Semiparametric Inference of Censored Data with Time-dependent Covariates

Chi Wing Chu

Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy under the Executive Committee of the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2021
Abstract
Semiparametric Inference of Censored Data with Time-dependent Covariates
Chi Wing Chu

This thesis develops two semiparametric methods for censored survival data when the covariates involved are time-dependent. Respectively in the two parts of this thesis, we introduce an interquantile regression model and a censored quantile regression model that account for the commonly observed time-dependent covariates in survival analysis. The proposed quantile-based techniques offer a greater model flexibility comparing to the Cox proportional hazards model and the accelerated failure time model.

The first half of this thesis introduces a censored interquantile regression model with time-dependent covariates. Conventionally, censored quantile regression stipulates a specific, pointwise conditional quantile of the survival time given covariates. Despite its model flexibility and straightforward interpretation, the pointwise formulation oftentimes yields rather unstable estimates across neighbouring quantile levels with large variances. In view of this phenomenon, we propose a new class of censored interquantile regression models with time-dependent covariates that can capture the relationship between the failure time and the covariate processes of a target population that falls within a specific quantile bracket. The pooling of information within a homogeneous neighbourhood facilitates more efficient estimates hence more consistent conclusion on statistical significances of the variables concerned. This new formulation can also be regarded as a generalization of the accelerated failure time model for survival data in the sense that it relaxes the assumption of global homogeneity for the error at all quantile levels. By introducing a class of weighted rank-based estimation procedure, our framework allows a quantile-based inference on the covariate effect with a less restrictive set of assumptions. Numerical studies demonstrate that the proposed estimator outperforms existing alternatives under various settings in terms of smaller empirical bias and standard deviation. A perturbation-based resampling method is also developed to reconcile the asymptotic distribution of the parameter estimates. Finally, consistency and weak convergence of the proposed estimator are established via empirical process theory.
In the second half of this thesis, we propose a class of censored quantile regression models for right censored failure time data with time-dependent covariates that only requires a standard conditionally independent censorship. Upon a quantile based transformation, a system of functional estimating equations for the quantile parameters is derived based on the martingale construction. While time-dependent covariates naturally arise in time to event analysis, the few existing literature requires either an independent censoring mechanism or a fully observed covariate process even after the event has occured. The proposed formulation extends the existing censored quantile regression model so that only the covariate history up to the observed event time is required as in the Cox proportional hazards model for time-dependent covariates. A recursive algorithm is developed to evaluate the estimator numerically. Asymptotic properties including uniform consistency and weak convergence of the proposed estimator as a process of the quantile level is established. Monte Carlo simulations and numerical studies on the clinical trial data of the AIDS Clinical Trials Group is presented to illustrate the numerical performance of the proposed estimator.
# Table of Contents

List of Tables ................................................................. iii

List of Figures ............................................................... iv

Acknowledgments ............................................................. v

Chapter 1: Introduction and Background ............................... 1
  1.1 Accelerated Failure Time Model .................................... 1
  1.2 Censored Quantile Regression Model ................................. 2
  1.3 External and Internal Time-dependent Covariates .................... 4

Chapter 2: Censored Interquantile Regression with Time-dependent Covariates ...... 8
  2.1 Introduction ................................................................. 8
  2.2 Methodology ............................................................... 11
  2.3 Asymptotic Results ....................................................... 15
  2.4 Inference ................................................................. 16
    2.4.1 Resampling Scheme .................................................. 16
    2.4.2 Two Stage Testing Procedure for Homogeneity .................. 17
  2.5 Simulations ............................................................. 19
  2.6 Data Analysis ........................................................... 28
List of Tables

2.1 Simulation Results with Homogeneous Error ........................................... 21
2.2 Simulation Results with Homogeneity assumed for $\tau \leq 0.25$ ................. 22
2.3 Simulation Results with Homogeneity assumed for $0.25 \leq \tau \leq 0.75$ ....... 23
2.4 Empirical Rejection Rate of the Test for Homogeneity ............................. 27
2.5 Simulation Results with Homogeneous Error and Independent Censoring .... 27
2.6 Parameter Estimates of the Stanford Heart Transplant Data ...................... 31
3.1 Simulation Results: Homogeneous Setup ................................................. 53
3.2 Simulation Results: Heterogeneous Setup ............................................... 54
3.3 Mean Goodness of Fit Statistics for the Monte Carlo Simulation ($\times 10^{-4}$) .... 59
3.4 Estimation Results for ACTG 320 Dataset .............................................. 60
List of Figures

3.1  Estimates and Pointwise Confidence Intervals for the ACTG 320 Dataset . . . . . . .  62
3.2  Estimated Log Quantile Functions for the Patient with Identifier 23424. . . . . . . .  63
Acknowledgements

I would like to take this opportunity to express my sincere gratitude to my advisor, Professor Zhiliang Ying, for his invaluable guidance and generous support throughout my research study. This dissertation would be nowhere without his enlightening ideas and kind encouragement. It is my great honor and good fortune to have him as my advisor.

I would like to extend my appreciation to Professor Sumit Mukherjee, Professor Marco Avella, Professor Hammou El Barmi and Professor Liming Peng for serving on my dissertation committee and sharing their insightful comments. I am also deeply indebted to Professor Tony Sit for his precious advice and constant support throughout my entire study path.

Studying at Columbia has been a worthwhile experience that benefits me a lot. I would like to thank the faculty members, classmates, and staff in the Department of Statistics for creating a comfortable and promising atmosphere. I am particularly grateful to Dood Kalicharan and Anthony Cruz for their generous help.

Lastly, I would like to thank my family for their unconditional giving and support. I am also grateful to my good friends for their continual encouragement. I am most thankful to my beloved, Mancy, for her faithful love and patience in my pursuits of graduate studies.
Chapter 1: Introduction and Background

Survival analysis has been an important area in statistics with wide applications in health sciences, econometrical analysis and industrial life testing. This chapter introduces some statistical techniques for censored data when the covariates are time-independent. In particular, we focus on the semiparametric regression models, where the failure time and the covariates are related by some straightforward functional forms, while model flexibility can be maintained with no distributional assumption on the error or baseline. As an example, the Cox proportional hazards model ([10],[11]) is probably the most renowned semiparametric survival model because of its straightforward implementation. In this chapter, instead of modeling the hazard function of the failure time, we discuss in details the accelerated failure time model and the censored quantile regression model, which model the failure time itself and its quantile respectively. Later, we discuss the two types of time-dependent covariates and the relevant extensions and difficulties under such scenarios.

1.1 Accelerated Failure Time Model

As an alternative to the Cox proportional hazards model, which assumes a multiplicative covariate effect on the hazard function, the accelerated failure time model stipulates a direct relationship between the failure time $T$ and the $p$-dimensional covariate $Z$, thus offers a more straightforward model interpretation. More precisely, consider the log-linear model

$$\log(T) = Z^T \beta + \log(\varepsilon), \quad (1.1)$$

or equivalently, $T \exp(-Z^T \beta) = \varepsilon$, where $\beta$ is the $p$-dimensional regression parameter and the error term $\varepsilon$ represents the baseline failure time when the covariate $Z$ is zero. We assume that $\varepsilon$ is
independent of $Z$ but leave the distribution of $\epsilon$ unspecified, thus model (1.1) is semiparametric in nature. If instead a parametric family is imposed on $\epsilon$, then (1.1) is simply a log-linear regression model, hence the least square method can be employed by replacing the censored observations with the conditional tail expectation. Note that the covariate effect in (1.1) is to accelerate or decelerate the time to event, hence the model name. Some important references for the semiparametric accelerated failure time model include [6], [46], [52], [54], [33], [59], amongst others.

Suppose we observe $n$ i.i.d. copies of $f_{\tilde{T} - \Delta - Z}$, denoted by $f_{\tilde{T}_i - \Delta_i - Z_i}$ for $i = 1, \ldots, n$, where $\tilde{T} = T \land C$ is the censored failure time and $\Delta = I(T \leq C)$ is the censoring indicator. Here, $C$ is the censoring time which is conditionally independent of $T$ given $Z$, and $I(\cdot)$ denotes the indicator function. A popular approach to estimate $\beta$ is the rank-based procedure, with the estimating equation given by

$$\frac{1}{n} \sum_{i=1}^{n} \Psi(e_i(\beta); \beta) \left\{ \begin{array}{l} Z_i - \frac{S^{(1)}(e_i(\beta); \beta)}{S^{(0)}(e_i(\beta); \beta)} \end{array} \right\} = 0, \quad (1.2)$$

where $e_i(\beta) = T_i \exp(-Z_i^\top \beta)$ is the residual term, $S^{(0)}(t; \beta) = n^{-1} \sum_{i=1}^{n} I(e_i(\beta) \geq t)$, $S^{(1)}(t; \beta) = n^{-1} \sum_{i=1}^{n} Z_i I(e_i(\beta) \geq t)$, and $\Psi(t, \beta)$ is a weight function. The cases of $\Psi(t, \beta) = 1$ and $\Psi(t, \beta) = S^{(0)}(t; \beta)$ are referred to as the log-rank and Gehan weight function respectively. However, $U_n(\beta)$ is discontinuous of $\beta$ and may not be monotone in $\beta$. It is monotone when the Gehan weight function is considered [37]. In practice, the estimator of $\beta_0$, denoted by $\hat{\beta}$, can be obtained by minimizing the $L_1$-norm of $U_n(\beta)$. Under some regularity conditions, $n^{1/2}(\hat{\beta} - \beta)$ is shown to be asymptotically normal ([52, 59]), where the covariance matrix depends on the unspecified distribution of $\epsilon$. For the inference procedure, a resampling technique is developed to estimate the unknown covariance, see [37] and [23], for example.

1.2 Censored Quantile Regression Model

The quantile regression model introduced by [29] is another widely used technique for survival data. The model assumes the covariate effect is associated with the quantile of the failure time, hence allowing a quantile-based inference on the regression model. Let $Q_T(\tau \mid Z)$ denote the $\tau$-th
conditional quantile of $T$ given $Z$, i.e. $Q_T(\tau \mid Z) = \inf\{t : P(T \leq t \mid Z) \geq \tau\}$. Note that any monotone transform is order preserving, therefore the quantile function is invariant to monotone transform, i.e. we have $Q_{h(T)}(\tau \mid Z) = h(Q_T(\tau \mid Z))$ for monotone $h(\cdot)$. Now, for some quantile level $\tau \in (0, 1)$, we consider

$$Q_{\log T}(\tau \mid Z) = Z^{T} \beta(\tau), \tag{1.3}$$

or equivalently, $Q_T(\tau \mid Z) = \exp(Z^{T} \beta(\tau))$, because of the equivariance property. In other words, (1.3) assumes a log-linear model on the conditional quantile of $T$ given $Z$, instead of the failure time $T$ itself as in the accelerated failure time model. Indeed, the censored quantile regression model can be viewed as a generalization of the accelerated failure time model by allowing a quantile specific covariate effect. More precisely, suppose (1.3) holds for all $\tau \in (0, 1)$, $Z = (1, \tilde{Z}^{T})^{T}$ and $\beta(\tau) = (\log q(\tau), \tilde{\beta}^{T})^{T}$ for some $q(\tau) \in (0, \infty)$, i.e. $\beta(\tau)$ is a constant function of $\tau$ except the intercept term. As a result, for $\tau \in (0, 1)$ we have $Q_{\log T}(\tau \mid Z) = Z^{T} \beta + \log q(\tau)$, or equivalently, $Q_{T \exp(-Z^{T} \beta)}(\tau \mid Z) = q(\tau)$. It follows that $T \exp(-Z^{T} \beta)$ has a homogeneous quantile function which is free of $Z$, i.e. we have $T \exp(-Z^{T} \beta) = \varepsilon$ for some $\varepsilon$ with its distribution unspecified. In other words, the semiparametric accelerated failure time model is a special case of the censored quantile regression model when the regression parameters of the covariates are constant. Some important references for the censored quantile regression model include [58], [43], [41], [53], [21], amongst others.

In the absence of censoring, the regression parameter $\beta(\tau)$ in model (1.3) can be estimated by minimizing the check function, i.e. analogy of the absolute error loss for general quantile level. When censoring is involved, an important development is due to [41], which observe the correspondence between the quantile process and the time scale. More precisely, assuming (1.3) holds over $\tau \in (0, \tau_U] \subset (0, 1)$, they use the martingale structure in the counting process of censored data to construct estimating equation, and transform the hazard-based martingale to the quantile domain via a change of variable. In particular, they consider the following estimating
equation

\[
\frac{1}{n} \sum_{i=1}^{n} \left[ Z_i \left\{ \Delta_i (\log \tilde{T}_i \leq Z_i^\top \beta) - \int_0^\tau I(\log \tilde{T}_i \geq Z_i^\top \beta(u)) dH(u) \right\} \right] = 0, \tag{1.4}
\]

where \( H(x) = -\log(1-x) \) and \( \{ \tilde{T}_i, \Delta_i, Z_i \} \) for \( i = 1, \ldots, n \) are i.i.d. copies of \( \{ \tilde{T}, \Delta, Z \} \) as defined. They propose to obtain the estimate of \( \beta(\tau) \) for \( \tau \in (0, \tau_U) \) recursively over a pre-specified grid of quantile levels, with the starting point set to be zero at the zero quantile. Under some regularity conditions, it is shown that \( \hat{\beta}(\tau) \) is uniformly consistent for \( \beta(\tau) \) and \( n^{1/2} \{ \hat{\beta}(\tau) - \beta(\tau) \} \) converges weakly to a mean-zero Gaussian process. The limiting covariance process is approximated by a perturbation-based resampling technique in the spirit of [23].

1.3 External and Internal Time-dependent Covariates

As survival event occurs over time, time-dependent covariates could arise naturally in several causes. The first examples are functions of time, say the interaction term between a fixed covariate and time. Also, measurements may be repeated made during the course of a study because of its relation to survival. As an example, the blood pressure of patients could be measured at some regular intervals, and it is natural to believe the blood pressure of an individual at a specified time point rather than the initial value at entry is more correlated to the current failure rate. Meanwhile, time-dependent covariates can also be used to model the effect of treatment switching. More precisely, a process of indicator function over time can be defined to track the individual history in the treatment group or in the control group. Finally, if the time-dependent covariate depend on its own history of failures, censoring and other features, it is called evolutionary. Examples include the number of failures and the size of the risk set over time. Further details and examples could be found in Chapter 8 of [9] and Chapter 6 of [26].

Time-dependent covariates fall into two classes, namely external and internal covariates. Let \( Z(t) \) denote the value of the time-dependent covariate process at time \( t \), and \( \tilde{Z}(t) = \{ Z(s), s \in \tau \} \).
be the covariate history up to time $t$. If a time-dependent covariate satisfies

$$P\{T \in [u, u + du] \mid T \geq u, \tilde{Z}(u)\} = P\{T \in [u, u + du] \mid T \geq u, \tilde{Z}(t)\}$$

(1.5)

for all $u, t$ such that $u \in [0, t]$, then it is an external covariate. It is an internal covariate otherwise. Note that the left hand side of (1.5) is the hazard rate conditional on the present covariate history, while the right hand side is conditioning on the covariate process up to some future time point. This indicates that the path of an external covariate does not have direct implications on the failure status of the subject, hence its future path does not alter the present hazard rate. Alternatively, (1.5) is equivalent to

$$P\{\tilde{Z}(t) \mid \tilde{Z}(u), T \geq u\} = P\{\tilde{Z}(t) \mid \tilde{Z}(u), T = u\}$$

(1.6)

for all $u, t$ such that $u \in [0, t]$, hence the future path of an external covariate is not affected by the failure at present. It is easy to see from definition that time-independent covariates are external. More generally, if the entire path of a covariate is predetermined for each individual in advance, such as stress applied in industrial life testing, then it is external. Another type of external covariates is called ancillary, where the path of which is based on external factors but not the individual. An example of this kind would be the air pollution level as a covariate for asthma attacks. In contrast, an internal covariate are usually those repeated measurements taken on the subject, such as blood pressure of the patients, hence its existence necessarily implies the survival of the individual. If our study defines failure to be death of the individual, the path of such internal covariates would have direct information on the failure status. For example, a non-zero measurement of the blood pressure at time $t$ indicates that the patient is alive for all $u \leq t$.

As the hazard function with time-dependent covariates can be naturally defined by the left hand side of (1.5), the Cox proportional hazards model can be readily generalized to accommodate time-dependent covariates. However, if instead the quantile function is concerned, the extension to internal covariates is not as straightforward as external covariates. As pointed out in [15] and [26], one needs to differentiate between external and internal covariates because of the subtle difference.
in the interpretation. For external covariates, we can define the conditional hazard rate as time $t$ as

\[ \lambda_T(t \mid \bar{Z}(t)) \, dt = P\{T \in [t, t + dt] \mid T \geq t, \bar{Z}(t)\}, \]

and the distribution function of $T$ given the entire covariate path $\bar{Z} = \bar{Z}(\infty)$ as $F_T(t \mid \bar{Z}) = P(T \leq t \mid \bar{Z})$. Because of condition (1.5), we have

\[ F_T(t \mid \bar{Z}) = 1 - \exp \left(-\int_0^t P\{T \in [u, u + du] \mid T \geq u, \bar{Z}\} \, du\right) = 1 - \exp \left(-\int_0^t \lambda_T(u \mid \bar{Z}(u)) \, du\right), \]

i.e. we resemble the usual relationship between the distribution function and the hazard function.

On the other hand, such observation is no longer true for internal covariates because condition (1.5) does not hold. Indeed, the above definition of $F_T(t \mid \bar{Z})$ would become degenerated for internal covariates as the mere existence of the covariate value implies the individual has not died, i.e. we have $P(T \leq t \mid Z(t) \neq 0) = 0$. It is therefore problematic to model such distribution function or the corresponding quantile function directly.

There have been some attempts in the literature that extend the above semiparametric techniques under the scenario of time-dependent covariates. However, as the usual relationship between the ordinary distribution function and the hazard function breaks down when internal covariates are involved, caution has to be taken in model construction and interpretation. In particular, [47] and [38] extended the accelerated failure time model with a rank-based procedure to accommodate external time-dependent covariates. For censored quantile regression, [18] considered an inverse probability weighted estimating equation under the more restrictive independent censorship assumption. Nevertheless, these proposals have no discussion on the treatment of internal covariates in their models because of the subtlety involved.

In the sequel, we discuss two semiparametric regression models for censored data with time-dependent covariates. In Chapter 2, we present a hybrid model between the accelerated failure time model and the censored quantile regression model that examines the covariate effect on the failure time over a designated range of quantiles. A rank-based estimation procedure analogous to the (1.2) is derived in a similar fashion. In Chapter 3, we propose a censored quantile regression model that generalizes [41] to the setup of time-dependent covariates. A function estimating equation is
derived based on martingale theory in the spirit of (1.4). We provide a different perspective in the model construction so that internal covariates can be incorporated.
Chapter 2: Censored Interquantile Regression with Time-dependent Covariates

2.1 Introduction

Censored quantile regression has been a popular alternative to Cox proportional hazards model ([10, 11]) in survival analysis. This broad class of models allows a greater degree of flexibility as it can capture the heterogeneity of covariate effects without assuming that the relative risks are constant with time. It also provides better interpretability as the model directly relates a specific, pointwise quantile of the failure time with a set of candidate covariates. There has been rich literature discussing quantile regression models on censored survival data. In particular, [44, 45] studied the least absolute deviation estimation for the censored regression model. Under the independent censoring assumption, [58] derived the median regression procedure for survival data. With a less restrictive assumption on conditionally independent censoring, [43]'s and [41]'s proposals provided novel adaptations for Kaplan-Meier (KM) and Nelson-Aalen (NA) estimators, respectively, upon which corresponding inference procedures for censored quantiles can be implemented. To avoid recursive algorithms required in the aforementioned procedures, [53] insightfully developed a locally weighted procedure that adopts the redistribution-of-mass idea for a kernel-based, local reweighting scheme. All these approaches were, however, designed for time-independent covariates.

In many real applications, since survival events occur over time, there is a natural need to handle time-dependent covariates such as heart blood pressures or the drug dosages received by individual patients. Statistical theory of the Cox proportional hazards model can be readily extended to handle time-dependent covariates, see, for example, [3], [26] and [39] amongst others. The classic accelerated failure time model has also been extended in [38], [60] and [51], amongst others to
accommodate time-dependent covariates. For censored quantile regression with time-dependent covariates, there has only been few results to the best of our knowledge. One interesting work is [18], whose estimating equation was developed upon an inverse probability weighting scheme under the independent censoring assumption.

As pointed out in [41], censored quantile regression models can be reduced to the accelerated failure time model ([47, 38]) if the regression parameters are constant for all quantile values $\tau \in (0, 1)$. In other words, when the homogeneity assumption holds for all $\tau$’s, the two models are equivalent. This observation motivates us to explore the uncharted region between the two ends of the spectrum. Indeed, our new proposal attempts to seek a reasonable compromise between model flexibility and estimation efficiency. Because of the homogeneity assumption, the accelerated failure time model can pool the information provided by all the observations for estimating a reduced set of model parameters. As we can see in Example 4 in Section 2.5, if one can delicately leverage the local homogeneity shared amongst the specific quantile level concerned and its neighbouring levels, the problem of obtaining volatile point estimates for quantile regression parameters, which is frequently observed in many existing studies, can be alleviated.

Recall that the censored quantile regression model studies the influence of the covariates on the failure time at designated quantile levels, such as the median and the quartiles, while the accelerated failure time model examines the overall covariate effect provided the aforementioned homogeneity assumption holds for all quantile levels. None of these model, however, can quantify the covariate effects over a specific group in the sample, say the top 25% or the bottom 25% of a population in term of length of survival after covariate adjustment. For instance, to measure the effectiveness of a new drug, investigators are interested in quantifying its effect on the survival times of individuals who are among the shortest lived 25% in the population given the covariates instead of its effect on an individual who represents the lower quartile of the failure time distribution. In such scenarios, a localized censored quantile regression model with parameters representing the treatment effect over the designated range of quantiles should appear more attractive, whereas the classic quantile regression model only quantifies treatment effect on particular levels of quantiles.
The main contribution of this work is the development of a quantile-based version of the accelerated failure time model for censored data with time-dependent covariates under which we relax the homogeneity requirement under the conditional independent censoring assumption. Moreover, our approach also provides an alternative perspective to the study of quantile regression problems. The pooling effect enjoyed by our model can remediate the drawback that many existing proposals for censored quantile regression may share. Besides numerical stability, our model also offers an alternative perspective for researchers so as to best match their primary goals of quantifying the covariate effects on the failure time for an interval of quantiles rather than a specific level. It should be emphasized that although there are similar works in literature that discuss methods of combining neighbouring quantile coefficient estimates to yield more stable estimates, to the best of our knowledge, none can be straightforwardly extended to our specific problem. In [61], for example, the aggregation is carried out via the penalty term based on an overall sum of the magnitudes of the neighbouring coefficients concerned, but the parameter estimates remain capturing the covariates effects on the response variable at specific quantile levels. Same as [61], [62] and [25] are also designed for uncensored observations in which case the celebrated check function can be straight-forwardly applied to provide initial estimates upon which smoothing or other aggregation is introduced to achieve the inference goals.

The rest of this paper is organized as follows: Section 2.2 outlines the model framework and establishes the corresponding estimation equation for the model parameters. Section 2.3 outlines the asymptotic results of the proposed estimator, while Section 2.4 provides a perturbation based resampling method and a two stage testing procedure for inference. A set of simulation results is presented in Section 2.5, followed by a data analysis of the Stanford heart transplant data in Section 2.6. A short concluding discussion is given in Section 2.7. All proofs of the asymptotic results and justification of the resampling method are presented in the appendix, i.e. Section 2.8.
2.2 Methodology

For \( i = 1, \ldots, n \), let \( T_i \) denote the actual failure time of the \( i \)-th individual, \( \mathbf{Z}_i = \{ Z_i(s), s \geq 0 \} = \{(Z_{i1}(s), \ldots, Z_{ip}(s))^\top, s \geq 0 \} \) be the \( p \)-dimensional multivariate random covariate process corresponding to this person, and \( \boldsymbol{\beta}_0 \) be the \( p \)-dimensional vector of coefficients. Let \( \varepsilon_i = h_i(T_i; \boldsymbol{\beta}_0) \) be a monotone transform of \( T_i \), where

\[
\varepsilon_i = \int_0^{T_i} \exp\{ -Z_i(s)^\top \boldsymbol{\beta} \} \, ds,
\]

in which \( \varepsilon_i \) represents the baseline failure time of the \( i \)-th individual when all of its covariates equal zero. Our model assumes that for all \( u \in [\tau_L, \tau_U] \subset [0, 1] \), the \( u \)-th quantile of \( \varepsilon_i \) conditional on the covariate process \( \mathbf{Z}_i \) is homogeneous for all \( i = 1, \ldots, n \). Mathematically, it stipulates that

\[
Q_{h_i(T_i; \boldsymbol{\beta}_0)}(u \mid \mathbf{Z}_i) = q(u), \quad (u \in [\tau_L, \tau_U] ; \ i = 1, \ldots, n), \tag{2.1}
\]

where \( q(u) \) is a function independent of \( \mathbf{Z}_i \) representing the common conditional quantile, and \( Q_{h_i(T_i; \boldsymbol{\beta}_0)}(\cdot \mid \mathbf{Z}_i) \) denotes the conditional quantile function of \( h_i(T_i; \boldsymbol{\beta}_0) \) given \( \mathbf{Z}_i \). Our goal is to estimate \( \boldsymbol{\beta}_0 \) under model (2.1).

Indeed, model (2.1) can be viewed as a generalization of the accelerated failure time model with time-dependent covariates proposed by [38]. Recall their model assumes that \( h_i(T_i; \boldsymbol{\beta}_0) = \varepsilon_i \) possesses a common baseline failure time distribution conditional on \( \mathbf{Z}_i \), which is equivalent to say \( Q_{h_i(T_i; \boldsymbol{\beta}_0)}(u \mid \mathbf{Z}_i) = q(u) \) for all \( u \in [0, 1] \), i.e. the form of (2.1) with \( \tau_L = 0 \) and \( \tau_U = 1 \). Our new formulation relaxes such requirement so that only a sub-interval of quantiles satisfies (2.1). An important special case is when \( \mathbf{Z}_i(t) \equiv \mathbf{Z}_i \) is time-independent, then model (2.1) says that \( h_i(T_i; \boldsymbol{\beta}_0) = T_i \exp\{ -Z_i^\top \boldsymbol{\beta}_0 \} \), or equivalently, \( \log T_i - Z_i^\top \boldsymbol{\beta}_0 \), has a homogeneous conditional quantile function given \( \mathbf{Z}_i \) over \( [\tau_L, \tau_U] \). The model parameter can be interpreted as the covariate effect over the designated quantiles of the failure time distribution.

Suppose our data are i.i.d. copies of \( \{ \tilde{T}_i, \Delta_i, \mathbf{Z}_i(\tilde{T}_i) \} \) for \( i = 1, \ldots, n \), where \( \tilde{T}_i = T_i \land C_i \) is the
observed event time, \( \Delta_i = I(T_i \leq C_i) \) is the censoring indicator, and \( \tilde{Z}_i(\tilde{T}_i) = \{ Z_i(s), 0 \leq s \leq \tilde{T}_i \} \) is the covariate history up to the survival time. Here, \( C_i \) is the censoring time which is conditionally independent of \( T_i \), given \( \tilde{Z}_i \), and \( I(\cdot) \) denotes the indicator function. Define the conditional cumulative hazard function, counting process and at-risk process associated with \( h_i(T_i; \beta_0) \) to be \( \Lambda_i(t; \beta_0) \), \( N_i(t; \beta_0) = \Delta_i I(h_i(\tilde{T}_i; \beta_0) \leq t) \), \( Y_i(t; \beta_0) = I(h_i(\tilde{T}_i; \beta_0) \geq t) \), respectively. Note that model (2.1) implies that \( h_i(T_i; \beta_0) \) has a homogeneous conditional distribution function for \( t \in [q(\tau_L), q(\tau_U)] \), so we can write \( \Lambda_i(t; \beta_0) = \Lambda(t; \beta_0) \) for such \( t \). Now, based on the classical counting process theory [16], we have
\[
E \left[ dN_i(t; \beta_0) - Y_i(t; \beta_0) d\Lambda(t; \beta_0) \mid \tilde{Z}_i \right] = 0, \tag{2.2}
\]
and also
\[
E \left[ Z_i(h_i^{-1}(t; \beta_0)) \left\{ dN_i(t; \beta_0) - Y_i(t; \beta_0) d\Lambda(t; \beta_0) \right\} \mid \tilde{Z}_i \right] = 0, \tag{2.3}
\]
for \( t \in [q(\tau_L), q(\tau_U)] \), where \( h_i^{-1}(\cdot; \beta) \) is the inverse function of \( h_i(\cdot; \beta) \). By solving the empirical counterpart of (2.2), we obtain a Nelson-Aalen type estimator of \( d\Lambda(t; \beta_0) \), which is \( \sum_{i=1}^n dN_i(t; \beta_0) / \sum_{i=1}^n Y_i(t; \beta_0) \) for \( t \in [q(\tau_L), q(\tau_U)] \). Meanwhile, with a slight abuse of notation, we denote the \( \tau \)-th unconditional quantile of \( h_i(T_i; \beta) \) by \( q(\tau; \beta) \), i.e. \( q(\tau; \beta) \) is the inverse function of \( F(t; \beta) \), where \( F(t; \beta) \) is the unconditional, or marginal, distribution function of \( h_i(T_i; \beta) \). Model (2.1) implies the homogeneous conditional quantile \( q(\tau) \) coincides with \( q(\tau; \beta_0) \) for \( \tau \in [\tau_L, \tau_U] \), so for \( \tau = \tau_L \) and \( \tau = \tau_U \), \( q(\tau) \) can be estimated by \( \hat{q}(\tau; \beta_0) \), the inverse function of \( \hat{F}(t; \beta_0) \), where \( \hat{F}(t; \beta) \) is an estimator of \( F(t; \beta) \). In particular, we take \( \hat{F}(t; \beta) \) to be the Kaplan-Meier estimator based on \( \{ h_i(\tilde{T}_i; \beta_0), \Delta_i \} \). As mentioned, this estimation works because the conditional quantile and the unconditional quantile are equivalent for \( \tau = [\tau_L, \tau_U] \) when \( \beta = \beta_0 \).

Incorporate the two estimators of \( d\Lambda(t; \beta_0) \) and \( q(\tau) \) into the empirical counterpart of (2.3) and integrate over the range of \( t \), it is natural to consider the weighted estimating equation
\[
U_n(\beta; \tau_L, \tau_U) = \]
0, where

\[
U_n(\beta; \tau_L, \tau_U) = \frac{1}{n} \sum_{i=1}^{n} \int_{\hat{q}(\tau_L; \beta)}^{\hat{q}(\tau_U; \beta)} \Psi(t, \beta) \left( \mathbf{Z}_i(h_i^{-1}(t; \beta)) - \frac{S(1)(t; \beta)}{S(0)(t; \beta)} \right) dN_i(t; \beta).
\]

Here, \(S(0)(t; \beta) = n^{-1} \sum_{i=1}^{n} Y_i(t; \beta), S(1)(t; \beta) = n^{-1} \sum_{i=1}^{n} \mathbf{Z}_i(h_i^{-1}(t; \beta)) Y_i(t; \beta),\) and \(\Psi(t, \beta)\) is a weight function that possibly depends on the data \(\{\tilde{T}_i, \Delta_i, \tilde{Z}_i(\tilde{T}_i), i = 1, \ldots, n\}.\) Note that these data-driven functionals are dependent on \(n\) but we suppress the subscript in the notations for brevity. Our proposed estimator of \(\beta_0,\) denoted by \(\hat{\beta},\) is obtained as the root of \(U_n(\beta; \tau_L, \tau_U) = 0.\) To this end, we simplify \(U_n(\beta; \tau_L, \tau_U)\) into a more familiar form by recalling that \(N_i(t; \beta) = \Delta_i I(h_i(\tilde{T}_i; \beta) \leq t).\) The above display can thus be rewritten as

\[
U_n(\beta; \tau_L, \tau_U) = \frac{1}{n} \sum_{i=1}^{n} \Psi(h_1(\tilde{T}_i; \beta); \beta) \left( \mathbf{Z}_i(\tilde{T}_i) - \frac{S(1)(h_1(\tilde{T}_i; \beta); \beta)}{S(0)(h_1(\tilde{T}_i; \beta); \beta)} \right) \times \left\{N_i(\hat{q}(\tau_U; \beta); \beta) - N_i(\hat{q}(\tau_L; \beta); \beta)\right\}. \tag{2.4}
\]

Here, \(\hat{q}(\tau_L; \beta)\) reduces to 0 if \(\tau_L = 0\) and \(\hat{q}(\tau_U; \beta)\) reduces to \(\infty\) if \(\tau_U = 1.\) Observe that if \(h_1(\tilde{T}_i; \beta) > h_j(\tilde{T}_j; \beta),\) then \(Y_j(h_1(\tilde{T}_i; \beta); \beta) = 0;\) otherwise if \(h_1(\tilde{T}_i; \beta) \leq h_j(\tilde{T}_j; \beta),\) then we have \(h_j^{-1}(h_1(\tilde{T}_i; \beta); \beta) \leq \tilde{T}_j.\) Therefore, the quantity \(S(1)(h_1(\tilde{T}_i; \beta); \beta)\) is observable for any \(\beta\) given the data, regardless of the censoring status of the subject. As a result, despite we condition on the entire covariate process \(\tilde{Z}_i\) in the derivation, the estimating equation (2.4) remains computable provided that only \(\tilde{Z}_i(\tilde{T}_i)\) is known, i.e. the covariate value up to the observed survival time instead of the entire process. We emphasize that, as in the accelerated failure time model and the quantile regression model for time-dependent covariates such as [38] and [18], although internal covariates could possibly be fit into (2.4), one has to be cautious in interpreting the model parameters as opposed to external covariates; see Chapter 6 of [26].

Observe that \(U_n(\beta; \tau_L, \tau_U)\) is neither continuous nor monotone in general, we solve its root by minimizing the \(L_1\) norm. The choice of \([\tau_L, \tau_U]\) in (2.4) depends on one’s primarily interested range of quantiles. As an example, if one is interested in the covariate effects on the bottom 25%
of population after covariate adjustment, then \([\tau_L, \tau_U]\) should be chosen to be \([0, 0.25]\). More discussion on the appropriate options for the range can be found in Section 2.4.

To shed light on the connection between our formulation with existing estimators, one could view \(\Psi(h_i(\bar{T}_i; \beta); \beta)\{N_i(\hat{q}(\tau_U; \beta); \beta) - N_i(\hat{q}(\tau_L; \beta); \beta)\}\) as a single subject specific weight, then \(U_n(\beta; \tau_L, \tau_U)\) falls into the class of weighted log rank statistics for time-dependent covariate setting, with \(\Psi(h_i(\bar{T}_i; \beta); \beta)\{N_i(\hat{q}(\tau_U; \beta); \beta) - N_i(\hat{q}(\tau_L; \beta); \beta)\}\) as the modified weight. Indeed, when \(\Psi(h_i(\bar{T}_i; \beta); \beta) = 1, \tau_L = 0\) and \(\tau_U = 1\), the expression reduces to \(\Delta_i\) and thus (2.4) recovers the estimating equation of [38].

Typical renowned weight functions include \(\Psi(t; \beta) = 1\) and \(\Psi(t; \beta) = S^{(0)}(t; \beta)\), which we refer to as Gehan weight. Due to the extra complication introduced by the quantile specific inference and the time-dependent covariates, the two choices do not allow for an easier numerical implementation as in [23]. Based on our numerical experience as shown in Section 2.5, the performance using the two different weight functions does not differ considerably. Unless otherwise specified, we shall assume the estimating equation is unweighted when we adopt (2.4) in the numerical studies.

Before concluding this section, we stress here that despite model (2.1) resembles the formulation considered in [18], they are fundamentally different in the following two senses. First, [18] developed their inference procedure based on inverse probability weighting approach because the event time and the censoring time concerned are assumed to be independent. In particular, they made use of the observation that \(P(h_i(T_i; \beta, \tau) \leq 1 \mid \hat{Z}_i) = \tau\) to derive their estimating equation, namely

\[
\frac{1}{n} \sum_{i=1}^{n} \frac{\Delta_i \tilde{Z}_i(0)}{\tilde{G}(\bar{T}_i)} \{I(h_i(T_i; \beta, \tau) \leq 1) - \tau\} = 0,
\]

where \(\tilde{G}(\cdot)\) denotes the Kaplan-Meier estimate of the survival function of the censoring distribution. On the contrary, in this paper, we relax this assumption and adopt the martingale formulation as in (2.2) through which we establish our inference procedure. Second, in [18], specific quantile parameters are estimated independently with no pooling of information across adjoining quantile levels. As we witness in our numerical studies, overlooking such homogeneity results in volatile
estimates, see Example 4 in Section 2.5. This observation indeed motivates the development of our methodology.

2.3 Asymptotic Results

In this section, we state the asymptotic results of the proposed estimator and the corresponding regularity conditions, while the proofs are deferred to the appendix. The required conditions are listed as follows:

**Condition 2.1.** \( |Z_{ik}(0)| + \int_0^\infty |dZ_{ik}(t)| < \infty \) for each \( i = 1, \ldots, n \) and \( k = 1, \ldots, p \).

**Condition 2.2.** \( \beta_0 \in \mathcal{B} \), where \( \mathcal{B} \) is a bounded convex region.

**Condition 2.3.** \( P(T > L | \tilde{Z}) > 0 \) almost surely for all \( \tilde{Z} \), where \( L = \sup_{\beta \in \mathcal{B}} h^{-1}(t; \beta) \).

**Condition 2.4.** \( \inf_{t \geq 0, \beta \in \mathcal{B}} f(t; \beta) > 0 \), where \( f(t; \beta) \) denotes the density function of the unconditional distribution of \( h(T; \beta) \).

**Condition 2.5.** \( \sup_{\beta \in \mathcal{B}} |\Psi(0; \beta)| + \int_0^\infty |d\Psi(t; \beta)| < \infty \), and there exists a function \( \psi(t; \beta) \) such that \( \sup_{t \in [0; \tilde{\tau}_L; \tilde{\tau}_U]; \beta \in \mathcal{B}} |\Psi(t; \beta) - \psi(t; \beta)| \to 0 \) almost surely.

**Condition 2.6.** \( A_n(\beta_0; \tau_L, \tau_U) \) is invertible, where \( A_n(\beta; \tau_L, \tau_U) \) is the gradient of \( \tilde{U}_n(\beta; \tau_L, \tau_U) \) with respect to \( \beta \), and \( \tilde{U}_n(\beta; \tau_L, \tau_U) \) is a smoothed analogy of \( U_n(\beta; \tau_L, \tau_U) \) as defined in the appendix.

Condition 2.1 is a mild smoothness requirement on the covariate processes. In particular, if \( Z_{ik}(t) \) has a uniformly bounded derivative, then it is absolutely continuous and hence of bounded total variation. More generally, \( Z_{ik}(t) \) is allowed to have jumps with sojourn times having uniformly bounded densities. Conditions 2.2, 2.3 and 2.4 are standard assumptions in survival analysis. In particular, using the idea of the functional delta method, see for example, Section 12 of [31], Conditions 2.3 and 2.4 imply the uniform consistency of the Kaplan-Meier estimator \( \hat{F}(t; \beta) \) and hence the quantile estimator \( \hat{q}(\tau; \beta) \) over \( \beta \in \mathcal{B} \). Condition 2.5 imposes restriction on the
weight function so as to guarantee that it is well behaved to achieve the consistency and weak convergence of the proposed estimator. This condition is satisfied for a broad class of weight functions including, but not limited to, the constant and Gehan weights discussed previously. Condition 2.6 is essential to ensure the uniqueness of $\beta_0$ and to recover the asymptotic normality of the estimator from that of the estimating equation.

With the above conditions, we now establish the consistency and weak convergence of the proposed estimator through the following theorems. Corresponding proofs are deferred to the appendix.

**Theorem 2.1.** If Conditions 2.1-2.6 hold, then there exists a fixed neighbourhood of $\beta_0$, denoted by $N(\beta_0)$, such that for any $\hat{\beta} \in N(\beta_0)$, we have $||\hat{\beta} - \beta_0|| \to 0$ almost surely.

**Theorem 2.2.** If Conditions 2.1-2.6 hold, then $n^{1/2}(\hat{\beta} - \beta_0)$ converges weakly to a mean zero multivariate normal with covariance matrix $A(\beta_0; \tau_L, \tau_U)^{-1}V(\beta_0; \tau_L, \tau_U)A(\beta_0; \tau_L, \tau_U)^{-1}$, where $A(\beta_0; \tau_L, \tau_U)$ is the limit of $A_n(\beta_0; \tau_L, \tau_U)$ and $V(\beta_0; \tau_L, \tau_U)$ is defined in the appendix.

In the appendix, we prove Theorems 2.1 and 2.2 by approximating $U_n(\beta; \tau_L, \tau_U)$ with its smoothed analogy $\tilde{U}_n(\beta; \tau_L, \tau_U)$ and hence showing its local asymptotic linearity in $\beta$. In the spirit of [59] and [38], the lack of monotonicity in rank-based estimating equations results in large sample properties in a local sense. Similar conclusion can also be seen in pointwise censored quantile regression models such as [41]. Global consistency might be justified under particular special cases, see, for example, [23] for a scenario where the accelerated failure time model with time-independent covariate and Gehan weight is considered.

### 2.4 Inference

#### 2.4.1 Resampling Scheme

As shown in the proof of Theorem 2.2, the covariance matrix of the limiting distribution of $n^{1/2}(\hat{\beta} - \beta_0)$ involves the unknown $\Lambda(\tau; \beta_0)$. To estimate the variances of the model parameters, we apply the resampling technique proposed in [24] by perturbing the minimand of the estimating
equations. Accordingly, we let \( \{\zeta_1, \ldots, \zeta_n\} \) be i.i.d. non-negative random variables with mean 1 and variance 1, such as exponential random variable with mean 1. As shown in the proof of Theorem 2.3, such distributional requirements on \( \{\zeta_1, \ldots, \zeta_n\} \) are essential to resemble the same limiting distribution upon resampling. We fix the data \( \{\tilde{T}_i, \Delta_i, \bar{Z}_i(\tilde{T}_i), i = 1, \ldots, n\} \) and generate \( \{\zeta_1, \ldots, \zeta_n\} \), then consider the following perturbation of \( U_n(\beta; \tau_L, \tau_U) \), i.e.

\[
U_n^*(\beta; \tau_L, \tau_U) = \frac{1}{n} \sum_{i=1}^{n} \zeta_i \Psi(h_i(\tilde{T}_i; \beta); \beta) \left\{ Z_i(\tilde{T}_i) - \frac{S^{(1)}(h_i(\tilde{T}_i; \beta); \beta)}{S^{(0)}(h_i(\tilde{T}_i; \beta); \beta)} \right\} \times \left\{ N_i(\hat{q}(\tau_U; \beta); \beta) - N_i(\hat{q}(\tau_L; \beta); \beta) \right\}.
\]

Again we find the root of \( U_n^*(\beta; \tau_L, \tau_U) \), denoted by \( \beta^* \), by minimizing its \( L_1 \) norm. While keeping the data fixed, we repeat the above procedure \( B \) times, i.e. for each \( r = 1, \ldots, B \), generate a set of \( \{\zeta_1, \ldots, \zeta_n\} \) and obtain a corresponding realization of \( \beta^* \), denoted by \( \beta^*_r \). Now, we state the following theorem, which will be proven in the appendix via a direct application of the results in [24].

**Theorem 2.3.** *If Conditions 2.1-2.6 hold, then the conditional distribution of \( n^{1/2}(\beta^* - \hat{\beta}) \) given the observed data converges weakly to the same limiting distribution of \( n^{1/2}(\hat{\beta} - \beta_0) \).*

As a consequence of Theorem 2.3, we may use \( \{\beta^*_r, r = 1, \ldots, B\} \) to estimate the limiting distribution of the parameter estimates, i.e. the variance of \( \hat{\beta} \) can be approximated by the sample variance of \( \{\beta^*_r, r = 1, \ldots, B\} \), and the confidence interval for \( \beta_0 \) can be constructed using the empirical quantile of \( \{\beta^*_r, r = 1, \ldots, B\} \) or by normal approximation.

### 2.4.2 Two Stage Testing Procedure for Homogeneity

As discussed previously, the choice of \( [\tau_L, \tau_U] \) essentially depends on the primary interest of individual users. Based on the derivation of (2.4), the estimating equation remains valid as long as the homogeneity condition (2.1) holds over \( [\tau_L, \tau_U] \). In such cases, the covariate effects represented by the regression coefficient are identical over the range of quantiles in which (2.1) holds. To yield greater numerical stability, we naturally ask the question if (2.1) remains valid
for an extended range of quantiles in lieu of the predefined interval chosen by the user. For cases where global homogeneity holds, i.e., there exists $\beta_0$ such that (2.1) holds all over $[0, 1]$, it suffices to adopt the conventional accelerated failure time model. In general, however, it is essential to test the constancy of $\beta_0$ to check whether model (2.1) holds for a pre-specified region of $[\tau_L, \tau_U]$. Accordingly, we have two concerns to address, namely (i) whether or not the accelerated failure time model should be adopted, and (ii) whether or not the designated range of quantiles can be extended given that (i) is rejected.

For a specific interval $[\tau_L, \tau_U] \subset [0, 1]$, suppose we want to test the constancy of $\beta_0$ within the range. A straightforward approach is to divide the interval into two halves, namely $[\tau_L, \tau_M]$ and $[\tau_M, \tau_U]$ where $\tau_M = (\tau_L + \tau_U)/2$, and test whether or not the corresponding parameters in the two non-overlapping regions are equal. More precisely, let $\beta_A$ denote the model parameter when the homogeneity condition (2.1) is assumed over a set $A$, and $\hat{\beta}_A$ be the corresponding estimate using (2.4), it suffices to test if $\beta_{[\tau_L, \tau_M]} = \beta_{[\tau_M, \tau_U]}$. One possible test is to reject the null hypothesis when the absolute deviation statistic $T = \sum_{k=1}^p |\hat{\beta}^{(k)}_{[\tau_L, \tau_M]} - \hat{\beta}^{(k)}_{[\tau_M, \tau_U]}|$ is large. As a result of Theorem 2.3, the rejection cutoff can be determined by the empirical quantile of the resampling-based $T^* = \sum_{k=1}^p \left( (\beta^{(k)}_{[\tau_L, \tau_M]} - \beta^{(k)}_{[\tau_M, \tau_U]}) - (\hat{\beta}^{(k)}_{[\tau_L, \tau_M]} - \hat{\beta}^{(k)}_{[\tau_M, \tau_U]}) \right)$. Theoretically, the test works for any arbitrary interval provided the sample size is large enough. However, because of the practical concern on the sample size of our data analysis, we propose a two stage testing procedure below that would address our two concerns with reasonable statistical powers.

In particular, if one concerns about the lower quantiles $[0, 0.25]$ of the failure time, she may question if the covariate effect is in fact homogeneous over $[0, 1]$, and if not, whether it is same as that of $[0.25, 0.5]$. Denote $A_1 = [0, 0.25]$, $A_2 = [0.25, 0.5]$, $A_3 = [0.5, 0.75]$ and $A_4 = [0.75, 1]$, then the two questions translate to testing the two null hypotheses of $\beta_{A_1} = \beta_{A_2} = \beta_{A_3} = \beta_{A_4}$ and $\beta_{A_1} = \beta_{A_2}$ respectively. To this end, we first test $\beta_{A_1} = \beta_{A_2} = \beta_{A_3} = \beta_{A_4}$ using the test statistic $T_1 = \sum_{k=1}^p \sum_{i \neq j} |\hat{\beta}^{(k)}_{A_i} - \hat{\beta}^{(k)}_{A_j}|$, where the rejection cutoff is taken to be the empirical quantile of $T^*_1 = \sum_{k=1}^p \sum_{i \neq j} \left( (\beta^{(k)}_{A_i} - \beta^{(k)}_{A_j}) - (\hat{\beta}^{(k)}_{A_i} - \hat{\beta}^{(k)}_{A_j}) \right)$. Suppose the null hypothesis is not rejected, we may simply apply the accelerated failure time model to obtain an estimate of global covariate
effects. On the other hand, provided that the first-stage test is rejected, we shall proceed to test the second hypothesis of $\beta_{A_1} = \beta_{A_2}$. Similarly, we consider the test statistic $T_2 = \sum_{k=1}^{p} | \hat{\beta}_{A_1}^{(k)} - \hat{\beta}_{A_2}^{(k)} |$ and use the empirical quantiles of $T_2^* = \sum_{k=1}^{p} | (\beta_{A_1}^{(k)} - \beta_{A_2}^{(k)}) - (\hat{\beta}_{A_1}^{(k)} - \hat{\beta}_{A_2}^{(k)}) |$ as cutoff. If this second-stage test is rejected, we shall proceed with $[0, 0.25]$ for inference, otherwise $[0, 0.5] = A_1 \cup A_2$ can be used to improve numerical stability.

The choices of $A_j$ depends on the quantile range of interest. If the upper quantiles $[0.75, 1]$ are concerned, we pick $A_1 = [0.75, 1]$, $A_2 = [0.5, 0.75]$, $A_3 = [0.25, 0.5]$ and $A_4 = [0, 0.25]$. Meanwhile, $A_1 = [0.375, 0.625]$, $A_2 = [0.25, 0.375] \cup [0.625, 0.75]$, $A_3 = [0, 0.25]$ and $A_4 = [0.75, 1]$ can be chosen if the median covariate effect is of interest. Because of the two-stage nature of this test, we adopt a Bonferroni type adjustment to the significance level for each stage. Alternative modification to the significance level such as an adaptive distribution to each stage may also be considered for more sophisticated testing powers, but the discussion is beyond the scope of the paper. In spite of the rather conservative Bonferroni correction, we manage to observe some reasonable empirical powers in our numerical studies. Also, one may generalize the two-stage test by slicing $[0, 1]$ into finer sub-regions and consider a multi-stage test so as to find out the region for which homogeneity holds more precisely, provided reasonable power can be achieved with the given sample size.

### 2.5 Simulations

In the subsequent simulation examples, we consider a covariate process that changes in values exactly $d - 1$ times at the time nodes $\{W_{i1}, \ldots, W_{i,d-1}\}$ where $0 = W_{i0} < W_{i1} < \cdots < W_{i,d-1} < W_{id} = \infty$. In particular, let $Z_i(t) = (Z_{i1}(t), Z_{i2}(t))^\top$ be a two-dimensional covariate process with $d = 3$ subintervals. Specifically, let

\[
Z_{i1}(t) = \sum_{s=1}^{d} X_{i1s} I(W_{i,s-1} \leq t < W_{is}),
\]

\[
Z_{i2}(t) = \sum_{s=1}^{d} X_{i2s} I(W_{i,s-1} \leq t < W_{is}),
\]
where $X_{i1s}, X_{i2s} \sim \text{Gamma}(2, 1/2)$ independently for each $i = 1, \ldots, n$ and $s = 1, \ldots, d$. The true regression coefficient is $\beta_0 = (1, 1)^T$ and the common conditional quantile function $\log q(\tau) = Q_{\varepsilon}(\tau)$ where $\varepsilon \sim N(0, 1)$. For $i = 1, \ldots, n$ and $s = 1, \ldots, d - 1$, the time nodes are recovered recursively through $W_{is} = W_{i,s-1} + S_{is}$, where $S_{is}$ (the sojourn times between the time nodes) are independently distributed as $U(0, 1)$. Recall that the covariate processes $\{Z_i(s), 0 \leq s \leq \tilde{T}_i\}$ can only be observed up to the survival times, so we generate the sojourn time in a way that the simulated failure times $\{T_i, i = 1, \ldots, n\}$ spread through the subintervals derived from $\{W_{i0}, \ldots, W_{id}\}$. One can also notice that the resulting conditional quantile function of the failure time remains monotone increasing for any realization of the covariate process.

Example 1: Homogeneous Error

Given the above setup, the first numerical example demonstrates that the proposed method resembles the classic accelerated failure time model when the homogeneous conditional quantile assumption (2.1) is satisfied for $[\tau_L, \tau_U] = [0, 1]$. To this end, we generate the censoring times $\{C_i, i = 1, \ldots, n\}$ from i.i.d. exponential distribution with mean $\exp\{c + 100Z_{i1}(0) + 100Z_{i2}(0)\}$ so that the censoring times depend on the covariate processes, where we set $c = -150$ or $c = -110$ to achieve a censoring rate of 40% or 20% respectively. With $\beta_0 = (1, 1)^T$ and $\log q(\tau) = Q_{\varepsilon}(\tau)$, we have (2.1) holds for $[\tau_L, \tau_U] = [0, 1]$, and thus for a subinterval $[\tau_L, \tau_U] = [0, 0.5]$. We use (2.4) with $[\tau_L, \tau_U] = [0, 0.5]$ to obtain our parameter estimates $\hat{\beta}$. We generate 500 Monte-Carlo datasets with $n = 200$ or $n = 400$ sample points each, and adopt the aforementioned resampling procedure with $B = 200$ and $\zeta_i$ exponentially distributed with mean $1$ to reconcile the asymptotic distribution of the parameter estimates. The numerical optimization is carried out via the differential evolution algorithm in MATLAB. We adopt existing quantile regression techniques with time-independent covariates, such as [53]’s approach, to obtain preliminary initial estimates. The average value over time of the covariate process observed in the study is considered as a time-independent proxy to be used in their methods. Our numerical experience suggests that the monotonicity of [53]’s estimating equation with respect to the parameter of interest serves as a
Table 2.1: Simulation Results with Homogeneous Error

<table>
<thead>
<tr>
<th>Censoring Rate</th>
<th>$n = 200$</th>
<th></th>
<th>$n = 400$</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>20%</td>
<td>40%</td>
<td>20%</td>
<td>40%</td>
<td>20%</td>
<td>40%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\hat{\beta}^{(1)}$</td>
<td>$\hat{\beta}^{(2)}$</td>
<td>$\hat{\beta}^{(1)}$</td>
<td>$\hat{\beta}^{(2)}$</td>
<td>$\hat{\beta}^{(1)}$</td>
<td>$\hat{\beta}^{(2)}$</td>
<td></td>
</tr>
<tr>
<td>Proposed (Unweighted)</td>
<td>Bias</td>
<td>-0.009</td>
<td>-0.011</td>
<td>-0.008</td>
<td>-0.011</td>
<td>0.000</td>
<td>-0.007</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>0.149</td>
<td>0.150</td>
<td>0.162</td>
<td>0.171</td>
<td>0.105</td>
<td>0.105</td>
</tr>
<tr>
<td></td>
<td>SE</td>
<td>0.151</td>
<td>0.152</td>
<td>0.168</td>
<td>0.168</td>
<td>0.105</td>
<td>0.105</td>
</tr>
<tr>
<td></td>
<td>CP</td>
<td>0.946</td>
<td>0.942</td>
<td>0.954</td>
<td>0.938</td>
<td>0.940</td>
<td>0.936</td>
</tr>
<tr>
<td>Proposed (Gehan weighted)</td>
<td>Bias</td>
<td>-0.009</td>
<td>-0.009</td>
<td>-0.010</td>
<td>-0.013</td>
<td>0.000</td>
<td>-0.007</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>0.146</td>
<td>0.147</td>
<td>0.160</td>
<td>0.167</td>
<td>0.103</td>
<td>0.104</td>
</tr>
<tr>
<td></td>
<td>SE</td>
<td>0.146</td>
<td>0.148</td>
<td>0.163</td>
<td>0.165</td>
<td>0.103</td>
<td>0.102</td>
</tr>
<tr>
<td></td>
<td>CP</td>
<td>0.938</td>
<td>0.952</td>
<td>0.954</td>
<td>0.946</td>
<td>0.946</td>
<td>0.938</td>
</tr>
<tr>
<td>LY</td>
<td>Bias</td>
<td>-0.006</td>
<td>-0.009</td>
<td>-0.002</td>
<td>-0.011</td>
<td>-0.004</td>
<td>-0.005</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>0.132</td>
<td>0.127</td>
<td>0.149</td>
<td>0.146</td>
<td>0.092</td>
<td>0.093</td>
</tr>
<tr>
<td>GGR ($\tau = 0.25$)</td>
<td>Bias</td>
<td>-0.767</td>
<td>-0.763</td>
<td>-0.849</td>
<td>-0.817</td>
<td>-0.722</td>
<td>-0.717</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>0.959</td>
<td>1.078</td>
<td>0.701</td>
<td>0.743</td>
<td>0.564</td>
<td>0.507</td>
</tr>
<tr>
<td>GGR ($\tau = 0.50$)</td>
<td>Bias</td>
<td>-0.768</td>
<td>-0.807</td>
<td>-0.910</td>
<td>-0.783</td>
<td>-0.700</td>
<td>-0.743</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>1.082</td>
<td>1.901</td>
<td>1.452</td>
<td>1.540</td>
<td>1.154</td>
<td>0.421</td>
</tr>
<tr>
<td>GGR ($\tau = 0.75$)</td>
<td>Bias</td>
<td>-0.748</td>
<td>-0.706</td>
<td>-0.766</td>
<td>-0.529</td>
<td>-0.789</td>
<td>-0.804</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>2.029</td>
<td>2.251</td>
<td>1.869</td>
<td>6.726</td>
<td>0.391</td>
<td>0.453</td>
</tr>
<tr>
<td>Composite - GGR</td>
<td>Bias</td>
<td>-0.839</td>
<td>-0.847</td>
<td>-0.898</td>
<td>-0.890</td>
<td>-0.770</td>
<td>-0.751</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>0.983</td>
<td>1.157</td>
<td>0.721</td>
<td>0.770</td>
<td>0.604</td>
<td>0.530</td>
</tr>
</tbody>
</table>

*Approaches LY and GGR denote [38]'s and [18]'s proposals respectively. Bias, SD, SE, CP denote empirical bias, empirical standard deviation, average standard error, and empirical coverage probability of the 95% confidence intervals, respectively.*
Table 2.2: Simulation Results with Homogeneity assumed for $\tau \leq 0.25$

| Censoring Rate | $n = 200$ | | $n = 400$ | | |
|---------------|-----------|---|-----------|---|
|               | $20\%$ | $40\%$ | $20\%$ | $40\%$ | |
|               | $\hat{\beta}^{(1)}$ | $\hat{\beta}^{(2)}$ | $\hat{\beta}^{(1)}$ | $\hat{\beta}^{(2)}$ | $\hat{\beta}^{(1)}$ | $\hat{\beta}^{(2)}$ |
| Proposed (Unweighted) Bias | -0.017 | -0.014 | -0.023 | -0.023 | -0.001 | -0.007 | -0.001 | -0.005 |
| SD | 0.201 | 0.203 | 0.238 | 0.230 | 0.134 | 0.136 | 0.163 | 0.156 |
| SE | 0.198 | 0.201 | 0.233 | 0.233 | 0.137 | 0.135 | 0.154 | 0.155 |
| CP | 0.944 | 0.952 | 0.946 | 0.950 | 0.948 | 0.934 | 0.926 | 0.954 |
| Proposed (Gehan weighted) Bias | -0.016 | -0.013 | -0.017 | -0.016 | 0.000 | -0.005 | 0.001 | -0.000 |
| SD | 0.201 | 0.202 | 0.230 | 0.222 | 0.135 | 0.134 | 0.155 | 0.150 |
| SE | 0.196 | 0.198 | 0.227 | 0.229 | 0.134 | 0.133 | 0.152 | 0.150 |
| CP | 0.950 | 0.950 | 0.954 | 0.946 | 0.944 | 0.932 | 0.930 | 0.946 |
| LY Bias | 0.809 | 0.804 | 0.810 | 0.796 | 0.822 | 0.820 | 0.819 | 0.820 |
| SD | 0.183 | 0.181 | 0.213 | 0.220 | 0.129 | 0.130 | 0.155 | 0.145 |
| GGR ($\tau = 0.25$) Bias | -0.609 | -0.650 | -0.685 | -0.753 | -0.495 | -0.495 | -0.596 | -0.592 |
| SD | 0.626 | 0.633 | 0.516 | 0.492 | 0.427 | 0.431 | 0.409 | 0.418 |
| GGR ($\tau = 0.50$) Bias | -0.753 | -0.802 | -0.845 | -0.863 | -0.735 | -0.741 | -0.790 | -0.786 |
| SD | 0.614 | 0.548 | 0.533 | 0.571 | 0.513 | 0.514 | 0.531 | 0.482 |
| GGR ($\tau = 0.75$) Bias | -1.701 | -1.733 | -1.750 | -1.782 | -1.445 | -1.493 | -1.421 | -1.415 |
| SD | 0.769 | 0.722 | 0.670 | 0.677 | 0.772 | 0.812 | 0.838 | 0.828 |
| Composite - GGR Bias | -0.688 | -0.727 | -0.758 | -0.811 | -0.551 | -0.551 | -0.627 | -0.659 |
| SD | 0.740 | 0.733 | 0.535 | 0.575 | 0.484 | 0.559 | 0.476 | 0.488 |

*Approaches LY and GGR denote [38]’s and [18]’s proposals respectively. Bias, SD, SE, CP denote empirical bias, empirical standard deviation, average standard error, and empirical coverage probability of the 95% confidence intervals, respectively.*
<table>
<thead>
<tr>
<th>Approach</th>
<th>Censoring Rate</th>
<th>n = 200</th>
<th>n = 400</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>20%</td>
<td>40%</td>
<td>20%</td>
</tr>
<tr>
<td></td>
<td>( \hat{\beta}^{(1)} )</td>
<td>( \hat{\beta}^{(2)} )</td>
<td>( \hat{\beta}^{(1)} )</td>
</tr>
<tr>
<td>Proposed (Unweighted)</td>
<td>Bias</td>
<td>0.079</td>
<td>0.079</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>0.341</td>
<td>0.348</td>
</tr>
<tr>
<td></td>
<td>SE</td>
<td>0.314</td>
<td>0.315</td>
</tr>
<tr>
<td></td>
<td>CP</td>
<td>0.930</td>
<td>0.940</td>
</tr>
<tr>
<td>Proposed (Gehan weighted)</td>
<td>Bias</td>
<td>0.028</td>
<td>0.015</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>0.227</td>
<td>0.218</td>
</tr>
<tr>
<td></td>
<td>SE</td>
<td>0.199</td>
<td>0.199</td>
</tr>
<tr>
<td></td>
<td>CP</td>
<td>0.960</td>
<td>0.966</td>
</tr>
<tr>
<td>LY</td>
<td>Bias</td>
<td>0.228</td>
<td>0.203</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>0.363</td>
<td>0.358</td>
</tr>
<tr>
<td>GGR (( \tau = 0.25 ))</td>
<td>Bias</td>
<td>-1.318</td>
<td>-1.350</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>0.811</td>
<td>0.781</td>
</tr>
<tr>
<td>GGR (( \tau = 0.50 ))</td>
<td>Bias</td>
<td>-0.827</td>
<td>-0.805</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>0.721</td>
<td>0.726</td>
</tr>
<tr>
<td>GGR (( \tau = 0.75 ))</td>
<td>Bias</td>
<td>-0.804</td>
<td>-0.740</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>0.606</td>
<td>0.656</td>
</tr>
<tr>
<td>Composite - GGR</td>
<td>Bias</td>
<td>-0.837</td>
<td>-0.826</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>0.697</td>
<td>0.676</td>
</tr>
</tbody>
</table>

\(^{a}\)Approaches LY and GGR denote [38]'s and [18]'s proposals respectively. Bias, SD, SE, CP denote empirical bias, empirical standard deviation, average standard error, and empirical coverage probability of the 95% confidence intervals, respectively.
reasonable guideline for initial value selection.

Table 2.1 reports the simulation results for our proposal in terms of empirical bias (Bias), empirical standard deviation (SD), average standard error (SE) based on the resampling method, and coverage rate (CP) of the 95% confidence intervals constructed by normal approximation. The results obtained by our proposal using the different weight functions are quite similar, so the estimation is not sensitive with respect to the specific choice of weight function. The empirical performance is compared against [38]’s and [18]’s proposals. Under the current setting where the accelerated failure time model holds, the proposed method produces competitive results in empirical biases and standard deviations when compared to [38]. Moreover, the standard errors are close to the empirical standard deviations and the coverage probabilities are close to the nominal level of 95%, validating the aforementioned resampling method. Our proposal loses some efficiency against [38]’s method as reflected in slightly larger empirical standard deviations. This is reasonable because we only consider (2.4) over a subinterval \([\tau_L, \tau_U] = [0, 0.5]\) in which (2.1) holds to obtain \(\hat{\beta}\), while they consider the whole region \([\tau_L, \tau_U] = [0, 1]\). On the other hand, despite [18]’s method only requires (2.1) to hold for a single value of \(\tau\), their estimates demonstrate larger empirical biases and standard deviations, especially when the censoring rate is high. One reason is that their methodology only works for cases with independent censoring, which obviously does not hold in this example as the censoring distribution depends on the covariate processes.

The composite quantile regression introduced in [62] also shares the pooling effect that our methodology provides, although the two ideas have different modeling motivations. We include their possible extension to censored data for comparison, which we refer to as Composite-[18]. This is because the composite feature could possibly improves the numerical performance of the ordinary quantile regression, it involves a higher dimensional optimization procedure, hence the numerical problem turns out to be more challenging here. It is shown that the Composite-[18]’s method shares similar empirical bias.
with [18] while its standard deviations are slightly inferior, which are both not comparable to the proposed method. The main reason is that the estimates being composited requires an independent censoring assumption which does not hold in our examples.

Example 2: Heterogeneity in Upper Tail Quantiles

The second example is more illustrative to show that our proposed method generalizes the accelerated failure time model proposed by [38], in the sense that it relaxes homogeneous conditional quantile assumption (2.1) for all quantile levels. With the same covariate processes described above, we now generate the failure times based on $Q_{T_i} (\tau \mid \tilde{Z}_i) = h_i^{-1}(q(\tau); \beta_0 + I(\tau > 0.5))$. Here, $\beta_0 + I(\tau > 0.5)$ reduces to $\beta_0$ for $\tau \leq 0.5$, which implies that model (2.1) holds for $[\tau_L, \tau_U] = [0, 0.5]$. But the expression becomes $\beta_0 + 1$, not the same constant $\beta_0$, for $[\tau_L, \tau_U] = [0.5, 1]$, therefore model (2.1) is not valid over the entire $[\tau_L, \tau_U] = [0, 1]$, Owing to the heterogeneity across the two subregions of quantiles, the accelerated failure time model assumed by [38] does not hold here. Nevertheless, suppose we are only interested in the covariate effect on the lower quantiles $[0, 0.25]$, then our proposal of using (2.4) with $[\tau_L, \tau_U] = [0, 0.25]$ to compute $\hat{\beta}$ will still be a justified methodology. Under the same simulation setting as in the first example, Table 2.2 reports the simulation results in the same format as Table 2.1 does. It is clear that the proposed estimator produces reasonable results as in the homogeneous setting described in the first example, because our model (2.1) holds for $[\tau_L, \tau_U] = [0, 0.25]$ here. Meanwhile, it is reasonable that [38]’s method produces substantially larger empirical bias and standard deviation than it does in the first example, because of the violation of their model assumption. The performance of [18]’s method and its composite counterpart are worse than our proposal as in the first example, which is due to the same reason of the violation of the independent censoring assumption.
Example 3: Heterogeneity in Both Lower and Upper Tail Quantiles

The third example considers failure times with quantile function

$$Q_{T_i}(\tau | \tilde{Z}_i) = h_i^{-1}(q(\tau); \beta_0 - Q_e(\tau)I(\tau < 0.25) + Q_e(\tau)I(\tau > 0.75)).$$

The second argument of the last display reduces to $\beta_0$ for $\tau \in [0.25, 0.75]$. Therefore, model (2.1) holds for $[\tau_L, \tau_U] = [0.25, 0.75]$ whereas such equivalence does not hold beyond that. Suppose we want to examine the covariate effect on the failure time around the median, we may consider a subinterval of quantiles around the median, say $[\tau_L, \tau_U] = 0.5 \pm 0.25 = [0.25, 0.75]$ for the estimating equation (2.4). As shown in Table 2.3, the empirical biases and standard deviations of our method can mostly demonstrate the advantage as compared to the competitors. Same as the previous scenario, [38]'s method, [18]'s proposal and its composite modification yield significantly larger biases because of violation of the homogeneous conditional quantile assumption over all quantiles and the independent censoring assumption, respectively.

With a significance level of $10\%$, Table 2.4 reports the empirical rejection rate of the two-stage test described in Section 2.4 for the previous three simulation examples with the unweighted estimating equation. Note that after the Bonferroni adjustment, the significance level of each stage will remain at $5\%$. We assume $A_1 = [0, 0.25]$, $A_2 = [0.25, 0.5]$, $A_3 = [0.5, 0.75]$ and $A_4 = [0.75, 1]$ for the first two examples; $A_1 = [0.375, 0.625]$, $A_2 = [0.25, 0.375] \cup [0.625, 0.75]$, $A_3 = [0, 0.25]$ and $A_4 = [0.75, 1]$ for the third example. It is shown that the test has empirical type I errors well controlled around the Bonferroni adjusted significance level, with reasonable powers under the alternative hypothesis. The test is seen to provided enough statistical power for rejection in the data analysis of Section 2.6 despite the possibly conservative Bonferroni correction.

Example 4: Independent Censoring

Despite the proposed method assumes a local homogeneity around the interested quantile level compared to the quantile regression method described by [18], it effectively exploits more infor-
Table 2.4: Empirical Rejection Rate of the Test for Homogeneity

<table>
<thead>
<tr>
<th>Censoring Rate</th>
<th>20%</th>
<th>40%</th>
<th>20%</th>
<th>40%</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Situations where the null is true</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Example 1, $T_1$</td>
<td>0.040</td>
<td>0.030</td>
<td>0.028</td>
<td>0.030</td>
</tr>
<tr>
<td>Example 2, $T_2$</td>
<td>0.034</td>
<td>0.028</td>
<td>0.052</td>
<td>0.050</td>
</tr>
<tr>
<td>Example 3, $T_2$</td>
<td>0.048</td>
<td>0.052</td>
<td>0.050</td>
<td>0.070</td>
</tr>
<tr>
<td>Example 3, $T_3$</td>
<td>0.008</td>
<td>0.020</td>
<td>0.018</td>
<td>0.026</td>
</tr>
<tr>
<td><strong>Situations where the null is false</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Example 2, $T_1$</td>
<td>0.904</td>
<td>0.806</td>
<td>0.988</td>
<td>0.976</td>
</tr>
<tr>
<td>Example 3, $T_1$</td>
<td>0.632</td>
<td>0.492</td>
<td>0.778</td>
<td>0.644</td>
</tr>
</tbody>
</table>

Table 2.5: Simulation Results with Homogeneous Error and Independent Censoring

<table>
<thead>
<tr>
<th>Estimate</th>
<th>95% CI</th>
<th>Estimate</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>$\hat{\beta}^{(1)}$</strong></td>
<td></td>
<td><strong>$\hat{\beta}^{(2)}$</strong></td>
<td></td>
</tr>
<tr>
<td>Proposed ([$\tau_L, \tau_U$] = [0.25, 0.75])</td>
<td>0.777 (0.444, 1.110)</td>
<td>0.776 (0.450, 1.103)</td>
<td></td>
</tr>
<tr>
<td>GGR ($\tau = 0.25$)</td>
<td>0.771 (−0.007, 1.549)</td>
<td>0.804 (−0.042, 1.649)</td>
<td></td>
</tr>
<tr>
<td>GGR ($\tau = 0.30$)</td>
<td>0.749 (−0.011, 1.508)</td>
<td>0.814 (−0.025, 1.652)</td>
<td></td>
</tr>
<tr>
<td>GGR ($\tau = 0.35$)</td>
<td>0.788 (0.021, 1.555)</td>
<td>0.772 (−0.031, 1.576)</td>
<td></td>
</tr>
<tr>
<td>GGR ($\tau = 0.40$)</td>
<td>0.740 (−0.016, 1.496)</td>
<td>0.747 (−0.014, 1.509)</td>
<td></td>
</tr>
<tr>
<td>GGR ($\tau = 0.45$)</td>
<td>0.736 (0.049, 1.422)</td>
<td>0.734 (0.070, 1.399)</td>
<td></td>
</tr>
<tr>
<td>GGR ($\tau = 0.50$)</td>
<td>0.731 (−0.015, 1.476)</td>
<td>0.754 (0.030, 1.478)</td>
<td></td>
</tr>
<tr>
<td>GGR ($\tau = 0.55$)</td>
<td>0.725 (−0.015, 1.464)</td>
<td>0.710 (−0.005, 1.426)</td>
<td></td>
</tr>
<tr>
<td>GGR ($\tau = 0.60$)</td>
<td>0.694 (0.011, 1.377)</td>
<td>0.699 (−0.079, 1.476)</td>
<td></td>
</tr>
<tr>
<td>GGR ($\tau = 0.65$)</td>
<td>0.672 (0.005, 1.339)</td>
<td>0.676 (0.050, 1.302)</td>
<td></td>
</tr>
<tr>
<td>GGR ($\tau = 0.70$)</td>
<td>0.677 (−0.004, 1.358)</td>
<td>0.647 (−0.010, 1.305)</td>
<td></td>
</tr>
<tr>
<td>GGR ($\tau = 0.75$)</td>
<td>0.658 (−0.003, 1.318)</td>
<td>0.649 (−0.083, 1.381)</td>
<td></td>
</tr>
</tbody>
</table>

*Approach GGR denotes [18]'s proposal.

27
mation than the latter to allow a more numerically stable estimation, provided such assumption holds. Here, we present a simulation example to demonstrate that the stability in the proposed method offers a more reliable inference across a range of quantiles even under the independent censoring setup. We consider covariate processes as described in the previous examples, while the true regression coefficients are now \( \beta_0 = (0.8,0.8)^\top \). The failure times are generated based on the conditional quantile function of \( Q_{T_i}(\tau \mid \tilde{Z}_i) = h_i^{-1}(q(\tau); \beta_0) \), while the censoring times are simulated from i.i.d. exponential distribution with mean \( \exp(3.6) \) to yield a censoring rate of 20\%. The failure times and the censoring times are then marginally independent in which case the condition assumed in [18] holds. Their method is implemented for every quantile level from 0.25 to 0.75 with an interval of 0.05, while the proposed method is adopted with \( [\tau_L, \tau_U] = [0.25,0.75] \) for comparison.

Table 2.5 shows that [18]’s proposal does not provide consistent results over this range of quantiles despite the independent censoring assumption holds, which in turn lead to erroneous conclusions. The empirical biases are worse for higher quantile levels and the empirical confidence intervals for both parameter estimates are sensitive and fluctuate across neighbourhooding quantiles. When a certain level of quantile is considered, the confidence interval may or may not cover zero due to that particular choice, which results in inconsistent conclusions. For instance, the first covariate would be concluded as insignificant when \( \tau = 0.25,0.5,0.75 \) are considered. On the other hand, provided that the homogeneity assumption holds over \( [\tau_L, \tau_U] \), our proposal pools information over this range to produce a more reliable estimate, in the sense that the empirical bias is small and the empirical confidence intervals do not cover zero, meaning that the covariates are significant as predicted.

2.6 Data Analysis

In this section, we demonstrate the proposed method with the Stanford heart transplant data. The study is described in details in [8] and [12]; see also [1] for a brief literature review on the statistical analysis of the data. [38] analyzed the effect of heart transplantation on the lifetime
using the accelerated failure time model with time-dependent covariates. Their results suggest that younger patients with lower mismatch scores are more likely to benefit from transplantation, while it is less beneficial for older patients with higher mismatch scores to receive transplant. Here, in particular, we are interested in examining the effect of heart transplantation in a neighbourhood of the median lifetime, and those in the lower quantile and the upper quantile.

Same as [38], we consider a three-dimensional time-dependent covariate process in our analysis, namely transplant status, age at transplant, and mismatch score, where the last two covariates are assumed to be zero before transplant. Mathematically, let $W_i$ denote the waiting time to transplant of the $i$-th patient, i.e. the duration from the date of acceptance into the program to the date of transplant. We take $W_i$ to be $\infty$ if the $i$-th patient never received transplantation. If we denote $Z_i(t) = \{Z_{i1}(t), Z_{i2}(t), Z_{i3}(t)\}^\top$ to be the three-dimensional covariate process of the $i$-th patient, then

\[
Z_{i1}(t) = I(t \geq W_i),
Z_{i2}(t) = (\text{age at transplant} - 35)I(t \geq W_i),
Z_{i3}(t) = (\text{mismatch score} - 0.5)I(t \geq W_i).
\]

Here, we centralize $Z_{i2}(t)$ and $Z_{i3}(t)$ so that the regression coefficient associated with $Z_{i1}(t)$ can be interpreted as the effect of transplantation for a patient who is 35 years old at the time of transplant and has a mismatch score of 0.5.

We start with analyzing the covariate effects around the median lifetime. Assuming a 10% level of significance and the Bonferroni adjustment, we follow the two-stage testing procedure described in Section 2.4 to check for heterogeneity and find out the range of quantiles around the median for which homogeneity holds. The empirical $p$-value of the first stage, i.e. the empirical proportion of $\mathcal{T}_1^*$ exceeding $\mathcal{T}_i$, is 0, suggesting that the error is not homogeneous and so the classical accelerated failure model shall not be directly applied. Meanwhile, the second stage produces an empirical $p$-value of 0.046, which means the covariate effects of $[0.375, 0.625]$ and $[0.25, 0.375] \cup [0.625, 0.75]$ are different. As a result, if we have to examine the covariate effects
around the median lifetime, we should work with \([\tau_L, \tau_U] = [0.375, 0.625]\). We repeat the above two-stage test for the cases when the lower quantiles or the upper quantiles are concerned. Suppose the lifetime in the lower quantiles is of interest, the empirical \(p\)-values of the two stages are 0 and 0.028 respectively. We conclude that heterogeneity exists, and the covariate effects over \([0, 0.25]\) and those of \([0.25, 0.5]\) are different, so only \([\tau_L, \tau_U] = [0, 0.25]\) should be considered if we want to examine the covariate effects on the lower quantiles. Finally, when the lifetime in the upper quantiles is of interest, the empirical \(p\)-values of the two stages are 0 and 0.132 respectively, i.e. the covariate effects over \([0.75, 1]\) and those of \([0.5, 0.75]\) are not distinguishable, so we would use \([\tau_L, \tau_U] = [0.5, 1]\) in the proposed method if we are focusing on the covariate effects on the upper quantiles.

Table 2.6 reports the parameter estimates and their corresponding 95% confidence intervals computed using our proposal with \([\tau_L, \tau_U] = [0, 0.25], [0.375, 0.625]\) and \([0.5, 1]\) in (2.4) (respectively the ranges determined from the above tests for the lower quantiles, neighbourhood of median and upper quantiles). The 95% confidence intervals are constructed by normal approximation under the aforementioned resampling scheme with \(B = 500\). Estimation results based on the accelerated failure time model, the Cox model, and the quantile model (at \(\tau = 0.25, 0.5, 0.75\)), each under the framework of time-dependent covariates, are also provided as in [18]. With a significance level of 5%, both the accelerated failure time model and the Cox model conclude that the transplant status and the age at transplant are significant, whereas the mismatch score is not. Based on [18]’s approach, the quantile regression models at 25th, 50th and 75th percentiles reveal that the covariate effects are not significant apart from the exception of transplant status at 25th and 50th percentiles, which might be due to the higher variance associated with their estimator as shown in the previous simulation examples. On the other hand, while Composite-[18]’s method enhances the numerical performance of [18] for the lower quantiles in terms of reducing the standard errors, its estimates for the larger quantiles are noticeably unstable. The possible causes could be the higher dimensional optimization problem involved in the composite method, and that [18]’s procedure is intrinsically volatile across extreme tail quantiles in nature. On the other hand, the
Table 2.6: Parameter Estimates of the Stanford Heart Transplant Data

<table>
<thead>
<tr>
<th></th>
<th>Transplant Status</th>
<th>Age at Transplant–35</th>
<th>Mismatch Score–0.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proposed (([\tau_L, \tau_U]) = [0, 0.25])</td>
<td>Estimate</td>
<td>0.772</td>
<td>-0.147</td>
</tr>
<tr>
<td></td>
<td>95% CI</td>
<td>(0.247, 1.296)</td>
<td>(-0.216, -0.078)</td>
</tr>
<tr>
<td>Proposed (([\tau_L, \tau_U]) = [0.375, 0.625])</td>
<td>Estimate</td>
<td>1.059</td>
<td>-0.225</td>
</tr>
<tr>
<td></td>
<td>95% CI</td>
<td>(0.084, 2.033)</td>
<td>(-0.293, -0.158)</td>
</tr>
<tr>
<td>Proposed (([\tau_L, \tau_U]) = [0.5, 1])</td>
<td>Estimate</td>
<td>2.529</td>
<td>-0.066</td>
</tr>
<tr>
<td></td>
<td>95% CI</td>
<td>(1.920, 3.139)</td>
<td>(-0.089, -0.042)</td>
</tr>
<tr>
<td>GGR ((\tau = 0.25))</td>
<td>Estimate</td>
<td>2.553</td>
<td>0.083</td>
</tr>
<tr>
<td></td>
<td>95% CI</td>
<td>(0.885, 4.220)</td>
<td>(-0.310, 0.475)</td>
</tr>
<tr>
<td>GGR ((\tau = 0.5))</td>
<td>Estimate</td>
<td>2.452</td>
<td>-0.062</td>
</tr>
<tr>
<td></td>
<td>95% CI</td>
<td>(0.637, 4.267)</td>
<td>(-0.238, 0.114)</td>
</tr>
<tr>
<td>GGR ((\tau = 0.75))</td>
<td>Estimate</td>
<td>2.013</td>
<td>-0.009</td>
</tr>
<tr>
<td></td>
<td>95% CI</td>
<td>(-0.399, 4.425)</td>
<td>(-0.455, 0.437)</td>
</tr>
<tr>
<td>Composite - GGR (([0, 0.25]))</td>
<td>Estimate</td>
<td>2.469</td>
<td>0.033</td>
</tr>
<tr>
<td></td>
<td>95% CI</td>
<td>(2.312, 2.627)</td>
<td>(0.011, 0.055)</td>
</tr>
<tr>
<td>Composite - GGR (([0.375, 0.625]))</td>
<td>Estimate</td>
<td>2.142</td>
<td>-0.019</td>
</tr>
<tr>
<td></td>
<td>95% CI</td>
<td>(-3.032, 7.316)</td>
<td>(-0.626, 0.587)</td>
</tr>
<tr>
<td>Composite - GGR (([0.5, 1]))</td>
<td>Estimate</td>
<td>-0.045</td>
<td>1.146</td>
</tr>
<tr>
<td></td>
<td>95% CI</td>
<td>(-4.929, 4.839)</td>
<td>(-0.690, 2.982)</td>
</tr>
<tr>
<td>LY</td>
<td>Estimate</td>
<td>1.986</td>
<td>-0.096</td>
</tr>
<tr>
<td></td>
<td>(p)-value</td>
<td>0.028</td>
<td>0.003</td>
</tr>
<tr>
<td>Cox Model</td>
<td>Estimate</td>
<td>1.031</td>
<td>-0.055</td>
</tr>
<tr>
<td></td>
<td>(p)-value</td>
<td>0.033</td>
<td>0.015</td>
</tr>
</tbody>
</table>

*Approaches LY and GGR denote [38]’s and [18]’s proposals respectively.
parameter estimates of the proposed method roughly share consistent signs with those given by [38] and the Cox model. For the transplant status, the estimates produced by the proposed method are all significant across the three regions of quantiles, despite the difference in magnitude. The estimated coefficient for transplant status increases with the range of quantile considered, probably suggesting that a transplantation is more beneficial to those patients who live longer than their counterparts with similar covariate background. Meanwhile, the age at transplant has a significant negative influence on the failure time over all three regions of quantiles. For the mismatch score, the estimated coefficient flips from positive to negative and has a decreasing trend across quantile. It implies a transplant mismatch would be more unfavourable to the patients surviving longer after covariate adjustment, which again infers that a transplantation is more influential to that cohort. Such observations could be reasonable in nature, because it would take time for the transplantation to become effective. Comparing the estimated coefficients for mismatch score across the three regions, the difference can be an indication that the mismatch score having an insignificant overall effect over all quantile levels as estimated by [38].

2.7 Discussion

In view of this new perspective in survival analysis, further studies may also be of interest for generalizing the proposed methodology to accommodate different variations of failure time data in practice. Particularly, modification can be made without disrupting the framework of rank-based estimation procedure so as to handle length-biased data and case-cohort studies. On the other hand, following the exploration of quantile localized covariate effects, one may abandon the assumption of universal covariate effects but assume a hierarchical structure of the regression parameters instead. Being viewed as a hybrid between our setup and the censored quantile regression model, the current proposal can be generalized when the regression parameters are presumed to be with certain structure, especially provided prior knowledge on the covariate effects. Investigation of these new directions shall be handled in separate papers.
2.8 Appendix: Proofs

2.8.1 Consistency

Without any loss of generality, we shall deal with the scenario when \([\tau_L, \tau_U] = [0, \tau]\) in this and the following appendices. The case of general \([\tau_L, \tau_U]\) is similar and follows directly. Accordingly, if there is no ambiguity, we will suppress the first zero for the sake of notation simplicity, i.e. for example, we shall write \(U_n(\beta; \tau)\) instead of \(U_n(\beta; 0, \tau)\).

Denote \(dM_i(t; \beta) = dN_i(t; \beta) - Y_i(t; \beta) d\Lambda(t; \beta)\), then using simple algebra, it is easy to verify that

\[
U_n(\beta; \tau) = \frac{1}{n} \sum_{i=1}^{n} \int_{0}^{q(\tau; \beta)} \Psi(t; \beta) \left\{ Z_i(h_i^{-1}(t; \beta)) - \frac{S(1)(t; \beta)}{S(0)(t; \beta)} \right\} dN_i(t; \beta)
= \frac{1}{n} \sum_{i=1}^{n} \int_{0}^{q(\tau; \beta)} \Psi(t; \beta) \left\{ Z_i(h_i^{-1}(t; \beta)) - \frac{S(1)(t; \beta)}{S(0)(t; \beta)} \right\} dM_i(t; \beta).
\]

We shall show that \(U_n\) and \(\tilde{U}_n\) are asymptotically close in some sense, where

\[
\tilde{U}_n(\beta; \tau) = \frac{1}{n} \sum_{i=1}^{n} \int_{0}^{q(\tau; \beta)} \psi(t; \beta) \left\{ Z_i(h_i^{-1}(t; \beta)) - \frac{s(1)(t; \beta)}{s(0)(t; \beta)} \right\} dM_i(t; \beta).
\]

In particular, we will claim the following convergence results, i.e.

\[
U_n(\beta_0; \tau) - \tilde{U}_n(\beta_0; \tau) = o(1) \quad a.s., \tag{2.5}
\]

\[
U_n(\hat{\beta}; \tau) - \tilde{U}_n(\hat{\beta}; \tau) = o(1 + \|\hat{\beta} - \beta_0\|) \quad a.s., \tag{2.6}
\]

\[
n^{1/2}U_n(\beta_0; \tau) - n^{1/2}\tilde{U}_n(\beta_0; \tau) = o_p(1), \tag{2.7}
\]

\[
n^{1/2}U_n(\hat{\beta}; \tau) - n^{1/2}\tilde{U}_n(\hat{\beta}; \tau) = o_p(1 + n^{1/2}\|\hat{\beta} - \beta_0\|). \tag{2.8}
\]

Equations (2.5) and (2.6) are used to establish the consistency while (2.7) and (2.8) are for the weak convergence of \(\hat{\beta}\). Note the subtle difference between the two pairs, i.e. the convergence results hold in probability instead of almost surely when there is an additional factor of \(n^{1/2}\). As the term
$n^{1/2}$ is included, the contrast arises because we need to exploit the strong representation theorem (Theorem 9.4 of [42]) on some weak convergence results, so that the convergence holds almost surely in another probability space. As a result, the resulting claims will only hold in probability in the original probability space, which, however, are still sufficient to prove the weak convergence of $\hat{\beta}$.

To this end, we first claim that

$$\sup_{t \geq 0, \beta \in \mathcal{B}} \left| S^{(0)}(t; \beta) - s^{(0)}(t; \beta) \right| = o(1) \quad a.s., \quad \sup_{t \geq 0, \beta \in \mathcal{B}} \left\| S^{(1)}(t; \beta) - s^{(1)}(t; \beta) \right\| = o(1) \quad a.s.,$$

(2.9)

where $s^{(0)}(t; \beta) = E[Y(t; \beta)]$ and $s^{(1)}(t; \beta) = E[Z(h^{-1}(t; \beta))Y(t; \beta)]$. Recall $S^{(0)}(t; \beta) = n^{-1} \sum_{i=1}^{n} Y_i(t; \beta)$ and $S^{(1)}(t; \beta) = n^{-1} \sum_{i=1}^{n} Z_i(h^{-1}_i(t; \beta))Y_i(t; \beta)$, it suffices to show $\{Y_i(t; \beta)\}$ and $\{Z_i(h^{-1}_i(t; \beta))Y_i(t; \beta)\}$ are manageable, and then apply Theorem 8.3 of [42] to show (2.9). Accordingly, note that indicator function and monotone function have pseudo-dimension one and hence manageable (see for example, Section 13 of [42] or Section 11.4 of [31]), therefore $\{Y_i(t; \beta)\}$ is manageable because it is an indicator. Also, since $Z_i(t)$ is of finite variation, it can be decomposed into $Z_i(t) = Z^+_i(t) - Z^-_i(t)$, where $Z^+_i(t)$ is non-negative and monotone increasing. Write

$$Z_i(h^{-1}_i(t; \beta))Y_i(t; \beta) = Z^+_i(h^{-1}_i(t; \beta))I(\bar{T}_i \geq h^{-1}_i(t; \beta)) - Z^-_i(h^{-1}_i(t; \beta))I(\bar{T}_i \geq h^{-1}_i(t; \beta)),$$

and note that the class $\{Z^+_i(t)I(\bar{T}_i \geq t) - Z^-_i(t)I(\bar{T}_i \geq t)\}$ is manageable because it consists of indicator functions and monotone functions in $t$. By Lemma 5.1 of [42], we conclude that $\{Z_i(h^{-1}_i(t; \beta))Y_i(t; \beta)\}$ has finite pseudo-dimension and hence manageable, therefore (2.9) holds. Also, note that Condition 2.3 implies $s^{(0)}(t; \beta)$ is uniformly bounded away from zero for large enough $n$. Using this fact and (2.9), we have

$$\sup_{t \geq 0, \beta \in \mathcal{B}} \left\| S^{(1)}(t; \beta) - S^{(0)}(t; \beta) \right\| = o(1) \quad a.s.$$  

(2.10)

Let $U_n^{(0)}(t) = n^{-1} \sum_{i=1}^{n} M_i(t; \beta_0)$ and $U_n^{(1)}(t) = n^{-1} \sum_{i=1}^{n} \int_{0}^{t} Z_i(h^{-1}_i(u; \beta_0))dM_i(u; \beta_0)$, then the
two processes are mean zero because of equations (2.2) and (2.3) respectively. Using the uniform strong law of large numbers (Theorem 8.3 of [42]), both \( U_n^{(0)} \) and \( U_n^{(1)} \) uniformly converge to zero almost surely. Indeed, we claim that \( n^{1/2}U_n^{(0)} \) and \( n^{1/2}U_n^{(1)} \) converge weakly to two mean zero Gaussian processes \( \{G_0(t), t \in [0, q(\tau)]\} \) and \( \{G_1(t), t \in [0, q(\tau)]\} \) respectively, with covariance functions

\[
E\{G_0(s)G_0^\top(t)\} = \int_0^{sn\Lambda} s^{(0)}(u; \beta_0)d\Lambda(u; \beta_0), \quad \text{and} \quad E\{G_1(s)G_1^\top(t)\} = \int_0^{sn\Lambda} s^{(2)}(u; \beta_0)d\Lambda(u; \beta_0),
\]

where \( s^{(2)}(t; \beta) = E[Z(h^{-1}(t; \beta))Z^\top(h^{-1}(t; \beta))Y(t; \beta)] \). This can be accomplished by invoking the function central limit theorem (Theorem 10.7 of [42] or Theorem 11.16 of [31]). We shall only check the first and the last conditions, i.e. manageability and equicontinuity, as the others are straightforward. Also, since \( U_n^{(0)} \) is a special case of \( U_n^{(1)} \) when \( Z_i(t) \equiv 1 \), it suffices to prove the result for \( U_n^{(1)} \) only. For the equicontinuity condition, define \( \rho_n(s, t) = E[\|n^{1/2}U_n^{(1)}(s) - n^{1/2}U_n^{(1)}(t)\|^2] \) and \( \rho(s, t) = E[\|n^{1/2}G_1(s) - n^{1/2}G_1(t)\|^2] \). We have

\[
\rho_n(s, t) \leq \frac{1}{n} \sum_{i=1}^n E \left[ \left\| \int_{h^{-1}(s; \beta_0)}^{h^{-1}(t; \beta_0)} Z_i(u)[d\Delta_i(\hat{T}_i \leq u) - I(\hat{T}_i \geq u)]d\Lambda(u) \right\|^2 \right]
\]

\[
= \frac{1}{n} \sum_{i=1}^n E \left[ \left\| Z_i(u) \right\|^2 I(\hat{T}_i \geq u) d\Lambda(u) \right],
\]

so \( \rho_n \) is equicontinuous. Also, it is not difficult to see that \( \rho_n \) converges pointwise to \( \rho \), therefore \( \rho_n \) converges uniformly to \( \rho \) because of its equicontinuity. It follows that if \( s_n \) and \( t_n \) are two sequences such that \( \rho(s_n, t_n) \to 0 \), then we have \( \rho_n(s_n, t_n) \to 0 \), so the equicontinuity condition holds. For the manageability condition, write

\[
U_n^{(1)}(t) = \frac{1}{n} \sum_{i=1}^n \left[ \int_0^t Z_i(h^{-1}_i(u; \beta_0))dN_i(u; \beta_0) - \int_0^t Z_i(h^{-1}_i(u; \beta_0))Y_i(u; \beta_0)d\Lambda(u; \beta_0) \right]
\]

\[
= \frac{1}{n} \sum_{i=1}^n \left[ Z_i(\hat{T}_i)N_i(t; \beta_0) - \int_0^t Z_i(h^{-1}_i(u; \beta_0))Y_i(u; \beta_0)d\Lambda(u; \beta_0) \right],
\]

from which it suffices to check \( \{Z_i(\hat{T}_i)N_i(t; \beta_0)\} \) and \( \{\int_0^t Z_i(h^{-1}_i(u; \beta_0))Y_i(u; \beta_0)d\Lambda(u; \beta_0)\} \) are
both manageable. But this is obvious, since we might assume $Z_i(t) \geq 0$ without loss of generality because $Z_i(t) = Z_i^+(t) - Z_i^-(t)$, so both of the above processes are monotone in $t$, which has pseudo-dimension 1 and hence manageable. Now, we have the weak convergence of $n^{1/2}U_n^{(0)}$ and $n^{1/2}U_n^{(1)}$ by the functional central limit theorem.

On the other hand, we show that $\sup_{\beta \in B} |\hat{q}(\tau; \beta) - q(\tau; \beta)| = o(1)$ almost surely. Using the results in Section 12.2.2 of [31], the Nelson-Aalen estimator of the marginal distribution of $h(T; \beta)$, denoted by $\hat{\Lambda}(t; \beta)$, is the integration map $(A, B) \mapsto \int_0^1 \frac{1}{\hat{r}} dA$ of $(\hat{N}(t; \beta), \hat{Y}(t; \beta)) = (n^{-1} \sum_{i=1}^n N_i(t; \beta), n^{-1} \sum_{i=1}^n Y_i(t; \beta))$. This map is Hadamard differentiable with a derivative of $(A, B) \mapsto \int \frac{1}{\hat{E}[Y(t; \beta)]} \{dA - B d\Lambda(t; \beta)\}$. As shown earlier, the classes $\{N_i(t; \beta)\}$ and $\{Y_i(t; \beta)\}$ are both manageable, so we have

$$\sup_{t \geq 0, \beta \in B} |\hat{N}(t; \beta) - E[N(t; \beta)]| = o(1) \quad a.s., \quad \sup_{t \geq 0, \beta \in B} |\hat{Y}(t; \beta) - E[Y(t; \beta)]| = o(1) \quad a.s.$$ 

It follows that

$$\hat{\Lambda}(t; \beta) - \Lambda(t; \beta) = \int_0^t \frac{1}{\hat{E}[Y(u; \beta)]} \left[ d\left\{\hat{N}(u; \beta) - E[N(u; \beta)]\right\} - \left\{\hat{Y}(u; \beta) - E[Y(u; \beta)]\right\} d\Lambda(u; \beta) \right].$$

Using Condition 2.3 and the uniform consistency of $\hat{N}(t; \beta)$ and $\hat{Y}(t; \beta)$, we can conclude that $\sup_{t \geq 0, \beta \in B} |\hat{\Lambda}(t; \beta) - \Lambda(t; \beta)| = o(1)$ almost surely. As an immediate consequence of the continuity of product integrals (Proposition II.8.7 of [4]), we have the uniform consistency of the Kaplan-Meier estimator, i.e. $\sup_{t \geq 0, \beta \in B} |\hat{F}(t; \beta) - F(t; \beta)| = o(1)$ almost surely. Finally, by the Hadamard differentiability of the inversion map (Section 12.2.4 of [31]), we arrive at

$$\sup_{\beta \in B} |\hat{q}(\tau; \beta) - q(\tau; \beta)| = \sup_{\beta \in B} \left| \frac{1}{\hat{f}\{q(\tau; \beta); \beta\}} \left[ \hat{f}\{q(\tau; \beta); \beta\} - \tau \right] \right| = o(1) \quad a.s. \quad (2.11)$$

A sufficient condition for the last equality is that $f\{q(\tau; \beta); \beta\}$ is uniformly bounded away from zero, which follows from Condition 2.4.
With (2.10), Condition 2.5 and the uniform consistency of \( U_n^{(0)} \) and \( U_n^{(1)} \), we can conclude via Lemma 2.8.3 of [5] that almost surely,

\[
\sup_{t \in [0,q(\tau;\beta_0)]} \left\| \frac{1}{n} \sum_{i=1}^{n} \int_0^t \{ \Psi(u;\beta_0) - \psi(u;\beta_0) \} Z_i(h_i^{-1}(u;\beta_0))dM_i(u;\beta_0) \right\| = o(1),
\]

\[
\sup_{t \in [0,q(\tau;\beta_0)]} \left\| \frac{1}{n} \sum_{i=1}^{n} \int_0^t \{ \Psi(u;\beta_0) \frac{S^{(1)}(u;\beta_0)}{S(0)(u;\beta_0)} - \psi(u;\beta_0) \frac{s^{(1)}(u;\beta_0)}{s(0)(u;\beta_0)} \} dM_i(u;\beta_0) \right\| = o(1).
\]

As we have shown that \( \sup_{\beta \in \mathcal{B}} |\hat{q}(\tau;\beta) - q(\tau;\beta)| = o(1) \) almost surely, it follows that \( U_n(\beta_0;\tau) - \hat{U}_n(\beta_0;\tau) = o(1) \) almost surely, i.e. (2.5). On the other hand, instead of using the uniform consistency, we could exploit the weak convergence of \( n^{1/2}U_n^{(0)} \) and \( n^{1/2}U_n^{(1)} \) shown above. The convergence holds almost surely in another probability space via the strong representation theorem (Theorem 9.4 of [42]). Therefore, in analogy to deriving the last display, we have

\[
\sup_{t \in [0,q(\tau;\beta_0)]} \left\| \frac{n^{1/2}}{n} \sum_{i=1}^{n} \int_0^t \{ \Psi(u;\beta_0) - \psi(u;\beta_0) \} Z_i(h_i^{-1}(u;\beta_0))dM_i(u;\beta_0) \right\| = o_p(1),
\]

\[
\sup_{t \in [0,q(\tau;\beta_0)]} \left\| \frac{n^{1/2}}{n} \sum_{i=1}^{n} \int_0^t \{ \Psi(u;\beta_0) \frac{S^{(1)}(u;\beta_0)}{S(0)(u;\beta_0)} - \psi(u;\beta_0) \frac{s^{(1)}(u;\beta_0)}{s(0)(u;\beta_0)} \} dM_i(u;\beta_0) \right\| = o_p(1).
\]

Note that we only have convergence in probability here in the original probability space, so we have \( n^{1/2}U_n(\beta_0;\tau) - n^{1/2}\hat{U}_n(\beta_0;\tau) = o_p(1) \), i.e. (2.7).

Now, using the law of large numbers and the multivariate central limit theorem, we know \( \hat{U}_n(\beta_0;\tau) = o(1) \) almost surely and \( n^{1/2}\hat{U}_n(\beta_0;\tau) \) converges to a mean zero multivariate normal distribution with a covariance matrix

\[
V(\beta_0;\tau) = \int_0^{q(\tau;\beta_0)} \psi^2(t;\beta_0) \left\{ s^{(2)}(t;\beta_0) - \frac{s^{(1)}(t;\beta_0)s^{(1)}(t;\beta_0)^\top}{s(0)(t;\beta_0)} \right\} d\Lambda(t;\beta_0).
\]

Meanwhile, consider \( n^{1/2}\hat{U}_n(\beta;\tau) = n^{1/2}U_n(\hat{\beta};\tau) - \{ n^{1/2}U_n(\hat{\beta};\tau) - n^{1/2}\hat{U}_n(\hat{\beta};\tau) \} \), and note
that $U_n(\hat{\beta}; \tau) = O(n^{-1})$ almost surely by construction. We now decompose

$$U_n(\hat{\beta}; \tau) - \tilde{U}_n(\hat{\beta}; \tau)$$

$$= \frac{1}{n} \sum_{i=1}^{n} \int_0^1 \psi(t; \hat{\beta}) \left\{ \mathbf{Z}_i(\hat{h}_i^{-1}(t; \hat{\beta})) - \frac{S^{(1)}(t; \hat{\beta})}{S^{(0)}(t; \hat{\beta})} \right\} dM_i(t; \hat{\beta})$$

$$- \frac{1}{n} \sum_{i=1}^{n} \int_0^1 \psi(t; \hat{\beta}) \left\{ \mathbf{Z}_i(\hat{h}_i^{-1}(t; \hat{\beta})) - \frac{S^{(1)}(t; \hat{\beta})}{S^{(0)}(t; \hat{\beta})} \right\} dM_i(t; \hat{\beta})$$

$$= \frac{1}{n} \sum_{i=1}^{n} \int_0^1 \hat{q}(t; \hat{\beta}) \psi(t; \hat{\beta}) \left\{ \mathbf{Z}_i(\hat{h}_i^{-1}(t; \hat{\beta})) - \frac{S^{(1)}(t; \hat{\beta})}{S^{(0)}(t; \hat{\beta})} \right\} d\left\{ M_i(t; \hat{\beta}) - E[M_i(t; \hat{\beta})|\mathbf{Z}_i] \right\}$$

$$- \frac{1}{n} \sum_{i=1}^{n} \int_0^1 \hat{q}(t; \hat{\beta}) \psi(t; \hat{\beta}) \left\{ \mathbf{Z}_i(\hat{h}_i^{-1}(t; \hat{\beta})) - \frac{S^{(1)}(t; \hat{\beta})}{S^{(0)}(t; \hat{\beta})} \right\} d\left\{ M_i(t; \hat{\beta}) - E[M_i(t; \hat{\beta})|\mathbf{Z}_i] \right\}$$

$$+ \frac{1}{n} \sum_{i=1}^{n} \int_0^1 \hat{q}(t; \hat{\beta}) \psi(t; \hat{\beta}) \left\{ \mathbf{Z}_i(\hat{h}_i^{-1}(t; \hat{\beta})) - \frac{S^{(1)}(t; \hat{\beta})}{S^{(0)}(t; \hat{\beta})} \right\} d\left\{ E[M_i(t; \hat{\beta})|\mathbf{Z}_i] - E[M_i(t; \beta_0)|\mathbf{Z}_i] \right\}$$

$$- \frac{1}{n} \sum_{i=1}^{n} \int_0^1 \hat{q}(t; \hat{\beta}) \psi(t; \hat{\beta}) \left\{ \mathbf{Z}_i(\hat{h}_i^{-1}(t; \hat{\beta})) - \frac{S^{(1)}(t; \hat{\beta})}{S^{(0)}(t; \hat{\beta})} \right\} d\left\{ E[M_i(t; \hat{\beta})|\mathbf{Z}_i] - E[M_i(t; \beta_0)|\mathbf{Z}_i] \right\}$$

$$= (I) + (II).$$

Analogous to the procedure in showing (2.5) and (2.7), i.e. by (2.10) and (2.11), Condition 2.5 and the use of Lemma 2.8.3 in [5], we could conclude that $(I) = o(1)$ almost surely. Coupled with the strong representation theorem, we also have $n^{1/2} \cdot (I) = o_p(1)$. As for (II), via a direct calculation, it could be shown that $E\{M_i(t; \hat{\beta}) - M_i(t; \beta_0)|\mathbf{Z}_i\} = t \cdot O(||\hat{\beta} - \beta_0||)$ almost surely. As the process $\{t\}$ is of finite total variation over the bounded interval $[0, q(\tau; \hat{\beta})]$, thus by Lemma 2.8.3 of [5], we have almost surely

$$(II) = O \left( \int_0^1 \hat{q}(t; \hat{\beta}) \psi(t; \hat{\beta}) \left\{ \mathbf{Z}_i(\hat{h}_i^{-1}(t; \hat{\beta})) - \frac{S^{(1)}(t; \hat{\beta})}{S^{(0)}(t; \hat{\beta})} \right\} dt \right)$$

$$= o(||\hat{\beta} - \beta_0||).$$

Combining the results of (I) and (II), we have $U_n(\hat{\beta}; \tau) - \tilde{U}_n(\hat{\beta}; \tau) = o(1 + ||\hat{\beta} - \beta_0||)$ almost
surely, i.e. (2.6), and \( n^{1/2} U_n(\hat{\beta}; \tau) - n^{1/2} \tilde{U}_n(\hat{\beta}; \tau) = o_p(1 + n^{1/2} \| \hat{\beta} - \beta_0 \|) \), i.e. (2.8).

As \( A_n(\beta_0; \tau) \) is invertible and \( A_n(\beta; \tau) \) has a uniformly bounded derivative with respect to \( \beta \), so there exists a neighbourhood of \( \beta_0 \), denoted by \( \mathcal{N}(\beta_0) \), such that \( A_n(\beta; \tau) \) is invertible for \( \beta \in \mathcal{N}(\beta_0) \). As a result, for any \( \hat{\beta} \in \mathcal{N}(\beta_0) \), there is an invertible \( A_n(\beta'; \tau) \) where \( \beta' \in \mathcal{N}(\beta_0) \) is in between \( \beta_0 \) and \( \hat{\beta} \), so that almost surely,

\[
\| A_n(\beta'; \tau)(\hat{\beta} - \beta_0) \| = \| \tilde{U}_n(\hat{\beta}; \tau) - \tilde{U}_n(\beta_0; \tau) \| \\
= \| U_n(\hat{\beta}; \tau) - \tilde{U}_n(\beta_0; \tau) - \{ U_n(\hat{\beta}; \tau) - \tilde{U}_n(\hat{\beta}; \tau) \} \| \\
= O(n^{-1}) + o(1) + o(1 + \| \hat{\beta} - \beta_0 \|) = o(1 + \| \hat{\beta} - \beta_0 \|).
\]

Hence we have \( \| \hat{\beta} - \beta_0 \| \to 0 \) almost surely.

### 2.8.2 Weak Convergence

As we now know \( \hat{\beta} \) converges to \( \beta_0 \) almost surely (and hence in probability), we have

\[
\{ A_n(\beta_0; \tau) + o_p(1) \} n^{1/2} (\hat{\beta} - \beta_0) \\
= n^{1/2} \{ \tilde{U}_n(\hat{\beta}; \tau) - \tilde{U}_n(\beta_0; \tau) \} \\
= n^{1/2} U_n(\hat{\beta}; \tau) - n^{1/2} \tilde{U}_n(\beta_0; \tau) - \{ n^{1/2} U_n(\hat{\beta}; \tau) - n^{1/2} \tilde{U}_n(\hat{\beta}; \tau) \} \\
= O_p(n^{-1/2}) - n^{1/2} U_n(\beta_0; \tau) + o_p(1 + n^{1/2} \| \hat{\beta} - \beta_0 \|) \\
= -n^{1/2} \tilde{U}_n(\beta_0; \tau) + o_p(1 + n^{1/2} \| \hat{\beta} - \beta_0 \|).
\]

As we have claimed above that \( n^{1/2} \tilde{U}_n(\beta_0; \tau) \) converges to a mean zero multivariate normal with covariance matrix \( V(\beta_0; \tau) \), it follows that \( n^{1/2}(\hat{\beta} - \beta_0) \) converges to a mean zero multivariate normal with covariance matrix \( A(\beta_0; \tau)^{-1} V(\beta_0; \tau) A(\beta_0; \tau)^{-1} \).
2.8.3 Justification of Resampling Method

Following the arguments in Section 2.8.1 and Section 2.8.2, and since \( E(\xi_i) = 1 \), we can similarly obtain that

\[
\{\mathbf{A}_n^*(\beta_0; \tau) + o_p(1)\} n^{1/2}(\beta^* - \hat{\beta}) = -n^{1/2} \bar{U}_n^*(\hat{\beta}; \tau) + o_p(1 + n^{1/2}\|\beta^* - \hat{\beta}\|),
\]

where \( \bar{U}_n^* \) and \( \mathbf{A}_n^* \) are respectively defined in analogy to \( \bar{U}_n \) and \( \mathbf{A}_n \). Now, we have \( n^{1/2} \bar{U}_n(\hat{\beta}; \tau) = o_p(1 + n^{1/2}\|\hat{\beta} - \beta_0\|) \)

\[
-n^{1/2} \bar{U}_n^*(\hat{\beta}; \tau) = n^{1/2} \{ \bar{U}_n(\hat{\beta}; \tau) - \bar{U}_n^*(\hat{\beta}; \tau) \} + n^{1/2} \{ U_n(\hat{\beta}; \tau) - U_n(\hat{\beta}; \tau) \} - n^{1/2} U_n(\hat{\beta}; \tau)
\]

\[
= n^{1/2} \{ \bar{U}_n(\hat{\beta}; \tau) - \bar{U}_n^*(\hat{\beta}; \tau) \} + o_p(1 + n^{1/2}\|\hat{\beta} - \beta_0\|) + O_p(n^{-1/2})
\]

\[
= n^{1/2} \frac{1}{n} \sum_{i=1}^{n} (1 - \xi_i) \kappa_i(\hat{\beta}; \tau) + o_p(1 + n^{1/2}\|\hat{\beta} - \beta_0\|),
\]

where \( \kappa_i(\beta; \tau) = \int q(\tau; \beta) \psi(t; \beta) \left\{ Z_i(h_i^1(t; \beta)) - \frac{r_i(t; \beta)}{s_i(t; \beta)} \right\} dM_i(t; \beta) \). Also, since \( \text{Var}(\xi_i) = 1 \), we have

\[
E(\bar{U}_n^*(\hat{\beta}; \tau) \bar{U}_n^*(\hat{\beta}; \tau)^\top \mid \{ \bar{T}_i, \Delta_i, \bar{Z}_i(\bar{T}_i) \}_{i=1}^n) = \frac{1}{n} \sum_{i=1}^{n} \kappa_i(\hat{\beta}; \tau) \kappa_i(\hat{\beta}; \tau)^\top \overset{\text{p}}{\rightarrow} \mathbf{V}(\beta_0; \tau).
\]

As argued in [36] and [41], \(-n^{1/2} \bar{U}_n^*(\hat{\beta}; \tau) \) converges to the same limiting distribution as that of \(-n^{1/2} \bar{U}_n(\beta_0; \tau) \) for almost all realizations of the data \( \{ \bar{T}_i, \Delta_i, \bar{Z}_i(\bar{T}_i) \}_{i=1}^n \). It follows that given the observed data \( \{ \bar{T}_i, \Delta_i, \bar{Z}_i(\bar{T}_i) \}_{i=1}^n \), the conditional distribution of \( n^{1/2}(\beta^* - \hat{\beta}) \) is asymptotically equivalent to the unconditional distribution of \( n^{1/2}(\hat{\beta} - \beta_0) \).
Chapter 3: Censored Quantile Regression with Time-dependent Covariates

3.1 Introduction

Since its introduction in the seminal paper of [29], quantile regression has become a widely adopted technique for modeling the conditional distribution of the outcome variable given covariates. In contrast to the classical ordinary least squares regression which relates the conditional mean of the response to the predictors, quantile regression models the conditional quantile of the outcome given the covariates, and therefore it can offer a more holistic description of the covariate effects on the response variable. Because of its greater model flexibility across different quantile levels, quantile regression has been widely adopted in various fields, such as economics ([30, 13]), environment modeling ([20, 50]), growth chart ([56, 55]), and in particular the modeling of failure time data. In the context of survival analysis, quantile regression is a valuable alternative to other semiparametric methods, including Cox proportional hazards model ([10, 11]) and accelerated failure time model ([52, 59]). Instead of focusing on the modelling of the hazard function concerned in Cox model, quantile regression examines the conditional quantile function of the failure time directly, hence leading to a more straightforward interpretation of the covariate effects. More importantly, quantile regression provides flexibility in accommodating heterogeneous effects of the covariates while such heterogeneity cannot be easily incorporated into the Cox model or the accelerated failure time model.

Because of these merits, censored quantile regression has attracted considerable attention in the literature. Assuming the censoring time is independent of the failure time and the covariates, [58] derived a median regression procedure which can also be readily generalized for any other arbitrary quantile level. Nevertheless, the more standard notion adopted in various literature on survival analysis, including [10, 11] and [6], is that the failure time and the censoring time are
conditionally independent given the covariates. In view of this, [43] and [53] relaxed the aforementioned marginally independent censoring assumption by exploiting the [14]’s redistribution to the right algorithm of the Kaplan-Meier estimator and proposed a locally weighted censored quantile regression model. Another important development is due to [41], which aptly connects the quantile process of interest with the naturally adopted counting process formulation. They derived a martingale based estimating equation in which the Nelson-Aalen type compensator is expressed in terms of a sequential sum of functions based on previously estimated quantile values. [34] relaxed the assumption made in the above framework that the linear quantile model holds globally for all quantiles by incorporating a kernel based estimator on the conditional survival function. Meanwhile, [7] considered an alternative approach by incorporating a transformation process into the quantile function, so that the global linear quantile model can be generalized.

While there have been various solutions proposed for censored quantile regression, most of which, however, are designed for time-independent covariates only. Nevertheless, time-dependent covariates appear naturally in many real applications as records of the time-dependent attributes of the subjects, including, for example, the blood pressure of the patients admitted, or the treatment status in a clinical follow-up. There have been research efforts on generalizing existing techniques for censored data to accommodate time-dependent covariate process. In particular, the hazard based Cox model can be readily generalized to accommodate time-dependent covariates; see, for example, [15] and [26]. Another line of research is the accelerated failure time model, which is generalized in [47] and [38] to allow for time-dependent covariates. For censored quantile regression model with time-dependent covariates, [18] proposed a procedure via the inverse probability weighting technique based on the independent censoring assumption. More recently, [22] adopted [27]’s formulation for censored data so that the panel covariates are reduced into time-independent proxies with spline approximation. However, apart from the use of splines that commonly results in extra effort and approximation error for estimation, their estimation requires that the entire covariate process has to be observable, which may not be possible in practice.

In this paper, we discuss a censored quantile regression model with time-dependent covari-
ates under the conditionally independent censoring assumption. Our contributions can be summarized in the following aspects. First, our model construction allows the discussion of internal time-dependent covariates, which have not been covered in existing quantile regression methods mentioned above. Also, our formulation of the quantile function is in line with those given in [38] and [18], extending [41]’s censored quantile regression model naturally. As a result, our algorithm requires the time-dependent covariate history only up to the observed survival time, which tremendously improves its applicability in terms of handling different types of censored data. Finally, our estimation procedure is built via a series of martingale based estimating equations, which only assumes the standard conditionally independent censorship, as opposed to the marginally independent censorship in existing alternatives.

The following of this paper is organized as follows: Section 3.2 outlines the model framework and describes the estimation procedure for the model parameters, while Section 3.3 outlines the asymptotic results and a perturbation based resampling method. Section 3.4 presents some numerical results based on Monte Carlo simulation, followed by analysis of the ACTG 320 dataset in Section 3.5. Section 3.6 concludes the paper, while all proofs of the asymptotic results and justification of the resampling method are postponed to the appendix, i.e. Section 3.7.

3.2 Methodology

Suppose $T$ is the failure time, $C$ is the censoring time, $\{Z(t), t \geq 0\}$ is a $p$-dimensional multivariate random process of covariates, $\beta_0(\tau)$ is a $p$-dimensional process of parameters indexed by $\tau$, where $\tau \in (0, 1)$ is the quantile level. Let $\hat{Z}(t) = \{Z(s), s \in [0, t]\}$ denote the covariate history up to time $t$, and write $\hat{Z}(\infty) = \{Z(s), s \geq 0\}$ as $\hat{Z}$. Consider the natural filtration $\mathcal{F}_t = \sigma\{I(T \leq s), Z(s), s \in [0, t]\}$, where $I(\cdot)$ is the indicator function. Note that $T$ is a stopping time with respect to the filtration $\{\mathcal{F}_t\}_{t \geq 0}$, hence the stopped sigma algebra is $\mathcal{F}_T = \{A \in \mathcal{F}_\infty : A \cap \{T \leq t\} \in \mathcal{F}_t, \forall t \geq 0\}$.

Define the conditional hazard function of $T$ given the covariate history at time $t$ as $\lambda_T(t \mid \hat{Z}(t))dt = P\{T \in [t, t + dt] \mid T \geq t, \hat{Z}(t)\}$, then we let $\Lambda_T(t \mid \hat{Z}(t)) = \int_0^t \lambda_T(s \mid \hat{Z}(s))ds$ and
\( F_T \{ t \mid \tilde{Z}(t) \} = 1 - \exp[\Lambda_T \{ t \mid \tilde{Z}(t) \} ] \). Let \( Q_T(\tau \mid \tilde{Z}) = \inf \{ t : F_T \{ t \mid \tilde{Z}(t) \} \geq \tau \} \) be the inverse function of \( F_T \{ t \mid \tilde{Z}(t) \} \), then \( Q_T(\tau \mid \tilde{Z}) \in \mathcal{F}_\infty \) is a stopping time with respect to \( \{ \mathcal{F}_t \}_{t \geq 0} \) because 
\( \{ Q_T(\tau \mid \tilde{Z}) \leq t \} = \{ F_T \{ t \mid \tilde{Z}(t) \} \geq \tau \} \in \mathcal{F}_t \) for all \( t \geq 0 \). Moreover, we can show that \( Q_T(\tau \mid \tilde{Z}) \) is \( \mathcal{F}_T \) measurable, which is done by verifying \( \{ Q_T(\tau \mid \tilde{Z}) \leq s \} \cap \{ T \leq t \} \in \mathcal{F}_t \) for all \( s, t \geq 0 \). This claim is true because

\[
\{ Q_T(\tau \mid \tilde{Z}) \leq s \} \cap \{ T \leq t \} = \{ F_T \{ s \mid \tilde{Z}(s) \} \geq \tau, T \leq t \} = \{ F_T \{ s \wedge t \mid \tilde{Z}(s \wedge t) \} \geq \tau, T \leq t \} \in \mathcal{F}_t,
\]

where \( s \wedge t \) denotes the minimum of \( s \) and \( t \). The last equality follows because on the set \( \{ T \leq t \} \) and for \( s > t \), we have \( \lambda_T \{ u \mid \tilde{Z}(u) \} = 0 \) for \( u \in [t, s] \) by definition, therefore \( F_T \{ s \mid \tilde{Z}(s) \} = F_T \{ t \mid \tilde{Z}(t) \} \). By characterizing \( Q_T(\tau \mid \tilde{Z}) \) in this regard, its \( \mathcal{F}_T \) measurability implies that we can determine \( Q_T(\tau \mid \tilde{Z}) \) given the information at the failure time \( T \). We call \( Q_T(\tau \mid \tilde{Z}) \) the (generalized) conditional quantile function of the failure time given the time-dependent covariates.

**Remark 3.1.** Recall from [26] that a time-dependent covariate \( Z(t) \) is called external when

\[
\lambda_T \{ t \mid \tilde{Z}(t) \} dt = P \{ T \in [t, t + dt] \mid T \geq t, \tilde{Z}(t) \} = P \{ T \in [t, t + dt] \mid T \geq t, \tilde{Z} \}.
\]

In other words, the future path of the covariate process has no impacts on the current hazard rate. If \( Z(t) \) consists of external covariates only, then \( F_T \{ t \mid \tilde{Z}(t) \} \) defined above coincides with the conditional distribution function of \( T \) given \( \tilde{Z} \), denoted by \( F_T(t \mid \tilde{Z}) \equiv P(T \leq t \mid \tilde{Z}) \). Indeed, the path of external covariate is not affected by the failure status, so we can image and condition on the entire covariate process. Accordingly, when only external covariates are involved, \( Q_T(\tau \mid \tilde{Z}) \) recovers the \( \tau \)-th conditional quantile of \( T \) given \( \tilde{Z} \). In particular, the quantile models with time-dependent covariates discussed in existing literature such as [18] and [22] are referring to this construction.

However, the previous interpretation of \( F_T \{ t \mid \tilde{Z}(t) \} \) is not straightforward to cases where
some $\tilde{Z}(t)$ are internal. Nevertheless, it can be shown that \cite{48}

$$P\{T \leq t \mid Z(0)\} = E_Z \left[ 1 - \exp \left\{ -\int_0^t \lambda_T(s \mid \tilde{Z}(s)) \, ds \right\} \right] = E[Z[F_T(t \mid \tilde{Z}(t))].$$

where the expectation $E_Z(\cdot)$ is taken with respect to the conditional distribution of the future covariate process given the survival and prior covariate history. Substitute $t = Q_T(\tau \mid \tilde{Z})$ and recall that $Q_T(\tau \mid \tilde{Z})$ is the inverse of $F_T(t \mid \tilde{Z}(t))$, we have

$$P\{T \leq Q_T(\tau \mid \tilde{Z}) \mid Z(0)\} = E_Z(\tau) = \tau,$$

for any realization of $Q_T(\tau \mid \tilde{Z})$, i.e. the event $\{T \leq Q_T(\tau \mid \tilde{Z})\}$ happens with probability $\tau$ when averaging over all possible failure times and future covariate paths. Indeed, unless we model the joint distribution of $T$ and $\tilde{Z}$, the interplay of the failure time and the internal covariates renders the usual definition of quantile unmeaningful. But with the above observation, $Q_T(\tau \mid \tilde{Z})$ can be interpreted as a generalized version of the conventional quantile function which satisfies the last display for all possible paths of $\tilde{Z}$.

Let $h(t; b) = \int_0^t \exp\{-Z(s)^Tb\} \, ds$ be a monotone transform of $t$ given $b$, then $h(t; b) \equiv t$ when $Z(t) \equiv 0$, so $h(t; b)$ can be regarded as a baseline transform of $t$ when the covariates are all zero. Here, we suppress the dependence of $h(t; b)$ on $\tilde{Z}(t)$ for brevity. As $h(t; b)$ is a monotone transform of $t$, we define $F_{h(T,b)}\{t \mid \tilde{Z}(t)\} = F_T\{h^{-1}(t; b) \mid \tilde{Z}(h^{-1}(t; b))\}$, where $F_T$ is as defined above and $h^{-1}(\cdot; b)$ is the inverse function of $t \mapsto h(t; b)$ with $b$ fixed. Let $Q_{h(T,b)}(\tau \mid \tilde{Z})$ be the inverse function of $F_{h(T,b)}\{t \mid \tilde{Z}(t)\}$ accordingly, it is straightforward to check that $Q_T(\tau \mid \tilde{Z})$ possesses the equivariance property of monotone transform, i.e. $Q_{h(T,b)}(\tau \mid \tilde{Z}) = h(Q_T(\tau \mid \tilde{Z}; b))$.

Now, our proposed model stipulates that, for $\tau \in (0, 1), \quad (3.1)$

$$Q_{h(T; b_0(\tau))}(\tau \mid \tilde{Z}) = 1,$$
or equivalently,

\[
Q_T(\tau | \tilde{Z}) = h^{-1}(1; \beta_0(\tau)).
\]

(3.2)

Since \( Z(t) \) may include a constant intercept term, (3.1) is equivalent to \( Q_{h(T; \beta_0(\tau))}(\tau | \tilde{Z}) = q(\tau) \) for some given \( q(\tau) \), because the constant function \( q(\tau) \) can be absorbed into the intercept of the parameter \( \beta_0(\tau) \). Without loss of generality, we require \( q(\tau) \equiv 1 \) to ensure identifiability.

To better illustrate (3.1) and (3.2), we consider two special cases. First, if \( Z(t) \equiv Z \) is time-independent, then \( Q_T(\tau | \tilde{Z}) \equiv Q_T(\tau | Z) \) represents the conditional quantile function of \( T \) given \( Z \). Moreover, (3.2) becomes \( Q_T(\tau | Z) = \exp\{Z^\top \beta_0(\tau)\} \) for \( \tau \in (0, 1) \) because \( h(t; b) = t \exp(-Z^\top b) \), so our model reduces to the censored quantile regression model proposed in [41]. Second, suppose \( Z(t) = (1, Z(t)^\top)^\top \) and \( \beta_0(\tau) = (\log q(\tau), \tilde{\beta}_0^\top)^\top \) for some function \( q(\tau) \in (0, \infty) \), i.e. only the intercept term is \( \tau \)-varying. Let \( \tilde{h}(t; b) = \int_0^t \exp\{-\tilde{Z}(s)^\top b\} \, ds \), we have

\[
h(T; \beta_0(\tau)) = \exp\{-\log q(\tau)\} \int_0^T \exp\{-\tilde{Z}(s)^\top \tilde{\beta}_0\} \, ds = \frac{1}{q(\tau)} \tilde{h}(T; \tilde{\beta}_0),
\]

thus model (3.1) reduces to \( Q_{h(T; \beta_0)}(\tau | \tilde{Z}) = q(\tau) \) for \( \tau \in (0, 1) \). In other words, the conditional quantile function of \( \tilde{h}(T; \tilde{\beta}_0) \) is free of \( \tilde{Z} \) for \( \tau \in (0, 1) \). This specification recovers the accelerated failure time model for time-dependent covariates proposed in [47] and [38], although the treatment of internal covariates is not discussed in their work. In this regard, model (3.2) is a natural generalization of the quantile regression model which accommodates time-dependent covariates. Indeed, by incorporating \( \tau \)-varying coefficients, model (3.2) relaxes the assumption of the accelerated failure time model that the covariates have identical impacts on the failure time across all quantiles, hence allowing quantile dependent covariate effects.

Suppose we observe \( \tilde{T} = T \wedge C, \Delta = I(T \leq C) \) and \( \tilde{Z}(|\tilde{T}) = \{Z(s), s \in [0, \tilde{T}]\} \), where \( C \) is the censoring time which is conditionally independent of \( T \) given the covariates in the sense that

\[
P\{T \in [t, t + dt] | T \geq t, \tilde{Z}(t)\} = P\{T \in [t, t + dt] | T \geq t, C \geq t, \tilde{Z}(t)\}.
\]
Let \( N(t) = \Delta I(\tilde{T} \leq t) \) and \( Y(t) = I(\tilde{T} \geq t) \), we write \( N\{h^{-1}(t; \mathbf{b})\} \) as \( N(t; \mathbf{b}) \) and \( Y\{h^{-1}(t; \mathbf{b})\} \) as \( Y(t; \mathbf{b}) \), so that \( N(t; \mathbf{b}) \) and \( Y(t; \mathbf{b}) \) denote the counting process and at-risk process of the baseline transform \( h(t; \mathbf{b}) \) respectively. Denote \( H(u) = -\log(1 - u) \), so that \( \Lambda_T \{ t | \bar{Z}(t) \} = H[F_T \{ t | \tilde{Z}(t) \}] \). Results from the classic counting process theory, see, for instance [16], suggests that

\[
0 = E \left[ \int_0^t Z(s) \left( dN(s) - Y(s) d\Lambda_T \{ s | \bar{Z}(s) \} \right) \right] = E \left[ \int_0^\infty Z(s) I(s \leq t) dN(s) \right] - E \left[ \int_0^t Z(s) Y(s) dH[F_T \{ s | \bar{Z}(s) \}] \right].
\]

(3.3)

The first term of (3.3) can be simplified into \( E[Z(\tilde{T})N(t)] \). With a change of variable \( s \mapsto Q_T(u | \tilde{Z}) \), the second term can also be rewritten as

\[
E \left[ \int_0^{F_T \{ t | \bar{Z}(t) \}} Z(Q_T(u | \tilde{Z})) Y(Q_T(u | \tilde{Z}) \) dH(u) \right] = E \left[ \int_0^{F_T \{ t | \bar{Z}(t) \}} Z\{h^{-1}(1; \beta_0(u))\} Y\{1; \beta_0(u)\} dH(u) \right],
\]

where the equality follows from model (3.2). Note that (3.3) holds for all \( t \), we may choose \( t = Q_T(\tau | \bar{Z}) = h^{-1}(1; \beta_0(\tau)) \), which leads to

\[
E \left[ Z(\tilde{T})N\{1; \beta_0(\tau)\} - \int_0^{\tau} Z\{h^{-1}(1; \beta_0(u))\} Y\{1; \beta_0(u)\} dH(u) \right] = 0.
\]

(3.4)

Due to the fact that \( Y\{1; \beta_0(u)\} = 0 \) when \( h(\tilde{T}; \beta_0(u)) < 1 \) and that \( h^{-1}(1; \beta_0(u)) \leq \tilde{T} \) when \( h(\tilde{T}; \beta_0(u)) \geq 1 \), the term \( Z\{h^{-1}(1; \beta_0(u))\} Y\{1; \beta_0(u)\} \) is known given the observed covariate history \( \tilde{Z}(\tilde{T}) \). As a result, (3.4) is fully observable and computable for given \( \beta_0(\tau), \tau \in (0, 1) \).

Now, suppose we have a sample of \( \{ \tilde{T}_i, \Delta_i, \tilde{Z}_i(\tilde{T}_i) \}_{i=1}^n \) that are i.i.d. copies of \( \{ \tilde{T}, \Delta, \tilde{Z}(\tilde{T}) \} \), we can apply the empirical counterpart of (3.4) to estimate \( \beta_0(\tau), \tau \in (0, 1) \). More precisely, we define our estimator \( \{ \tilde{\beta}(\tau), \tau \in (0, \tau_U) \} \) to be a right-continuous step function with jumps on a grid \( S_{L(n)} = \{ 0 = \tau_0 < \tau_1 < \cdots < \tau_{L(n)} = \tau_U < 1 \} \) and denote \( \| S_{L(n)} \| = \sup_{1 \leq j \leq L(n)} | \tau_j - \tau_{j-1} | \), with the number of grids \( L = L(n) \) depending on \( n \), which will be suppressed for brevity in the
sequel. Instead of estimating $\beta_0(\tau)$ for all $\tau \in (0, 1)$, we confine our attention to $\tau \in (0, \tau_U]$, where $\tau_U \in (0, 1)$ in order to tackle the identifiability issues due to censorship. Now, by definition we have $Q_\tau(T \mid Z) = 0$, so we take $\hat{\beta}(0)$ such that $h^{-1}(1; \hat{\beta}(0)) = 0$. Subsequently for each $\tau_j$, $j = 1, \cdots, L$, we obtain $\hat{\beta}(\tau_j)$ by solving the following estimating equation, which is a discretized version of the empirical counterpart of (3.4), i.e.

$$S_n(\beta, \tau_j) = \frac{1}{n} \sum_{i=1}^{n} \left[ Z_i(\tau_j)N_i(1; \beta_j) \right]$$

$$- \sum_{k=1}^{j} Z_i(h_i^{-1}(1; \hat{\beta}(\tau_{k-1})))Y_i(1; \hat{\beta}(\tau_{k-1})) \{ H(\tau_k) - H(\tau_{k-1}) \} = 0.$$  \(3.5\)

While (3.5) shares a similar structure as [41]'s estimating equation, the $p$-dimensional discontinuous estimating equations are now generally non-monotone with respect to $\beta(\tau)$ due to the extra complication caused by time-dependent covariates. In our numerical studies, we adopt the differential evolution algorithm in MATLAB to solve for the roots by minimizing $L_1$ norm of the estimating equations. According to our experience, the stochastic-based differential evolution algorithm performs generally better than other deterministic derivative-free methods such as simplex search in the sense that the estimators are less likely to be trapped in local minima.

### 3.3 Asymptotic Results

We introduce in this section the asymptotic properties of the proposed estimator. Define $\mu_1(b) = E[Z(\tilde{T})N(1; b)]$, $\mu_2(b) = E[Z(h^{-1}(1; b))Y(1; b)]$, $\nu_1(b) = \nabla_b \mu_1(b)$ and $\nu_2(b) = \nabla_b \mu_2(b)$, where $\nabla_b$ denotes the gradient operator with respect to $b$. Also, let $\hat{\mu}_1$, $\hat{\mu}_2$, $\hat{\nu}_1$ and $\hat{\nu}_2$ denote their empirical counterparts respectively, i.e. $\hat{\mu}_1(b) = n^{-1} \sum_{i=1}^{n} Z_i(\tilde{T}_i)N_i(1; b)$, etc.

Let $B(d, u) = \{ b \in \mathcal{R}^p : \inf_{\tau \in (0, u]} \| \mu_1(b) - \mu_1(\beta_0(\tau)) \| \leq d \}$, and write $B(d, \tau_U)$ as $B(d)$. Furthermore, we define $\psi_Z(b, \tau) = E \| Z\{ h^{-1}(1; b) \} - Z\{ h^{-1}(1; \beta_0(\tau)) \} \|$ and $\psi_Y(b, \tau) = E \| Y\{ 1; b \} - Y\{ 1; \beta_0(\tau) \} \|$. Now, we list the conditions required to establish the asymptotic properties of $\hat{\beta}(\tau)$. 

48
Condition 3.1. \[ \|Z(0)\| + \int_0^\infty \|dZ(t)\| < \infty \text{ almost surely element-wise.} \]

Condition 3.2. \( \psi_Z(b, \tau) \) is continuous at \( b = \beta_0(\tau) \) uniformly over \( \tau \in [\tau_L, \tau_U] \) for any \( \tau_L \in (0, \tau_U] \), i.e. \( \sup_{\tau \in [\tau_L, \tau_U] \setminus \{b - \beta_0(\tau)\}} |\psi_Z(b, \tau)| \to 0 \).

Condition 3.3. \( \psi_Y(b, \tau) \) is continuous at \( b = \beta_0(\tau) \) uniformly over \( \tau \in [\tau_L, \tau_U] \) for any \( \tau_L \in (0, \tau_U] \), i.e. \( \sup_{\tau \in [\tau_L, \tau_U] \setminus \{b - \beta_0(\tau)\}} |\psi_Y(b, \tau)| \to 0 \).

Condition 3.4. \( \beta_0(\tau) \) is a Lipschitz function of \( \tau \) for \( \tau \in (0, \tau_U] \).

Condition 3.5. Each element of \( \mu_1(\beta_0(\tau)) \) is a Lipschitz function of \( \tau \) for \( \tau \in (0, \tau_U] \).

Condition 3.6. Each element of \( v_2(b) \) is uniformly bounded for \( b \in B(d_0) \).

Condition 3.7. \( v_1(b) \) is invertible and each element of its inverse matrix is uniformly bounded for \( b \in B(d_0)/B(d_0, \tau_L) \) for any \( \tau_L \in (0, \tau_U] \).

Conditions 3.1 and 3.2 are essential regularity conditions on the covariate processes. Condition 3.1 requires that the covariate process as a function of time is of finite total variation, while Condition 3.2 enforces smoothness of the scaled covariate process in the expectation sense with respect to the regression coefficient in a shrinking neighbourhood of the true value. In particular, the two conditions hold when the covariate process has uniformly bounded derivatives almost everywhere except with possible jumps where the random inter-arrival times have uniformly bounded densities. Similar to Condition 3.2, Condition 3.3 is needed to establish the tightness results of the functionals. Note that if the covariates concerned are external, Condition 3.3 is implied by the uniform boundedness of the conditional density function of the failure time distribution given the covariates. In particular, [41] discusses the scenario of time-independent covariates. In general, when the time-dependent covariates are internal, such conditional density function may not exist and thus we resort to Condition 3.3. Condition 3.4 ensures the asymptotic results hold for all \( \tau \). When \( \beta_0(\tau) \) is not Lipschitz but there are finite discontinuity jump points, we can see from the proofs that the asymptotic results hold almost everywhere for \( \tau \) except on the jump points. Conditions 3.5 and 3.6 are mild requirements on the defined functionals which build the blocks for
the asymptotic properties of the estimator via a sequential approach. Condition 3.7 assumes the invertibility of the gradient matrix in a neighborhood of the true parameter value, i.e. it enforces a local convexity of the estimating equations in such neighborhood, and thus establishes the asymptotic results of the estimator from that of the functional. While the counterpart of Condition 3.7 in [41] is required to hold at the true value only, we further assume that it holds for a neighborhood because our estimating equation is not necessarily convex in nature.

Note that the additional requirement of Condition 3.7 is a result of our less stringent formulation which requires the covariate history observable only up to the survival time. Indeed, if we instead assume the entire process \( \tilde{Z} \) is observed as in [22], or more realistically, \( \tilde{Z}(t_{\text{end}}) \) is known for some \( t_{\text{end}} \) large enough, we recover the analogous convex problem of [41]. Under such circumstances, as an alternative to solving (3.5), one may obtain \( \hat{\beta}(\tau_j) \) for each \( j = 1, \cdots, L \) as a minimizer of the following quantity:

\[
\frac{1}{n} \sum_{i=1}^{n} \left[ \tilde{T}_i - h_i^{-1}(1; \beta(\tau_j)) \right] \left[ \sum_{k=1}^{j} Y_i \{ 1; \hat{\beta}(\tau_{k-1}) \} \{ H(\tau_k) - H(\tau_{k-1}) \} - N_i \{ 1; \beta(\tau_j) \} \right].
\] (3.6)

Note that its derivative with respect to \( \beta(\tau_j) \) is given by

\[
\frac{1}{n} \sum_{i=1}^{n} \nabla_{\beta} h_i^{-1}(1; \beta) \bigg|_{\beta=\beta(\tau_j)} \left[ N_i \{ 1; \beta(\tau_j) \} - \sum_{k=1}^{j} Y_i \{ 1; \hat{\beta}(\tau_{k-1}) \} \{ H(\tau_k) - H(\tau_{k-1}) \} \right].
\]

which can be shown to possess a zero mean for \( \beta(\cdot) = \beta_0(\cdot) \) as in (3.3). As a result, deriving \( \hat{\beta}(\tau_j) \) sequentially from (3.6) is justified, since \( \beta(\cdot) = \beta_0(\cdot) \) is the theoretical minimizer of (3.6). Indeed, we introduce (3.6) as an alternative in this special occasion because (3.6) is an \( L_1 \)-type convex problem, where the numerical solution can be easily solved. However, the minimand (3.6) is usually uncomputable in the general setting where only \( \tilde{Z}(\tilde{T}) \) is observed, because evaluating the quantity \( h^{-1}(1; \beta(\tau)) \) involves the entire process \( \tilde{Z} \). Under the circumstances where the alternative (3.6) can be used, because of the additional convex nature in the estimation, we may relax Condition 3.7 above to hold at the true parameter value only, similar to its counterpart in [41]. But in this paper, in contrary to [22], we consider a more practical, albeit theoretically challenging,
scenario that the covariate history is observed until the survival time, resulting in the less lenient version of Condition 3.7.

Now, the following two theorems establish the uniform consistency and the weak convergence of the proposed estimator.

**Theorem 3.1.** Under Conditions 3.1 to 3.7, if \( \|S_L\| \to 0 \), then \( \sup_{\tau \in [\tau_L, \tau_U]} \|\hat{\beta}(\tau) - \beta_0(\tau)\| \to_p 0 \) where \( \tau_L \in (0, \tau_U] \).

**Theorem 3.2.** Under Conditions 3.1 to 3.7, if \( n^{1/2}\|S_L\| \to 0 \), then \( n^{1/2} \{\hat{\beta}(\tau) - \beta_0(\tau)\} \) converges weakly to a Gaussian process for \( \tau \in [\tau_L, \tau_U] \) where \( \tau_L \in (0, \tau_U] \).

Theorem 3.1 states that provided the grid size in the numerical algorithm diminishes to 0 as \( n \to \infty \), the proposed estimator converges uniformly to the true value as a process of quantiles. Theorem 3.2 requires further that the grid size is of a finer order \( o(n^{-1/2}) \) so that the weak convergence of the proposed estimator to a Gaussian process can be obtained. Following the idea of [41], we consider the asymptotic properties of the function \( \mu_1\{\hat{\beta}(\tau)\} \) as a machinery. In particular, by exploiting the stepwise structure of the proposed estimator, we establish the uniform consistency of \( \mu_1\{\hat{\beta}(\tau)\} \) to \( \mu_1\{\beta_0(\tau)\} \) and the weak convergence of \( n^{1/2} [\mu_1\{\hat{\beta}(\tau)\} - \mu_1\{\beta_0(\tau)\}] \) via empirical process techniques, which in turn renders the asymptotic properties of \( \hat{\beta}(\tau) \). Detailed proofs of Theorems 3.1 and 3.2 are deferred to Sections 3.7.1 and 3.7.2, respectively.

As we shall see in the proof of the weak convergence, the limiting Gaussian process has a covariance process which is a function of the unknown conditional density functions of the underlying failure times \( T \). Therefore, Theorem 3.2 does not offer a direct inference of the estimator as in other semiparametric models. To yield a reliable estimate of the covariance process of the estimator, we propose a minimand perturbing based resampling scheme in the spirit of [24] and [41].

To this end, let \( \{\zeta_1, \ldots, \zeta_n\} \) be non-negative i.i.d. random variables with mean 1 and variance 1, such as exponentially distributed with mean 1. Given a dataset \( \{\tilde{T}_i, \Delta_i, \tilde{Z}_i(\tilde{T}_i)\}_{i=1}^n \), we fix the data and independently generate \( \{\zeta_1, \ldots, \zeta_n\} \). Similar to \( \hat{\beta}(\tau) \), we define \( \{\beta^*(\tau), \tau \in (0, \tau_U]\} \) to be
a right-continuous step function with jumps on $S_L$ and take $h^{-1}(1; \beta^*(0)) = 0$. Sequentially for $j = 1, \cdots, L$, we obtain $\beta^*(\tau_j)$ by solving the following perturbed version of (3.5), i.e.

$$S_n^*(\beta, \tau_j) = \frac{1}{n} \sum_{i=1}^{n} \zeta_i \left[ Z_i(\tilde{T}_i)N_i\{1; \beta(\tau_j)\} - \sum_{k=1}^{j} Z_i\{h_i^{-1}(1; \beta^*(\tau_{k-1}))\}Y_i\{1; \beta^*(\tau_{k-1})\}\{H(\tau_k) - H(\tau_{k-1})\}\right] = 0.$$  

We shall repeat the above procedure $B$ times to obtain $B$ realizations of $\beta^*(\tau)$, i.e. for each $r = 1, \cdots, B$, we generate a set of $\{\zeta_1, \cdots, \zeta_n\}$ and estimate $\beta^*_r(\tau_j)$ sequentially by minimizing the $L_1$ norm of $\tilde{S}_n(\beta, \tau_j)$. Using the result of Theorem 3.3, the standard deviation of $\hat{\beta}$ can then be approximated by the sample standard deviation of $\{\beta^*_r(\tau)\}_{r=1}^{B}$, while the confidence interval for $\beta(\tau)$ can be constructed by exploiting the empirical quantiles of $\{\beta^*_r(\tau)\}_{r=1}^{B}$ or by normal approximation.

**Theorem 3.3.** Under Conditions 3.1 to 3.7, $n^{1/2} \{\beta^*(\tau) - \hat{\beta}(\tau)\}$ given the observed data converges weakly to the same limiting process of $n^{1/2} \{\hat{\beta}(\tau) - \beta_0(\tau)\}$ for $\tau \in [\tau_L, \tau_U]$ where $\tau_L \in (0, \tau_U]$.

The justification of Theorem 3.3 is given in Section 3.7.3.

### 3.4 Simulations

In this section, we demonstrate the numerical performance of the proposed estimator via Monte Carlo simulated data sets. While our framework allows any covariate process, it suffices for us to consider piecewise constant covariate processes in terms of practical purpose. To this end, we assume in this section that the covariate process changes in values exactly $d - 1$ times at the time
Table 3.1: Simulation Results: Homogeneous Setup

<table>
<thead>
<tr>
<th></th>
<th>$\tau = 0.25$</th>
<th>$\tau = 0.50$</th>
<th>$\tau = 0.75$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\hat{\beta}^{(0)}$</td>
<td>$\hat{\beta}^{(1)}$</td>
<td>$\hat{\beta}^{(2)}$</td>
</tr>
<tr>
<td>$n = 200 &amp; 20%$ censoring</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Proposed Bias</td>
<td>0.014</td>
<td>-0.004</td>
<td>-0.004</td>
</tr>
<tr>
<td>SD</td>
<td>0.253</td>
<td>0.165</td>
<td>0.155</td>
</tr>
<tr>
<td>SE</td>
<td>0.258</td>
<td>0.162</td>
<td>0.165</td>
</tr>
<tr>
<td>CP</td>
<td>0.948</td>
<td>0.934</td>
<td>0.952</td>
</tr>
<tr>
<td>GGR Bias</td>
<td>0.669</td>
<td>-0.687</td>
<td>-0.700</td>
</tr>
<tr>
<td>SD</td>
<td>0.599</td>
<td>0.657</td>
<td>0.613</td>
</tr>
<tr>
<td>$n = 200 &amp; 40%$ censoring</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Proposed Bias</td>
<td>0.015</td>
<td>0.010</td>
<td>0.003</td>
</tr>
<tr>
<td>SD</td>
<td>0.293</td>
<td>0.190</td>
<td>0.182</td>
</tr>
<tr>
<td>SE</td>
<td>0.329</td>
<td>0.200</td>
<td>0.205</td>
</tr>
<tr>
<td>CP</td>
<td>0.972</td>
<td>0.964</td>
<td>0.972</td>
</tr>
<tr>
<td>GGR Bias</td>
<td>0.766</td>
<td>-0.705</td>
<td>-0.679</td>
</tr>
<tr>
<td>SD</td>
<td>0.638</td>
<td>0.643</td>
<td>0.627</td>
</tr>
<tr>
<td>$n = 500 &amp; 20%$ censoring</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Proposed Bias</td>
<td>-0.002</td>
<td>0.007</td>
<td>0.002</td>
</tr>
<tr>
<td>SD</td>
<td>0.153</td>
<td>0.101</td>
<td>0.099</td>
</tr>
<tr>
<td>SE</td>
<td>0.158</td>
<td>0.101</td>
<td>0.099</td>
</tr>
<tr>
<td>CP</td>
<td>0.944</td>
<td>0.930</td>
<td>0.944</td>
</tr>
<tr>
<td>GGR Bias</td>
<td>0.662</td>
<td>-0.379</td>
<td>-0.394</td>
</tr>
<tr>
<td>SD</td>
<td>0.537</td>
<td>0.397</td>
<td>0.376</td>
</tr>
<tr>
<td>$n = 500 &amp; 40%$ censoring</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Proposed Bias</td>
<td>-0.003</td>
<td>0.015</td>
<td>0.010</td>
</tr>
<tr>
<td>SD</td>
<td>0.175</td>
<td>0.115</td>
<td>0.112</td>
</tr>
<tr>
<td>SE</td>
<td>0.193</td>
<td>0.121</td>
<td>0.119</td>
</tr>
<tr>
<td>CP</td>
<td>0.958</td>
<td>0.940</td>
<td>0.948</td>
</tr>
<tr>
<td>GGR Bias</td>
<td>0.666</td>
<td>-0.406</td>
<td>-0.394</td>
</tr>
<tr>
<td>SD</td>
<td>0.544</td>
<td>0.414</td>
<td>0.400</td>
</tr>
</tbody>
</table>

Approach GGR denotes [18]'s proposal. Bias, SD, SE, CP denote empirical bias, empirical standard deviation, average standard error, and empirical coverage probability of the 95% confidence intervals, respectively.
### Table 3.2: Simulation Results: Heterogeneous Setup

<table>
<thead>
<tr>
<th></th>
<th>$\tau = 0.25$</th>
<th>$\tau = 0.50$</th>
<th>$\tau = 0.75$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\hat{\beta}^{(0)}$</td>
<td>$\hat{\beta}^{(1)}$</td>
<td>$\hat{\beta}^{(2)}$</td>
</tr>
<tr>
<td>Proposed Bias</td>
<td>0.053</td>
<td>0.060</td>
<td>0.048</td>
</tr>
<tr>
<td>SD</td>
<td>0.444</td>
<td>0.403</td>
<td>0.419</td>
</tr>
<tr>
<td>SE</td>
<td>0.459</td>
<td>0.423</td>
<td>0.458</td>
</tr>
<tr>
<td>CP</td>
<td>0.938</td>
<td>0.954</td>
<td>0.956</td>
</tr>
<tr>
<td>GGR Bias</td>
<td>0.998</td>
<td>1.175</td>
<td>1.310</td>
</tr>
<tr>
<td>SD</td>
<td>0.677</td>
<td>0.714</td>
<td>0.884</td>
</tr>
</tbody>
</table>

$n = 200$ & 20% censoring

<table>
<thead>
<tr>
<th></th>
<th>$\tau = 0.25$</th>
<th>$\tau = 0.50$</th>
<th>$\tau = 0.75$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\hat{\beta}^{(0)}$</td>
<td>$\hat{\beta}^{(1)}$</td>
<td>$\hat{\beta}^{(2)}$</td>
</tr>
<tr>
<td>Proposed Bias</td>
<td>0.093</td>
<td>0.080</td>
<td>0.041</td>
</tr>
<tr>
<td>SD</td>
<td>0.504</td>
<td>0.472</td>
<td>0.510</td>
</tr>
<tr>
<td>SE</td>
<td>0.545</td>
<td>0.495</td>
<td>0.538</td>
</tr>
<tr>
<td>CP</td>
<td>0.944</td>
<td>0.946</td>
<td>0.956</td>
</tr>
<tr>
<td>GGR Bias</td>
<td>1.087</td>
<td>1.254</td>
<td>1.442</td>
</tr>
<tr>
<td>SD</td>
<td>0.714</td>
<td>0.777</td>
<td>0.858</td>
</tr>
</tbody>
</table>

$n = 200$ & 40% censoring

<table>
<thead>
<tr>
<th></th>
<th>$\tau = 0.25$</th>
<th>$\tau = 0.50$</th>
<th>$\tau = 0.75$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\hat{\beta}^{(0)}$</td>
<td>$\hat{\beta}^{(1)}$</td>
<td>$\hat{\beta}^{(2)}$</td>
</tr>
<tr>
<td>Proposed Bias</td>
<td>0.014</td>
<td>0.020</td>
<td>0.020</td>
</tr>
<tr>
<td>SD</td>
<td>0.270</td>
<td>0.251</td>
<td>0.273</td>
</tr>
<tr>
<td>SE</td>
<td>0.282</td>
<td>0.256</td>
<td>0.280</td>
</tr>
<tr>
<td>CP</td>
<td>0.944</td>
<td>0.940</td>
<td>0.952</td>
</tr>
<tr>
<td>GGR Bias</td>
<td>0.847</td>
<td>1.150</td>
<td>1.539</td>
</tr>
<tr>
<td>SD</td>
<td>0.647</td>
<td>0.716</td>
<td>0.869</td>
</tr>
</tbody>
</table>

$n = 500$ & 20% censoring

<table>
<thead>
<tr>
<th></th>
<th>$\tau = 0.25$</th>
<th>$\tau = 0.50$</th>
<th>$\tau = 0.75$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\hat{\beta}^{(0)}$</td>
<td>$\hat{\beta}^{(1)}$</td>
<td>$\hat{\beta}^{(2)}$</td>
</tr>
<tr>
<td>Proposed Bias</td>
<td>0.034</td>
<td>0.034</td>
<td>0.028</td>
</tr>
<tr>
<td>SD</td>
<td>0.309</td>
<td>0.299</td>
<td>0.319</td>
</tr>
<tr>
<td>SE</td>
<td>0.333</td>
<td>0.305</td>
<td>0.331</td>
</tr>
<tr>
<td>CP</td>
<td>0.950</td>
<td>0.942</td>
<td>0.948</td>
</tr>
<tr>
<td>GGR Bias</td>
<td>0.899</td>
<td>1.198</td>
<td>1.570</td>
</tr>
<tr>
<td>SD</td>
<td>0.682</td>
<td>0.770</td>
<td>0.933</td>
</tr>
</tbody>
</table>

$n = 500$ & 40% censoring

Approach GGR denotes [18]'s proposal. Bias, SD, SE, CP denote empirical bias, empirical standard deviation, average standard error, and empirical coverage probability of the 95% confidence intervals, respectively.

54
nodes \{W_{i1}, \cdots, W_{i,d-1}\} where 0 = W_{i0} < W_{i1} < \cdots < W_{i,d-1} < W_{id} = \infty, then we can express

\[
h_i(t; \beta) = \sum_{s=1}^{d} \left[ \sum_{r=1}^{s-1} (W_{ir} - W_{i,r-1}) \exp \left\{ -Z_i(W_{i,r-1})^{\top} \beta \right\} + (t - W_{i,s-1}) \exp \left\{ -Z_i(W_{i,s-1})^{\top} \beta \right\} \right] \times I(W_{i,s-1} \leq t < W_{is}). \tag{3.7}
\]

Since \(h_i(t; \beta)\) is monotone in \(t\), its inverse function is given by

\[
h_i^{-1}(t; \beta) = \sum_{s=1}^{d} \left[ t - \sum_{r=1}^{s-1} (W_{ir} - W_{i,r-1}) \exp \left\{ -Z_i(W_{i,r-1})^{\top} \beta \right\} \right] \times \\
\exp\{Z_i(W_{i,s-1})^{\top} \beta \} + W_{i,s-1} \right] \times \\
I \left( \sum_{r=1}^{s-1} (W_{ir} - W_{i,r-1}) \exp \left\{ -Z_i(W_{i,r-1})^{\top} \beta \right\} \leq \\
t < \sum_{r=1}^{s} (W_{ir} - W_{i,r-1}) \exp \left\{ -Z_i(W_{i,r-1})^{\top} \beta \right\} \right). \tag{3.8}
\]

Recall our quantile regression model stipulates that \(Q_{T_i}(\tau \mid \tilde{Z}_i) = h_i^{-1}(1; \beta_0(\tau))\). When only external covariates are considered, \(Q_{T_i}(\tau \mid \tilde{Z}_i)\) coincides with the usual conditional quantile function of \(T\) given \(\tilde{Z}\), so we could use the last display to generate the desired lifetime.

In the subsequent simulation examples, we consider a three-dimensional covariate process \(Z_i(t) = (Z_{i0}(t), Z_{i1}(t), Z_{i2}(t))^\top\) which is piecewise linear with \(d = 3\) subintervals. Specifically, for \(t \geq 0\), we let \(Z_{i0}(t) = 1, \)

\[
Z_{i1}(t) = \sum_{s=1}^{d} X_{i1s} I(W_{i,s-1} \leq t < W_{is}),
\]

\[
Z_{i2}(t) = \sum_{s=1}^{d} X_{i2s} I(W_{i,s-1} \leq t < W_{is}),
\]

where \(X_{i1s}, X_{i2s}\) are i.i.d. Gamma(2, 1/2) for \(i = 1, \cdots, n\) and \(s = 1, \cdots, d\). Now, the first example demonstrates the numerical performance of the proposed method with a conditionally
independent censoring and a homogeneous error distribution. Let $\beta_0(\tau) = (Q_\varepsilon(\tau) + b_0, b_1, b_2)^T$, where $Q_\varepsilon(\tau)$ denotes the quantile function of $\varepsilon$, so that model (3.1) becomes

$$
\int_0^{Q_{T_i}(\tau|Z_i)} \exp\{-Q_\varepsilon(\tau) + b_0 + Z_{i1}(s)b_1 + Z_{i2}(s)b_2\} \, ds = 1, \quad \tau \in (0, 1).
$$

Here, we consider $b_0 = b_1 = b_2 = 2$ and $\varepsilon$ with a standard normal distribution. For $i = 1, \cdots, n$ and $s = 1, \cdots, d - 1$, the time nodes $W_{is}$ are recovered recursively through the equation $W_{is} = W_{i,s-1} + S_{is}$, where $S_{is}$, which represent the sojourn times between the time nodes, are i.i.d. $U(0, 1)$. Recall that the covariate processes can only be observed up to the survival times, so we generate the sojourn time in a way that the simulated failure times $\{T_1, \cdots, T_n\}$ spread through the subintervals derived from $\{W_{i0}, \cdots, W_{id}\}$, hence replicating a realistic scenario. Also, note that the constructed conditional quantile function of the failure time will be monotone increasing for any realization of the covariate process.

Note that if $Z_{i1}(s) = Z_{i1}$ and $Z_{i2}(s) = Z_{i2}$ are actually time-independent, then the last display reduces to $Q_{T_i}(\tau | Z_i) \exp\{-Q_\varepsilon(\tau) + b_0 + Z_{i1}b_1 + Z_{i2}b_2\} = 1$, i.e.

$$
Q_{\log T_i}(\tau | Z_i) = Q_{\varepsilon + b_0 + Z_{i1}b_1 + Z_{i2}b_2}(\tau), \quad \tau \in (0, 1).
$$

Hence we recover a form of the accelerated failure time model, i.e. $\log T_i = b_0 + Z_{i1}b_1 + Z_{i2}b_2 + \varepsilon_i$ where $\{\varepsilon_1, \cdots, \varepsilon_n\}$ are i.i.d. standard normal. Therefore, under the scenario of $\beta(\tau) = (Q_\varepsilon(\tau) + b_0, b_1, b_2)^T$, we may view the proposed model as a generalization of the accelerated failure time model with time-dependent covariates. To resemble a conditionally independent censoring scheme, we generate the censoring times $\{C_1, \cdots, C_n\}$ from i.i.d. exponential distribution with mean $\exp\{c + 100Z_{i1}(0) + 100Z_{i2}(0)\}$ so that the censoring times depend on the covariate processes. We set $c = -100$ or $c = -150$ to achieve a censoring rate of 20% or 40%, respectively.

We generate 500 Monte-Carlo datasets with $n = 200$ or $n = 500$ sample points each, and adopt the aforementioned resampling procedure with $B = 300$ and $\{\zeta_1, \cdots, \zeta_n\}$ from i.i.d. exponential distribution with mean 1 to reconcile the asymptotic distribution of the parameter estimates. We
implement the differential evolution algorithm in MATLAB to obtain solutions of the estimating equations.

Table 3.1 reports the simulation results for the proposed method and [18]'s proposal, in terms of empirical bias (Bias), empirical standard deviation (SD), average standard error (SE) based on the resampling method, and coverage rate (CP) of the 95% confidence intervals constructed by normal approximation. It is shown that the proposed method produces favourable empirical bias and empirical standard deviation with different settings of sample size and censoring rate. Moreover, the average standard error across all resampling trials is closed to the empirical standard deviation and the coverage probability is close to 95%, validating the aforementioned resampling method. Meanwhile, [18]'s method produces non-negligible empirical bias and larger standard deviation when compared to our proposal, because their framework requires the marginally independent censoring assumption instead of the conditional independence we assumed, which is violated in this example as the censoring mechanism depends on the covariate processes.

In the second example, we consider a setup where a heterogeneous error distribution is taken in account. To this end, with the same covariate processes as in the first example, we let \( \beta_0(\tau) = (Q_e(\tau) + b_0, b_1, b_2 + Q_e(\tau))^\top \), in which case model (3.1) becomes

\[
\int_0^{Q_{\tau_i}(\tau|\bar{Z}_i)} \exp[-((1 + Z_{i1}(s))Q_e(\tau) + b_0 + Z_{i1}(s)b_1 + Z_{i2}(s)b_2)] \, ds = 1, \quad \tau \in (0, 1).
\]

Again we have \( b_0 = b_1 = b_2 = 2 \) and \( e \) with a standard normal distribution. Also, for \( i = 1, \ldots, n \) and \( s = 1, \ldots, d - 1 \), the sojourn times \( S_{is} \), the time nodes \( W_{is} \) and the censoring time \( C_i \) are generated in the same fashion as previously. Note that with the current setup, even if \( Z_{i1}(s) \) and \( Z_{i2}(s) \) are time-independent processes, the model are not within the framework of the accelerated failure time model, because of the extra covariate dependent heterogeneity imposed on the error quantile. Nevertheless, as demonstrated in Table 3.2, our methodology provides reasonable numerical performance as in the homogeneous setting of the first example. This shows that the proposed censored quantile regression model could serve as a flexible machinery for survival data despite
the error heterogeneity when dealing with time-dependent covariates.

While [22]’s model requires the entire history of the covariate process and the different formulation makes direct comparison of the model parameters not as straightforward as it appears, we introduce a viable goodness of fit statistic to compare the two proposals. In the spirit of the residual cusum process, see, for example, [49], [35], [40], amongst others, we propose the quantity $R_n(\tau) = \frac{1}{n} \sum_{i=1}^{n} \{E_n(Z_i, \tau)\}^2$ as a measure of the goodness of fit, where

$$E_n(z, \tau) = \frac{1}{n} \sum_{j=1}^{n} I[Z_j(0) \leq z(0)] \left[ N_j \{1; \hat{\beta}(\tau)\} - \int_0^\tau Y_j \{1; \hat{\beta}(u)\} dH(u) \right].$$

Here, $Z_j(0) \leq z(0)$ if and only if the inequality holds for all entries, i.e. each component of $Z_j(0)$ is less than or equal to the corresponding component of $z(0)$. Provided that the model is correctly specified, the martingale residual within the square bracket of the last display will be approximately zero, hence a smaller statistic would naturally imply a better fit. Meanwhile, we define a similar statistic for [22]’s proposal, where the estimated quantile $h_i^{-1}(1; \hat{\beta}(\tau))$ is replaced by their corresponding analogy. The mean goodness of fit statistics for the Monte Carlo simulated datasets in the previous two examples are reported in Table 3.3. The results suggest that the proposed method provides a better fit in terms of the smaller statistics. The observation is reasonable because of the data generating mechanism and the extra numerical complication induced by the use of splines in [22]’s methodology. Meanwhile, we also consider the mean goodness of fit statistics for a set of Monte Carlo simulated data generated under [22]’s model. The data generating mechanism is similar to that in the first example, except that the conditional quantile function of the failure time is now

$$Q_T(\tau | \hat{Z}_t) = \exp \left\{ \int_0^{W_id} \{Q_\varepsilon(\tau) + 0.1 Z_{i1}(s) + 0.1 Z_{i2}(s)\} ds \right\},$$

where $\varepsilon$ has a $U(0, 1)$ distribution here. Table 3.3 reports that the mean goodness of fit statistics under such circumstances favor [22]’s model, which further validates the proposed statistic as a sound measure of the goodness of fit.
Table 3.3: Mean Goodness of Fit Statistics for the Monte Carlo Simulation ($\times 10^{-4}$)

<table>
<thead>
<tr>
<th></th>
<th>Proposed</th>
<th>[22]'s proposal</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\tau = 0.25$</td>
<td>$\tau = 0.50$</td>
</tr>
<tr>
<td><strong>Homogeneous Setup in the First Example</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$n = 200 &amp; 20%$ censoring</td>
<td>1.039</td>
<td>2.103</td>
</tr>
<tr>
<td>$n = 200 &amp; 40%$ censoring</td>
<td>0.560</td>
<td>1.098</td>
</tr>
<tr>
<td>$n = 500 &amp; 20%$ censoring</td>
<td>0.418</td>
<td>0.785</td>
</tr>
<tr>
<td>$n = 500 &amp; 40%$ censoring</td>
<td>0.223</td>
<td>0.415</td>
</tr>
<tr>
<td><strong>Heterogeneous Setup in the Second Example</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$n = 200 &amp; 20%$ censoring</td>
<td>1.079</td>
<td>2.125</td>
</tr>
<tr>
<td>$n = 200 &amp; 40%$ censoring</td>
<td>0.556</td>
<td>1.087</td>
</tr>
<tr>
<td>$n = 500 &amp; 20%$ censoring</td>
<td>0.421</td>
<td>0.790</td>
</tr>
<tr>
<td>$n = 500 &amp; 40%$ censoring</td>
<td>0.218</td>
<td>0.410</td>
</tr>
<tr>
<td><strong>Data Generated under [22]'s Model</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$n = 200 &amp; 20%$ censoring</td>
<td>1.212</td>
<td>2.399</td>
</tr>
<tr>
<td>$n = 200 &amp; 40%$ censoring</td>
<td>0.613</td>
<td>1.166</td>
</tr>
<tr>
<td>$n = 500 &amp; 20%$ censoring</td>
<td>0.506</td>
<td>1.060</td>
</tr>
<tr>
<td>$n = 500 &amp; 40%$ censoring</td>
<td>0.237</td>
<td>0.466</td>
</tr>
</tbody>
</table>
Table 3.4: Estimation Results for ACTG 320 Dataset

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>95% CI</th>
<th>Estimate</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Proposed Method</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \hat{\beta}^{(0)} )</td>
<td>0.915</td>
<td>(−2.987, 4.817)</td>
<td>4.104</td>
<td>(2.710, 5.498)</td>
</tr>
<tr>
<td>( \hat{\beta}^{(1)} )</td>
<td>−10.470</td>
<td>(−17.996, −2.943)</td>
<td>−4.971</td>
<td>(−7.489, −2.452)</td>
</tr>
<tr>
<td>( \hat{\beta}^{(2)} )</td>
<td>8.908</td>
<td>(5.245, 12.571)</td>
<td>5.865</td>
<td>(3.717, 8.014)</td>
</tr>
<tr>
<td><strong>[18]'s proposal</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \hat{\beta}^{(0)} )</td>
<td>1.210</td>
<td>(−1.036, 3.455)</td>
<td>2.377</td>
<td>(0.165, 4.589)</td>
</tr>
<tr>
<td>( \hat{\beta}^{(1)} )</td>
<td>−0.563</td>
<td>(−3.449, 2.324)</td>
<td>1.041</td>
<td>(−2.242, 4.324)</td>
</tr>
<tr>
<td>( \hat{\beta}^{(2)} )</td>
<td>0.498</td>
<td>(−1.118, 2.114)</td>
<td>−0.701</td>
<td>(−2.610, 1.208)</td>
</tr>
</tbody>
</table>

3.5 Data Analysis

In this section, we apply the proposed methodology to the ACTG 320 dataset [19]. The study conducted by the AIDS Clinical Trials Group is a randomized clinical trial of patients infected with type-1 HIV. The enrolled patients are randomly assigned to a regimen of three drugs, namely indinavir, zidovudine (or stavudine), and lamivudine. The dataset considered consists of the individual records of drug compliance, the plasma HIV RNA levels and the CD4 cell counts of 337 patients.

In accordance with clinical practice, we define the event time \( T_i \) for the \( i \)-th patient to be the first occasion where his CD4 cell count exceeds 200. The observation is treated as censored if the CD4 cell count did not exceed 200 until the end of the study period. The censoring rate of the dataset is approximately 50%. Consider a three dimensional covariate process \( \mathbf{Z}_i(t) = (Z_{i0}(t), Z_{i1}(t), Z_{i2}(t))^T \), where \( Z_{i0}(t) \equiv 1 \) for all \( t \geq 0 \) is the intercept term, \( Z_{i1}(t) \) is an indicator function of whether the individual takes all three drugs at time \( t \), and \( Z_{i2}(t) \) is the plasma HIV RNA levels at time \( t \).

Table 3.4 reports the estimation results of our proposed method for \( \tau = 0.25, 0.5 \). The estimates
are obtained by solving (3.5) sequentially for \( \tau \) from 0 to 0.6 with a step size of 0.01. The confidence intervals are constructed using the perturbation based resampling method with \( B = 500 \). The estimated coefficient \( \hat{\beta}^{(1)} \) suggests that the regimen of three drugs has a negative effect on the two quantiles of the event time, i.e. the medication shortens the time to cure. Meanwhile, we also conclude via the estimate \( \hat{\beta}^{(2)} \) that the plasma HIV RNA level has a positive effect on the concerned quantiles of the event time, i.e. a higher virus level prolongs the time for the CD4 cell count to return to normal level. It is noteworthy that the high censoring rate of the dataset renders a challenging numerical problem for the extreme quantiles especially. In particular, the initial searching algorithm we adopted, i.e. [41], fails to produce reasonable estimates for the higher level of quantiles. Table 3.4 also compares the results with [18]’s estimates, which float around zero and do not appear to be statistically significant. The results of [22]’s method are not tabulated as in Section 3.4, because their estimates are regression coefficients of the spline approximated proxies, render direct interpretation of the parameters nontrivial. As [22] requires the entire covariate history for estimation, we assume the last observed value of the covariate process presist. Figure 3.1 shows the estimates \( \hat{\beta}(\tau) \) and their corresponding pointwise 95% confidence intervals against \( \tau \). We plot the graphs with a step size of 0.05 to allow a cleaner presentation.

To assess the goodness of fit, we use the statistic \( R_n(\tau) \) defined in Section 3.4. For \( \tau = 0.25, 0.5 \) respectively, the values \((\times 10^{-3})\) are 0.750, 1.009 for our proposal and 3.913, 2.802 for [22], which appears to support the claim that the proposed method provides a better fit to the dataset. Meanwhile, Figure 3.2 shows the estimated log quantile functions against \( \tau \) for the patient with identifier 23424 using our proposal and [22]’s method. The two subplots demonstrate that the proposed method provides a better fit of the quantile function than [22], in the sense that the estimated quantile function resembles the inherent rising trend of the quantile function better as \( \tau \) increases.
Figure 3.1: Estimates and Pointwise Confidence Intervals for the ACTG 320 Dataset
Figure 3.2: Estimated Log Quantile Functions for the Patient with Identifier 23424
3.6 Conclusion

In this paper, we propose a class of censored quantile regression model to incorporate time-dependent covariates that requires only conditionally independent censorship. We develop estimating equations based on a more standard set of assumptions that the failure time and the censoring time are conditionally independent given the covariates, hence allowing greater model flexibility compared to the literature. On top of that, our formation naturally includes the time-dependent characteristic of the covariate process in the proposed estimating equation, while existing method has to rely on time-independent instrumental variable in constructing the corresponding estimating equation. In this perspective, our method provides a more straightforward estimation procedure that does not require spline approximation. The parameter estimates in our algorithm are computed recursively via a sequential approach, and thus the estimator can be viewed as a process of the quantile level. The performance of the proposed estimator has been demonstrated numerically to be highly competitive via Monte Carlo simulation and justified by its uniform consistency and weak convergence as a process. An interesting application of our procedure is also demonstrated via our real data example.

Our proposal can serve as a dynamic procedure of the censored quantile regression to handle time-dependent covariates under comparatively flexible model assumption. It is also an interesting question to further generalize the current formulation so that it can deal with survival data with more complex structure, say, length-biased data and case-cohort data, to name but a few. These problems will be handled in separate papers.

3.7 Appendix: Proofs

3.7.1 Uniform Consistency

Without loss of generality, we assume the \( \tau_j \)'s are equally spaced. Let \( a_n = \tau_{j+1} - \tau_j \), it is straightforward to show that \( H(\tau_{j+1}) - H(\tau_j) \leq a_n/(1 - \tau_U) = O(a_n) \). Also, Condition 3.1 implies that \( Z(t) \) is uniformly bounded, so we assume \( Z(t) \) is uniformly bounded by 1 without loss of
generality. By Taylor expansion of $\mu_1(\cdot)$, there exists $\beta'(\tau)$ in between $\hat{\beta}(\tau)$ and $\beta_0(\tau)$ such that

$$
\mu_1\{\hat{\beta}(\tau)\} - \mu_1\{\beta_0(\tau)\} = \nu_1\{\beta'(\tau)\}\{\hat{\beta}(\tau) - \beta_0(\tau)\}.
$$

By Condition 3.7, $\nu_1\{\beta'(\tau)\}$ is invertible for $\beta'(\tau) \in \mathcal{B}(d_0)/\mathcal{B}(d_0, \tau_U)$, so we can write

$$
\hat{\beta}(\tau) - \beta_0(\tau) = \left[\nu_1\{\beta'(\tau)\}\right]^{-1}\left[\mu_1\{\hat{\beta}(\tau)\} - \mu_1\{\beta_0(\tau)\}\right].
$$

Now, we show that $\mu_1\{\hat{\beta}(\tau)\}$ converges to $\mu_1\{\beta_0(\tau)\}$ uniformly for $\tau \in [0, \tau_U)$. To begin with, recall $\beta_0(\tau)$ satisfies

$$
\mu_1\{\beta_0(\tau)\} - \int_0^{\tau_j} \mu_2\{\beta_0(\tau)\} \, dH(u) = 0. \quad (3.9)
$$

Note that $\{Z_i(\tilde{T}_i)N_i\{1; b\}, b \in \mathcal{B}(d_0)\}$ is a Glivenko-Cantelli class, because the class of indicator functions is Glivenko-Cantelli and $Z_i(t)$ is uniformly bounded. Also, since $Z_i(t)$ is of finite variation by Condition 3.1, we can decompose $Z_i(t) = Z_i^+(t) - Z_i^-(t)$ such that $Z_i^\pm(t)$ are both non-negative and monotone increasing. We write $Z_i(t)Y_i(t) = Z_i^+(t)Y_i(t) - Z_i^-(t)Y_i(t)$, then both terms on the right hand sides are Glivenko-Cantelli because they are products of monotone functions, and the class of monotone functions is Glivenko-Cantelli. As a result, $\{Z_i(t)Y_i(t), t \geq 0\}$ is also Glivenko-Cantelli. It follows that $\{Z_i\{h_i^{-1}(1; b)\}Y_i\{1; b\}, b \in \mathcal{B}(d_0)\}$ is a Glivenko-Cantelli class. Using the Glivenko-Cantelli theorem, we have

$$
\sup_j \left\|\tilde{\mu}_1\{\hat{\beta}(\tau_j)\} - \mu_1\{\hat{\beta}(\tau_j)\}\right\| - \int_0^{\tau_j} \left[\tilde{\mu}_2\{\hat{\beta}(u)\} - \mu_2\{\hat{\beta}(u)\}\right] \, dH(u) < C_1,
$$

for any given $C_1 > 0$ with probability 1 when $n$ is large enough. On the other hand, by the definition of generalized solution [17] and the boundedness of $Z(t)$, the estimator $\hat{\beta}(\tau_j)$ satisfies

$$
\sup_j \left\|\hat{\mu}_1\{\hat{\beta}(\tau_j)\} - \int_0^{\tau_j} \hat{\mu}_2\{\hat{\beta}(u)\} \, dH(u)\right\| \leq \frac{1}{n}.
$$
Combining the last two displays, it follows that

\[
\left\| \mu_1(\hat{\beta}(\tau_j)) - \int_0^{\tau_j} \mu_2(\hat{\beta}(u)) \, dH(u) \right\| \leq C_1 + \frac{1}{n}. \tag{3.10}
\]

Meanwhile, recall from Condition 3.5 that \( \mu_1(\beta_0(\tau)) \) is a Lipschitz function of \( \tau \), so for \( j = 0, 1, \ldots, L - 1 \), we have

\[
\sup_{\tau \in [\tau_j, \tau_{j+1})} \left\| \mu_1(\beta_0(\tau)) - \mu_1(\beta_0(\tau_j)) \right\| \leq \sup_{\tau \in [\tau_j, \tau_{j+1})} C_3|\tau - \tau_j| = C_3|\tau_{j+1} - \tau_j| = C_3a_n. \tag{3.11}
\]

On the other hand, since \( Z(t) \) and indicator function are bounded by 1, we have

\[
\int_0^{\tau_1} \left\| \mu_2(\hat{\beta}(u)) - \mu_2(\beta_0(u)) \right\| \, dH(u) \leq \int_0^{\tau_1} 2 \, dH(u) = 2H(\tau_1) \leq \frac{2a_n}{1 - \tau_U}.
\]

Moreover, Conditions 3.6 and 3.7 imply that there exists a constant \( C_4 \) such that

\[
\left\| \nu_2(\beta''(u)) \left[ \nu_1(\beta'(u)) \right]^{-1} \mathbf{x} \right\| \leq C_4\|\mathbf{x}\|
\]

for any \( \mathbf{x} \in \mathbb{R}^p \) and \( \beta'(u), \beta''(u) \in B(d_0)/B(d_0, \tau_L) \). By Taylor expansion, we can write

\[
\begin{align*}
\int_0^{\tau_j} \left\| \mu_2(\hat{\beta}(u)) - \mu_2(\beta_0(u)) \right\| \, dH(u) & \leq \frac{2a_n}{1 - \tau_U} + \int_0^{\tau_1} \left\| \nu_2(\beta''(u)) \left[ \nu_1(\beta'(u)) \right]^{-1} \left[ \mu_1(\hat{\beta}(u)) - \mu_1(\beta_0(u)) \right] \right\| \, dH(u) \\
& \leq \frac{2a_n}{1 - \tau_U} + \sum_{k=1}^{j-1} \int_{\tau_k}^{\tau_{k+1}} C_4 \sup_{u \in [\tau_k, \tau_{k+1})} \left\| \mu_1(\hat{\beta}(u)) - \mu_1(\beta_0(u)) \right\| \, dH(u) \\
& \leq \frac{2a_n}{1 - \tau_U} + \frac{C_4a_n}{1 - \tau_U} \sum_{k=1}^{j-1} \sup_{u \in [\tau_k, \tau_{k+1})} \left\| \mu_1(\hat{\beta}(u)) - \mu_1(\beta_0(u)) \right\|. \tag{3.12}
\end{align*}
\]

Combining (3.9) to (3.12) and using the definition of \( \hat{\beta}(\tau) \), we have

\[
\sup_{\tau \in [\tau_0, \tau_1]} \left\| \mu_1(\hat{\beta}(\tau)) - \mu_1(\beta_0(\tau)) \right\| \leq \left\| \mu_1(\hat{\beta}(\tau_0)) - \mu_1(\beta_0(\tau_0)) \right\| + C_3a_n = C_3a_n,
\]
and for $j = 1, \ldots, L - 1$,

$$
\sup_{\tau \in [\tau_j, \tau_{j+1}]} \| \mu_1 \{ \hat{\beta}(\tau) \} - \mu_1 \{ \beta_0(\tau) \} \| \leq \| \mu_1 \{ \hat{\beta}(\tau_j) \} - \mu_1 \{ \beta_0(\tau_j) \} \| + C_3 a_n
$$

$$
\leq C_1 + \frac{1}{n} + C_3 a_n + \frac{2a_n}{1 - \tau_U} + \frac{C_4 a_n}{1 - \tau_U} \sum_{j=1}^{j-1} \sup_{\tau \in [\tau_k, \tau_{k+1}]} \| \mu_1 \{ \hat{\beta}(u) \} - \mu_1 \{ \beta_0(u) \} \|.
$$

Denote $\varepsilon_0 = C_3 a_n$ and $\varepsilon_j = C_1 + \frac{1}{n} + C_3 a_n + \frac{2a_n}{1 - \tau_U} + \frac{C_4 a_n}{1 - \tau_U} \sum_{k=0}^{j-1} \varepsilon_k$ for $j = 1, \ldots, L - 1$, then $\varepsilon_j$ will be an upper bound for $\sup_{\tau \in [\tau_j, \tau_{j+1}]} \| \mu_1 \{ \hat{\beta}(\tau) \} - \mu_1 \{ \beta_0(\tau) \} \|$. [41] shows that $\varepsilon_j$ can be bounded uniformly by $C_1 \exp \{ \frac{C_4 \tau_U}{1 - \tau_U} \}$ for $j = 0, 1, \ldots, L - 1$. Note that for large enough $n$, we may choose small enough $C_1$ such that $C_1 \exp \{ \frac{C_4 \tau_U}{1 - \tau_U} \} < d_0$, therefore $\hat{\beta}(\tau_j) \in \mathcal{B}(d_0)$ for $j = 0, 1, \ldots, L - 1$ by definition of $\mathcal{B}(d_0)$. Moreover, since $C_1$ can be arbitrarily small as $n$ grows, we have $\sup_{\tau \in [0, \tau_U]} \| \mu_1 \{ \hat{\beta}(\tau) \} - \mu_1 \{ \beta_0(\tau) \} \| \to_p 0$ as $n \to \infty$.

Using Condition 3.7, we have for $\tau \in [\tau_L, \tau_U]$ where $\tau_L \in (0, \tau_U]$,

$$
\| \hat{\beta}(\tau) - \beta_0(\tau) \| \leq \left( \| v_1 \{ \beta' (\tau) \} \|^{-1} \right) \| \mu_1 \{ \hat{\beta}(\tau) \} - \mu_1 \{ \beta_0(\tau) \} \| 
\leq C_5 \| \mu_1 \{ \hat{\beta}(\tau) \} - \mu_1 \{ \beta_0(\tau) \} \|.
$$

Note that $C_5$ does not depend on $\tau$, so the uniform consistency follows immediately.

### 3.7.2 Weak Convergence

We begin by showing two results on tightness, namely

$$
\sup_{\tau \in (0, \tau_U]} n^{1/2} \| \hat{\mu}_1 \{ \hat{\beta}(\tau) \} - \mu_1 \{ \beta_0(\tau) \} \| \to_p 0, \quad (3.13)
$$

and

$$
\sup_{\tau \in (0, \tau_U]} n^{1/2} \int_0^\tau \| \hat{\mu}_2 \{ \hat{\beta}(u) \} - \mu_2 \{ \beta_0(u) \} \| dH(u) \to_p 0. \quad (3.14)
$$
In line with the arguments of [2] and [32], define

$$\sigma_1^2(b) = \text{Var}[Z(\tilde{T})N(1; b) - Z(\tilde{T})N\{1; \beta_0(\tau)\} - \mu_1(b) + \mu_1\{\beta_0(\tau)\}]$$

and

$$\sigma_2^2(b) = \text{Var}[Z\{h^{-1}(1; b)\}Y(1; b) - Z\{h^{-1}(1; \beta_0(\tau))\}Y\{1; \beta_0(\tau)\} - \mu_2(b) + \mu_2\{\beta_0(\tau)\}].$$

To prove (3.13), it suffices to show the diagonal elements of $\sup_{\tau \in (0, \tau_1]} \sigma_1^2(\hat{\theta}(\tau)) \to_p 0$. By considering each entry separately, we may assume without loss of generality that $Z(t)$ is one-dimensional in the proof of (3.13) and (3.14). To this end, note that

$$\sup_{\tau \in (0, \tau_1]} \sigma_1^2(\hat{\theta}(\tau)) = \max \left\{ \sup_{\tau \in (0, \tau_1]} \sigma_1^2(\hat{\theta}(\tau)), \sup_{\tau \in (\tau_1, \tau_1]} \sigma_1^2(\hat{\theta}(\tau)) \right\}.$$

Now, $Z(t)$ is uniformly bounded by 1 implies that $Z^2(\tilde{T}) \leq Z(\tilde{T})$, so we have

$$\sup_{\tau \in (0, \tau_1]} \sigma_1^2(\hat{\theta}(\tau)) \leq \sup_{\tau \in (0, \tau_1]} E \left[ Z(\tilde{T})N\{1; \hat{\theta}(\tau)\} - Z(\tilde{T})N\{1; \beta_0(\tau)\} \right]^2$$

$$\leq \sup_{\tau \in (0, \tau_1]} E \left[ Z(\tilde{T})N\{1; \hat{\theta}(\tau)\} + Z(\tilde{T})N\{1; \beta_0(\tau)\} \right]$$

$$\leq \sup_{\tau \in (0, \tau_1]} \left[ \left| \mu_1\{\hat{\theta}(\tau)\} - \mu_1\{\beta_0(\tau)\} \right| + 2\mu_1\{\beta_0(\tau)\} \right]$$

$$\leq \sup_{\tau \in (0, \tau_1]} \left| \mu_1\{\hat{\theta}(\tau)\} - \mu_1\{\beta_0(\tau)\} \right| + 2C_3a_n \to_p 0,$$

where the last line follows from (3.11) and the result shown in the proof of uniform consistency.
that \( \sup_{\tau \in (0, \tau_U)} |\mu_1\{\hat{\beta}(\tau)\} - \mu_1\{\beta_0(\tau)\}| \to_p 0 \). On the other hand, we have

\[
\sup_{\tau \in (\tau_1, \tau_U]} \sigma^2_1 \{\hat{\beta}(\tau)\} \leq \sup_{\tau \in (\tau_1, \tau_U]} E \left[ N\{1; \hat{\beta}(\tau)\} - N\{1; \beta_0(\tau)\} \right]^2 \\
\leq \sup_{\tau \in (\tau_1, \tau_U]} E \left| Y\{1; \hat{\beta}(\tau)\} - Y\{1; \beta_0(\tau)\} \right|^2 \\
\leq \sup_{\tau \in (\tau_1, \tau_U]} E \left| Y\{1; \hat{\beta}(\tau)\} - Y\{1; \beta_0(\tau)\} \right| \to_p 0
\]

because of Condition 3.3. It follows that \( \sup_{\tau \in (0, \tau_U]} \sigma^2_1 \{\hat{\beta}(\tau)\} \to_p 0 \) and hence (3.13) holds.

To show (3.14), recall that \( n^{1/2} a_n \to 0 \), therefore

\[
\sup_{\tau \in (0, \tau_1]} n^{1/2} \int_0^\tau \left\| \hat{\mu}_2\{\hat{\beta}(u)\} - \hat{\mu}_2\{\beta_0(u)\} - \mu_2\{\hat{\beta}(u)\} + \mu_2\{\beta_0(u)\} \right\| \, dH(u) \\
\leq \sup_{\tau \in (0, \tau_1]} n^{1/2} \int_0^\tau 4 \, dH(u) = 4n^{1/2} H(\tau_1) \leq \frac{4n^{1/2} a_n}{1 - \tau_U} \to_p 0,
\]

which in turn implies that (3.14) is equivalent to

\[
\sup_{\tau \in (\tau_1, \tau_U]} n^{1/2} \int_{\tau_1}^\tau \left\| \hat{\mu}_2\{\hat{\beta}(u)\} - \hat{\mu}_2\{\beta_0(u)\} - \mu_2\{\hat{\beta}(u)\} + \mu_2\{\beta_0(u)\} \right\| \, dH(u) \to_p 0.
\]

Furthermore, note that

\[
\sup_{\tau \in (\tau_1, \tau_U]} n^{1/2} \int_{\tau_1}^\tau \left\| \hat{\mu}_2\{\hat{\beta}(u)\} - \hat{\mu}_2\{\beta_0(u)\} - \mu_2\{\hat{\beta}(u)\} + \mu_2\{\beta_0(u)\} \right\| \, dH(u) \\
\leq \sup_{\tau \in (\tau_1, \tau_U]} n^{1/2} \left\| \hat{\mu}_2\{\hat{\beta}(\tau)\} - \hat{\mu}_2\{\beta_0(\tau)\} - \mu_2\{\hat{\beta}(\tau)\} + \mu_2\{\beta_0(\tau)\} \right\| \, H(\tau_U),
\]

so it is equivalent to show that

\[
\sup_{\tau \in (\tau_1, \tau_U]} n^{1/2} \left\| \hat{\mu}_2\{\hat{\beta}(\tau)\} - \hat{\mu}_2\{\beta_0(\tau)\} - \mu_2\{\hat{\beta}(\tau)\} + \mu_2\{\beta_0(\tau)\} \right\| \to_p 0.
\]

To this end, it suffices to show that \( \sup_{\tau \in (\tau_1, \tau_U]} \sigma^2_2 \{\hat{\beta}(\tau)\} \to_p 0 \) again in the spirit of [2] and [32].
Now,

\[
{\sigma^2_2(\widehat{\beta}(\tau)) \leq E \left[ Z\{h^{-1}(1; \widehat{\beta}(\tau))\}Y\{1; \widehat{\beta}(\tau)\} - Z\{h^{-1}(1; \beta_0(\tau))\}Y\{1; \beta_0(\tau)\} \right]^2}
\]

\[
\leq 2E \left[ \left( Z\{h^{-1}(1; \widehat{\beta}(\tau))\} - Z\{h^{-1}(1; \beta_0(\tau))\} \right)^2 \right] Y\{1; \widehat{\beta}(\tau)\}
\]

\[
+ 2E \left( Z^2\{h^{-1}(1; \beta_0(\tau))\} \right) \left[ Y\{1; \widehat{\beta}(\tau)\} - Y\{1; \beta_0(\tau)\} \right]^2
\]

\[
\leq 2E \left[ Z\{h^{-1}(1; \widehat{\beta}(\tau))\} - Z\{h^{-1}(1; \beta_0(\tau))\} \right]^2 + 2E \left[ Y\{1; \widehat{\beta}(\tau)\} - Y\{1; \beta_0(\tau)\} \right]^2
\]

\[
\leq 4E \left[ Z\{h^{-1}(1; \widehat{\beta}(\tau))\} - Z\{h^{-1}(1; \beta_0(\tau))\} \right] + 2E \left[ Y\{1; \widehat{\beta}(\tau)\} - Y\{1; \beta_0(\tau)\} \right].
\]

The first term on the last line converges uniformly to 0 over \( \tau \in (\tau_1, \tau_U) \) because of Condition 3.2 and the uniform consistency of \( \widehat{\beta}(\tau) \), while the expectation in the second term also converges uniformly to 0 over \( \tau \in (\tau_1, \tau_U) \) because of Condition 3.3. It follows that \( \sup_{\tau \in (\tau_1, \tau_U)} \sigma^2_2(\widehat{\beta}(\tau)) \rightarrow_p 0 \) and hence (3.14) is true.

Let \( o_{p,I}(a_n) \) denote a term that converges to 0 uniformly for \( \tau \in I \) in probability after being divided by \( a_n \), and \( O_{p,I}(a_n) \) denote a term that is stochastically bounded uniformly for \( \tau \in I \) after being divided by \( a_n \). By the definition of \( \widehat{\beta}(\tau) \) and \( S_n(\widehat{\beta}, \tau) \), we have

\[
\sup_{\tau \in [\tau_j, \tau_{j+1})} n^{1/2} \left\| S_n(\widehat{\beta}, \tau) - S_n(\beta_0, \tau) \right\| = \sup_{\tau \in [\tau_j, \tau_{j+1})} n^{1/2} \left\| \int_{\tau_j}^{\tau} Z_i\{h_i^{-1}(1; \widehat{\beta}(u))\}Y_i\{1; \widehat{\beta}(u)\} \, dH(u) \right\|
\]

\[
\leq n^{1/2} \{ H(\tau_{j+1}) - H(\tau_j) \} = O_p(n^{1/2} a_n).
\]

Recall that \( n^{1/2} a_n \rightarrow 0 \), therefore \( n^{1/2} S_n(\widehat{\beta}, \tau) = o_p(0, \tau_U) \). Coupled with equations (3.13) and (3.14), we have

\[
-n^{1/2} S_n(\beta_0, \tau) = n^{1/2} \left[ \mu_1(\widehat{\beta}(\tau)) - \mu_1(\beta_0(\tau)) \right]
\]

\[
- n^{1/2} \int_0^\tau \left[ \mu_2(\widehat{\beta}(\tau)) - \mu_2(\beta_0(\tau)) \right] \, dH(u) + o_p(0, \tau_U)(1)
\]

\[
= n^{1/2} \left[ \mu_1(\widehat{\beta}(\tau)) - \mu_1(\beta_0(\tau)) \right]
\]

\[
- \int_0^\tau \left[ \nu_2(\beta_0(u)) \nu_1(\beta_0(u)) \right]^{-1} + o_p(0, \tau_U)(1) \times
\]

\[
n^{1/2} \left[ \mu_1(\beta_0(\tau)) - \mu_1(\beta_0(\tau)) \right] \, dH(u) + o_p(0, \tau_U)(1).
\]
The above equation is a stochastic differential equation for \( n^{1/2} \left[ \mu_1 \{ \hat{\beta}(\tau) \} - \mu_1 \{ \beta_0(\tau) \} \right] \), so using the production integration theory [4], we obtain

\[
n^{1/2} \left[ \mu_1 \{ \hat{\beta}(\tau) \} - \mu_1 \{ \beta_0(\tau) \} \right] = \phi \left\{ -n^{1/2} S_n(\beta_0, \tau) \right\} + o_{p,(0, \tau_U)}(1),
\]

where \( \phi : \mathcal{F} \to \mathcal{F} \) is such that

\[
\phi(g)(\tau) = \int_0^\tau \prod_{\mu \in \{s, \tau\}} \left( I_p + \nu_1 \{ \beta_0(u) \} \right) \left[ \nu_2 \{ \beta_0(u) \} \right]^{-1} dH(u) \, dg(s),
\]

for any \( g \in \mathcal{F} = \{ g : [0, \tau_U] \to \mathcal{R}^p, g \) is cadlag, \( g(0) = 0 \}. \)

Using similar arguments as in the proof of uniform consistency, and note that the class of indicator functions and the class of monotone functions are Donsker, we could show that

\[
\left\{ Z_i(\hat{T}_i) N_i \{ 1; \beta_0(\tau) \} - \int_0^\tau Z_i \{ 1; \beta_0(u) \} Y_i \{ 1; \beta_0(u) \} dH(u), \quad \tau \in (0, \tau_U) \right\}
\]

is a Donsker class. Indeed, \( Z_i(\hat{T}_i) \) is bounded and \( N_i \{ 1; \beta_0(\tau) \} \) is a class of indicator functions, therefore \( \{ Z_i(\hat{T}_i) N_i \{ 1; \beta_0(\tau) \}, \tau \in (0, \tau_U) \} \) is Donsker. Also, by decomposing \( Z_i(t) = Z_i^+(t) - Z_i^-(t) \), we see that \( \int_0^\tau Z_i^\pm \{ 1; \beta_0(u) \} Y_i \{ 1; \beta_0(u) \} dH(u) \) are both monotone functions of \( \tau \), hence \( \int_0^\tau Z_i \{ 1; \beta_0(u) \} Y_i \{ 1; \beta_0(u) \} dH(u), \tau \in (0, \tau_U) \} \) is Donsker. As a result, the above claim is true. By the Donsker theorem, we conclude that \( -n^{1/2} S_n(\beta_0, \tau) \) converges weakly to a tight Gaussian process \( G(\tau) \) for \( \tau \in (0, \tau_U) \) with mean 0 and covariance \( E \left[ \kappa_i(\beta_0, \tau) \kappa_i(\beta_0, \tau)^T \right] \), where \( \kappa_i(\beta, \tau) = Z_i(\hat{T}_i) N_i \{ 1; \beta(\tau) \} - \int_0^\tau Z_i \{ h_i^{-1}(1; \beta(u)) \} Y_i \{ 1; \beta(u) \} dH(u) \). It follows that \( \phi \{ G(\tau) \} \) is also a Gaussian process for \( \tau \in (0, \tau_U) \), because \( \phi \) is a linear operator. Finally, since \( \left[ \nu_1 \{ \beta_0(\tau) \} \right]^{-1} \) is uniformly bounded for \( \tau \in [\tau_L, \tau_U] \) where \( \tau_L \in (0, \tau_U) \) by Condition 3.7, and with the aid of Taylor expansion and continuous mapping theorem, we obtain that \( n^{1/2} \{ \hat{\beta}(\tau) - \beta_0(\tau) \} \) converges to a Gaussian process \( \left[ \nu_1 \{ \beta_0(\tau) \} \right]^{-1} \phi \{ G(\tau) \} \) for \( \tau \in [\tau_L, \tau_U] \).
3.7.3 Justification of Resampling Method

Following the arguments in Section 3.7.1 and Section 3.7.2, and since $E(\xi) = 1$, we can similarly obtain that $n^{1/2}S_n^*(\beta^*, \tau) = o_{p,(0,\tau_U)}(1)$ and

$$n^{1/2}\{\beta^*(\tau) - \hat{\beta}(\tau)\} = \left[V_1(\beta_0(\tau))\right]^{-1}\phi\left\{-n^{1/2}S_n^*(\hat{\beta}, \tau)\right\} + o_{p,(0,\tau_U)}(1).$$  \hfill (3.15)

Recall that $n^{1/2}S_n(\hat{\beta}, \tau) = o_{p,(0,\tau_U)}(1)$, so we have

$$-n^{1/2}S_n^*(\hat{\beta}, \tau) = n^{1/2}\{S_n(\hat{\beta}, \tau) - S_n^*(\hat{\beta}, \tau)\} + o_{p,(0,\tau_U)}(1)$$

$$= n^{1/2}\frac{1}{n} \sum_{i=1}^{n} (1 - \xi_i) \kappa_i(\hat{\beta}, \tau) + o_{p,(0,\tau_U)}(1).$$ \hfill (3.16)

Now, since $Var(\xi_i) = 1$, we have for $\tau_s, \tau_t \in [\tau_L, \tau_U]$,

$$E\left[S_n^*(\hat{\beta}, \tau_s)S_n^*(\hat{\beta}, \tau_t)^\top \mid \{\tilde{T}_i, \Delta_i, \tilde{Z}_i(\tilde{T}_i)\}_{i=1}^n\right]$$

$$= \frac{1}{n} \sum_{i=1}^{n} \kappa_i(\hat{\beta}; \tau_s)\kappa_i(\hat{\beta}; \tau_t)^\top \rightarrow_p E\left[\kappa_i(\beta_0, \tau_s)\kappa_i(\beta_0, \tau_t)^\top \right].$$

As demonstrated in [36] and [41], the limiting distribution of $-n^{1/2}S_n^*(\hat{\beta}, \tau)$ is the same as the limiting distribution of $-n^{1/2}S_n(\beta_0, \tau)$ for almost all realizations of the data $\{\tilde{T}_i, \Delta_i, \tilde{Z}_i(\tilde{T}_i)\}_{i=1}^n$. It follows from (3.15) and (3.16) that given the observed data $\{\tilde{T}_i, \Delta_i, \tilde{Z}_i(\tilde{T}_i)\}_{i=1}^n$, the limiting conditional distribution of $n^{1/2}\{\beta^*(\tau) - \hat{\beta}(\tau)\}$ is asymptotically equivalent to the limiting unconditional distribution of $n^{1/2}\{\hat{\beta}(\tau) - \beta_0(\tau)\}$. 
Chapter 4: Discussion

The preceding chapters discuss two semiparametric methods for censored data with time-dependent covariates. We first propose a generalization of the accelerate failure time model that allows for localized covariate effects over designated quantile brackets, which serves as a flexible alternative when one is interested in a particular subclass of the population. The estimator is obtained via a rank-based procedure, where its consistency and normality have been established. Applying the proposed method to the renowned Stanford heart transplant data, we observe different covariate effects across different subregions of quantiles, which justifies the results of the Cox proportional hazards model and the accelerated failure time model. Meanwhile, we propose a censored quantile regression model with time-dependent covariates that examines pointwise covariate effects as in ordinary quantile regression, while internal covariates can be incorporated in our model because of the generalized quantile formation. Exploiting the correspondence between the hazard-based martingale structure and the quantile process, we derive a functional estimating equation without resorting to the rather restrictive independent censoring assumption. The procedure naturally requests only the history of the time-dependent covariates up to the failure time instead of the entire covariate processes. A recursive estimator for the regression parameters is provided with its uniform consistency and weak convergence established via empirical process theory.

The proposed quantile formation can serve as a building block for survival analysis with time-dependent covariates, so that other quantile regression models for time-independent covariates can be extended accordingly. In particular, the recursive estimator proposed can be computational intensive and requires a linear quantile model globally for all quantiles. Analogous to [43] and [53], we may use a kernel-based method to approximate the unknown conditional hazard function so that a recursive algorithm and hence the globally linear assumption are not needed. On the
other hand, while we only consider the data generating mechanism for external covariates in the Monte Carlo simulation, the scenario for internal covariates shall be explored in the future. In particular, as the quantile function is no longer ordinary for internal covariates, we shall transform to the hazard domain and stipulate the relationship between the hazard function and the internal covariates. Finally, both proposals can be generalized to handle various types of survival data under general biased sampling schemes, such as length-biased data and case-cohort design. For time-independent covariates, such extensions of the accelerated failure time model and the censored quantile regression model to biased sampled data are done in [28] and [57], for example.
References


