 80–96.

Martinussen, T., Vansteelandt, S., Tchetgen, E. and Zucker, D. M. (2016), 'Instrumental variables estimation of exposure effects on a time-to-event response using structural cumulative survival models', arXiv preprint arXiv:1608.00818 .

Mathee, K., Ciofu, O., Sternberg, C., Lindum, P. W., Campbell, J. I., Jensen, P., Johnsen, A. H., Givskov, M., Ohman, D. E., Søren, M. et al. (1999), 'Mucoid conversion of pseudomonas aeruginos by hydrogen peroxide: a mechanism for virulence activation in the cystic fibrosis lung', Microbiology **145**(6), 1349–1357.

Moffett, K. S. (2010), 'Pseudomonas aeruginosa in patients with cystic fibrosis', Antimicrob. Ther **1**(1).

Neocleous, T., Branden, K. V. and Portnoy, S. (2006), 'Correction to censored regression quantiles by s. portnoy, 98 (2003), 1001–1012', Journal of the American Statistical Association **101**(474), 860–861.

Newey, W. K. (1994), 'Kernel estimation of partial means and a general variance estimator', Econometric Theory **10**(2), 1–21.

Nie, H., Cheng, J. and Small, D. S. (2011), 'Inference for the effect of treatment on survival probability in randomized trials with noncompliance and administrative censoring', Biometrics **67**(4), 1397–1405.

Oliveira, M., Reis, F., Monteiro, A. and Penna, F. (2002), 'Effect of meconium ileus on the clinical prognosis of patients with cystic fibrosis', Brazilian journal of medical and biological research **35**(1), 31–38.

Peng, L. (2012), 'Self-consistent estimation of censored quantile regression', Journal of Multivariate Analysis **105**(1), 368–379.

Peng, L. and Fine, J. P. (2007), 'Regression modeling of semicompeting risks data', Biometrics **63**(1), 96–108.

Peng, L. and Fine, J. P. (2009), 'Competing risks quantile regression', Journal of the American Statistical Association **104**(488), 1440–1453.

Peng, L. and Huang, Y. (2008), 'Survival analysis with quantile regression models', Journal of the American Statistical Association **103**(482), 637–649.

Peng, L., Jiang, H., Chappell, R. and Fine, J. (2008), 'An overview of the semi–competing risks problem', Statistical Advances in the Biomedical Sciences: Clinical Trials, Epidemiology, Survival Analysis, and Bioinformatics pp. 177–192.

Pepe, M. S. and Cai, J. (1993), 'Some graphical displays and marginal regression analyses for recurrent failure times and time dependent covariates', Journal of the American Statistical Association **88**(423), 811–820.

Pollard, D. (1990), Empirical processes: theory and applications, in 'NSF-CBMS regional conference series in probability and statistics', JSTOR, pp. i–86.

Portnoy, S. (2003), 'Censored regression quantiles', Journal of the American Statistical Association **98**(464), 1001–1012.

Portnoy, S. and Lin, G. (2010), 'Asymptotics for censored regression quantiles', Journal of Nonparametric Statistics **22**(1), 115–130.

Pouliot, G. A. (2019), 'Instrumental variables quantile regression with multivariate endogenous variable', Unpublished manuscript .

Powell, J. L. (1984), 'Least absolute deviations estimation for the censored regression model', Journal of Econometrics **25**(3), 303–325.

Prentice, R. L., Williams, B. J. and Peterson, A. V. (1981), 'On the regression analysis of multivariate failure time data', Biometrika **68**(2), 373–379.

Reid, N. (1994), 'A conversation with sir david cox', Statistical Science **9**(3), 439–455.

Robins, J. M. and Rotnitzky, A. (1992), Recovery of information and adjustment for dependent censoring using surrogate markers, in 'AIDS epidemiology', Springer, pp. 297–331.

Robins, J. M. and Tsiatis, A. A. (1991), 'Correcting for non-compliance in randomized trials using rank preserving structural failure time models', Communications in statistics-Theory and Methods **20**(8), 2609–2631.

Rosenfeld, M., Emerson, J., McNamara, S., Thompson, V., Ramsey, B. W., Morgan, W., Gibson, R. L., Group, E. S. et al. (2012), 'Risk factors for age at initial pseudomonas acquisition in the cystic fibrosis epic observational cohort', Journal of Cystic Fibrosis **11**(5), 446–453.

Russell, P., Hertz, P. and McMillan, B. (2012), Biology: The Dynamic Science, number v. 1 in 'Available Titles CourseMate Series', Cengage Learning.
**URL:** *https://books.google.com/books?id=UxkOTD5LNCwC*

Sawyer, S., Taylor, R., MacMahon, R. and Robertson, C. (1994), 'Meconium ileus in cystic fibrosis', Pediatric surgery international **9**(3), 180–184.

Schaubel, D. E. and Cai, J. (2005), 'Analysis of clustered recurrent event data with application to hospitalization rates among renal failure patients', Biostatistics **6**(3), 404–419.

Schaubel, D. E., Zeng, D. and Cai, J. (2006), 'A semiparametric additive rates model for recurrent event data', Lifetime Data Analysis **12**(4), 389–406.
**URL:** *http://dx.doi.org/10.1007/s10985-006-9017-x*

Schmittlein, D. C. and Morrison, D. G. (1981), 'The median residual lifetime: A characterization theorem and an application', Operations Research **29**(2), 392–399.

Sørensen, D. N., Martinussen, T. and Tchetgen, E. T. (2019), 'A causal proportional hazards estimator under homogeneous or heterogeneous selection in an iv setting', Lifetime data analysis **25**(4), 639–659.

Sovey, A. J. and Green, D. P. (2011), 'Instrumental variables estimation in political science: A readers' guide', American Journal of Political Science **55**(1), 188–200.

Spiekerman, C. F. and Lin, D. (1998), 'Marginal regression models for multivariate failure time data', Journal of the American Statistical Association **93**(443), 1164–1175.

Stefanski, L. A. and Carroll, R. J. (1987), 'Conditional scores and optimal scores for generalized linear measurement-error models', Biometrika **74**(4), 703–716.

Stock, J. H. and Yogo, M. (2002), Testing for weak instruments in linear iv regression, Technical report, National Bureau of Economic Research.

Stone, C. J. (1980), 'Optimal rates of convergence for nonparametric estimators', The annals of Statistics pp. 1348–1360.

Sun, X., Peng, L., Huang, Y. and Lai, H. J. (2016), 'Generalizing quantile regression for counting processes with applications to recurrent events', Journal of the American Statistical Association **111**(513), 145–156.

Tchetgen, E. J. T., Walter, S., Vansteelandt, S., Martinussen, T. and Glymour, M. (2015), 'Instrumental variable estimation in a survival context', Epidemiology (Cambridge, Mass.) **26**(3), 402.

Vaart, A. W. and Wellner, J. A. (1996), Weak convergence and empirical processes: with applications to statistics, Springer.

van der Vaart, A. W., Wellner, J. A. W. et al. (2007), Empirical processes indexed by

estimated functions, in 'Asymptotics: particles, processes and inverse problems', Institute of Mathematical Statistics, pp. 234–252.

Vansteelandt, S. and Goetghebeur, E. (2003), 'Causal inference with generalized structural mean models', Journal of the Royal Statistical Society: Series B (Statistical Methodology) **65**(4), 817–835.

Wang, M.-C. and Chang, S.-H. (1999), 'Nonparametric estimation of a recurrent survival function', Journal of the American Statistical Association **94**(445), 146–153.

Wang, M.-C., Qin, J. and Chiang, C.-T. (2001), 'Analyzing recurrent event data with informative censoring', Journal of the American Statistical Association **96**(455), 1057–1065.

Wei, L.-J., Lin, D. Y. and Weissfeld, L. (1989), 'Regression analysis of multivariate incomplete failure time data by modeling marginal distributions', Journal of the American statistical association **84**(408), 1065–1073.

Wooldridge, J. M. (2010), Econometric analysis of cross section and panel data, MIT press.

Yang, S. (1999), 'Censored median regression using weighted empirical survival and hazard functions', Journal of the American Statistical Association **94**(445), 137–145.

Ye, Y., Kalbfleisch, J. D. and Schaubel, D. E. (2007), 'Semiparametric analysis of correlated recurrent and terminal events', Biometrics **63**(1), 78–87.

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.

Yu, W., Chen, K., Sobel, M. E. and Ying, Z. (2015), 'Semiparametric transformation models for causal inference in time-to-event studies with all-or-nothing compliance', Journal of the Royal Statistical Society: Series B (Statistical Methodology) **77**(2), 397–415.

Zeng, D. and Cai, J. (2010), 'A semiparametric additive rate model for recurrent events with an informative terminal event', Biometrika **97**(3), 699–712.

Zeng, D. and Lin, D. (2007), 'Maximum likelihood estimation in semiparametric regression models with censored data', Journal of the Royal Statistical Society: Series B (Statistical Methodology) **69**(4), 507–564.

Zeng, D. and Lin, D. (2009), 'Semiparametric transformation models with random effects for joint analysis of recurrent and terminal events', Biometrics **65**(3), 746–752.

Zheng, C., Dai, R., Hari, P. N. and Zhang, M.-J. (2017), 'Instrumental variable with competing risk model', Statistics in medicine **36**(8), 1240–1255.