QUANTILE REGRESSION MODELS FOR INTERVAL-CENSORED FAILURE TIME DATA

Fang-Shu Ou

A dissertation submitted to the faculty at the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Biostatistics in the Gillings School of Global Public Health.

Chapel Hill
2015

Approved by:

Jianwen Cai

David Couper

Gerardo Heiss

Donglin Zeng

Haibo Zhou

ABSTRACT

Fang-Shu Ou: Quantile Regression Models for Interval-censored Failure Time Data
(Under the direction of Jianwen Cai and Donglin Zeng)

Quantile regression models the conditional quantile as a function of independent variables providing a complete association between the response and predictors. Quantile regression can describe the association at different quantiles yielding more information than the least squares method which only detects associations with the conditional mean. Quantile regression models have gained popularity in many disciplines including medicine, finance, economics, and ecology as they can accommodate heteroscedasticity.

A specific type of failure time data is called interval-censored where the failure time is only known to have occurred between certain observation times. Such data appears commonly in medical or longitudinal studies because disease onset is known to have occurred between scheduled visits but the exact time is unknown. Quantile regression has been extended to survival analysis with random censoring time. Most methods focus on survival analysis with right-censored data while a few were developed for data with other censoring mechanisms. Despite the fact that the development for censored quantile regression flourishes, limited work has been done to handle interval-censored failure time data under the quantile regression framework.

In this dissertation, we developed a new method to analyze interval-censored failure time data using conditional quantile regression models. Our method can handle both Case I and Case II interval-censored data and allow the censoring time to depend on covariates. We developed an estimation procedure that is computationally efficient and easy to implement with inference performed using a subsampling method. The consistency and asymptotic distribution of the resulting estimators were established using modern empirical process theory. The developed method was extended as a computational tool to analyze interval-censored data for accelerated failure time models. The estimators from different quantiles were combined to increase the efficiency of the estimators. The small sample performances were demonstrated via simulation studies. The proposed methods were illustrated with current status datasets, data from the Voluntary HIV-1 Counseling and Testing Efficacy Study Group and calcification study, and Case II interval-censored data, data from the Atherosclerosis Risk in Communities Study and breast cosmesis data.

To Nyan-Mei Wang, my mother and the source of unconditional love and support.
To Martin Andrew Heller, my husband and my best friend.
To Bucky and Mei-Mei, my Australian shepherd and old English sheepdog who provide endless entertainment.

ACKNOWLEDGMENTS

I would like to express my sincere gratitude to my advisers Dr. Jianwen Cai and Dr. Donglin Zeng, for their guidance and support throughout my graduate studies. They gave me the freedom to explore different possibilities, learn from my own mistakes, and pointed me in the right direction when I have gone astray. I cannot ask for better advisers than them. Additionally, I would like to thanks my committee members, Dr. David Couper, Gerardo Heiss, and Haibo Zhou, for their time and insightful comments.

I would like to offer thanks to my supervisors Drs. Daniela Sotres-Alvarez and Diane Catellier while working at the Collaborative Studies Coordinating Center and Dr. June Stevens in the Nutrition Department. They provided not only the financial support for my study but also the opportunity to work on interesting research projects.

Last but not the least, I want to extend my gratitude to one and all, who directly or indirectly, lent me their hands in this venture.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: INTRODUCTION

Regression analysis has typically been performed using the least squares method since the time of Gauss. The least squares method summarizes the relationship between the dependent variable and independent variables by a conditional mean function. When the homoscedasticity assumption fails; however, the least squares method cannot provide a complete picture of the relationship between the response and predictors.

Quantile regression, on the other hand, models the conditional quantile as a function of independent variables providing a solution when heteroscedasticity is present and offers a more detailed perspective regarding associations. The conditional median is less sensitive to skewness than the conditional mean making it a more reliable measure of the central tendency. When there is heteroscedasticity, quantile regression can describe the association at different quantiles yielding more information than the least squares method which only detects associations with the conditional mean. Quantile regression models have gained popularity in many disciplines including medicine, finance, economics, and ecology as they can accommodate heteroscedasticity. In survival analysis, ease of interpretation is particularly appealing since the estimates can be directly interpreted as the effect on survival time.

For failure time data, the distribution of the response variable tends to be highly right skewed with possible heterogeneity. In recent decades, quantile regression has been extended to survival analysis with random censoring time. Most methods focus on survival analysis with right-censored data while a few were developed for doubly censored data, left-truncated/right-censored data, and recurrent event data.

A specific type of failure time data is called interval-censored where the failure time is only known to have occurred between observation times. Such data appears commonly in medical or longitudinal studies because subjects schedule follow-ups where disease status is determined. The disease then is known to have manifested between scheduled visits but the exact time is unknown. There are two subtypes of interval-censored data, Case I and Case II. Case I interval-censored data refers to interval-censored failure time data where all observed intervals had either zero or infinity as one of the end points, i.e. all observations are either right- or left-censored. Case II interval-censored data is interval-censored failure time data where we know the event occurred either prior to the first observation time, between observation times, or after the final observation time. Despite the fact that the development for censored quantile regression flourishes, only one published manuscript discussed the method developed to handle interval-censored failure time data under the quantile regression framework. This existing publication did not provide any theoretical justification regarding the method proposed and can only handle categorical covariates in the model.

In this dissertation, we developed a new method to analyze interval-censored failure time data using conditional quantile regression models. Our method can handle both Case I and Case II interval-censored data and allow the censoring time to depend on covariates. We developed an estimation procedure that is computationally efficient and easy to implement with inference performed using a resampling based method. The consistency and asymptotic distribution of the resulting estimators will be established using modern empirical process theory. The small sample performances will be demonstrated via simulation studies and the methods developed will be illustrated with the females' data from The Voluntary HIV-1 Counseling and Testing Efficacy Study Group. The Voluntary HIV-1 Counseling and Testing Efficacy Study Group had two follow-up visits scheduled for all participants but the status of sexually transmitted disease was only determined in the first follow-up; therefore, it is a typical Case I interval censored failure time data.

The Atherosclerosis Risk in Communities Study is a prospective epidemiologic study with five follow-up visits. During each visit, the disease status, such as diabetes, hyperglycemia, and hypercholesterolemia, was determined using biomarkers. If a biomarker value exceed a certain threshold then the participant was diagnosed as having the disease. Accordingly, we only know the disease occurred between the last and the current follow-up but the exact onset time is unknown; thus, it is Case II interval-censored data. We extended our method developed for Case I interval-censored data to Case II data and applied it on the data from Atherosclerosis Risk in Communities Study.

The accelerated failure time (AFT) model is an alternative to the well established proportional hazards models. An AFT model assumes that the effect of a covariate is to accelerate or decelerate the course of a disease by some constant. Several methods have been developed for applying AFT models to interval-censored data; however, the implementations are computationally intensive and the asymptotic inference often involves nonparametric functional estimation. Under the AFT model assumption, the quantile estimates from different quantiles should have the same coefficients and differ only in intercepts. We used the quantile regression framework developed for interval-censored data to estimate the coefficients in AFT models then combine the estimates from different quantiles to increase efficiency of our estimates.

CHAPTER 2: LITERATURE REVIEW

2.1 Quantile Regression

The first attempt of regression analysis using least absolute deviations may be dated back to 1760 by the Croatian Jesuit Roger Boscovich who was interested in a problem concerning ellipticity of the earth and a geometric algorithm was proposed as the solution. Boscovich's proposed method is a peculiar hybrid of mean and median ideas, i.e. the intercept is estimated as a mean and the slope is estimated as a median. With the development of the least squares at the end of the 18th century, Boscovich's estimator faded into history. Until a century later, Francis Ysidro Edgeworth modified Boscovich's conditions and proposed to minimize the sum of absolute residuals. A geometric algorithm was developed for the bivariate case but the approach was rather awkward. The least absolute deviation method did not become practical on a large scale until the simplex algorithm for linear programming was developed in the late 1940s. Koenker (2005) and Portnoy et al. (1997) provide interesting historical introductions to least absolute deviations.

The τ^{th} quantile of a random variable, Y, is defined as

$$Q_\tau(Y) = \inf\{y : F_Y(y) \geq \tau\},$$

where $F_Y(y)$ denotes the cumulative distribution function of Y and $0 < \tau < 1$. Extend this ideal to a regression model with p-dimensional covariate $\boldsymbol{X}$ here the first component is one allowing an intercept in the model, the τ^{th} conditional quantile is defined as

$$Q_\tau(Y|\boldsymbol{X} = \boldsymbol{x}) = \inf\{y : F_Y(y|\boldsymbol{x}) \geq \tau\},$$

where $F_Y(y|\boldsymbol{x})$ denotes the conditional cumulative distribution function of Y given $\boldsymbol{X} = \boldsymbol{x}$. Similarly to the least squares method which fits the conditional means as a function of covariates, the linear quantile regression model fits the conditional quantile as a function of covariate, i.e.

$$Q_\tau(Y|\boldsymbol{X}) = \boldsymbol{X}'\boldsymbol{\beta}(\tau),$$

where $\boldsymbol{\beta}(\tau)$ is the quantile coefficient that may depend on τ. $\boldsymbol{\beta}(\tau)$ can be interpreted as the marginal change in the τ^{th} quantile caused by an increase in covariate values.

Several advantages of quantile regression are worth mentioning. Quantile regression provides a complete picture of the relationship between response variable and covariates; therefore, it can detect relationships which may be overlooked by the least squares method. It is robust to outliers in response variable and the estimation and inference are distribution-free. For example, Dunham et al. (2002) analyzed the relationship between the abundance of Lahontan cutthroat trout and the ratio of stream width to depth. While a least squares regression estimated no linear change in mean density across ratio, the quantile regression estimates shows a nonlinear, negative relationship of cutthroat trout densities across 13 streams and over 7 years in the upper quantiles (Figure 2.1).

Figure 2.2 shows a toy example to demonstrate the robustness of quantile regression when outliers are present. The data was generated from a linear regression model with iid normal error with one additional data point (solid dot) added as an outlier. The 0.5 quantile (median) estimate is denoted by the solid line and the least square estimate is denoted by the dash line. It is clear that the median fit was not influenced by the outlier.

Figure 2.3 and Figure 2.4 (taken from Koenker (2005)) show partial results from an

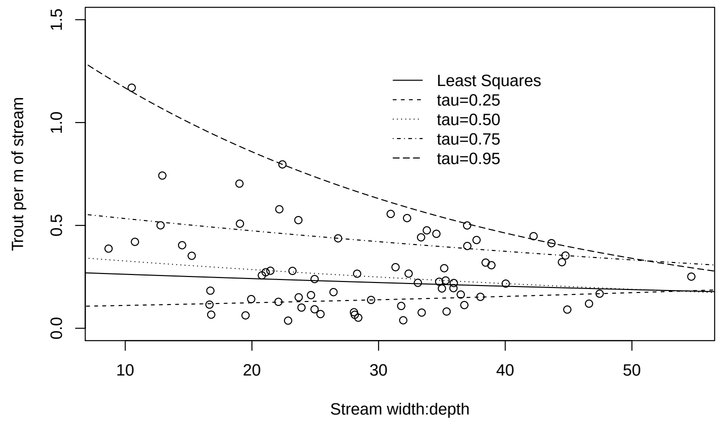

Figure 2.1: A scatterplot of 71 observations of stream width to depth ratio and trout densities with 0.95, 0.75, 0.50, and 0.25 quantiles (dash line) and least squares regression (solid line) estimates for the model $log(\text{trout densit}) = \beta_0 + \beta_1\text{width:depth} + \varepsilon$.

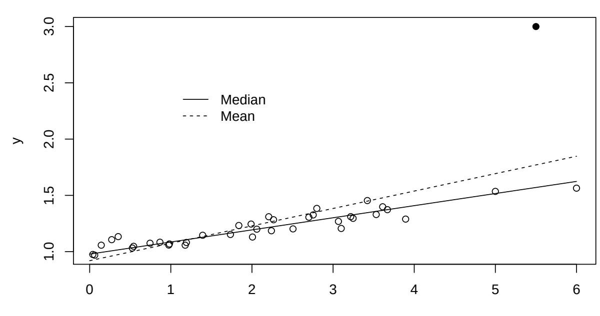

Figure 2.2: A toy example. A scatterplot of 38 observations and median (solid line) and least squares (dash line) estimates. The solid dot is served as an outlier.

investigation carried out by Abrevaya (2002). The outcome of interest was infant birth weight and the covariates included were various demographic characteristics and maternal behavior. Please see Koenker (2005) for a detailed description of the data source, covariates used, and quantiles estimated. The data has been centered to yield a reference group, a girl born to an unmarried, white mother with less than a high school education, who is 27 years old and had a weight gain of 30 pounds, did not smoke, and had her first prenatal visit in the first trimester of pregnancy. Since the data has been centered, the intercept of the model (far left panel of Figure 2.3) can be interpreted as the estimated conditional quantile function of the birth-weight distribution of the reference group. The median birth-weight of the reference group is about 3300 grams and the 5th percentile of the birth-weight is about 2500 grams which is the conventional definition of a low-birth-weight baby.

The far right panel of Figure 2.3 shows the difference in birth-weight of infants born to black versus white mothers. The birth-weight of infants born to black mothers is significantly less than those born to white mothers, especially in the lower tail of the distribution. At the 5th percentile of the conditional distribution, infants born to black mother are more than 300 grams lighter than infants born to white mother. The horizontal line indicates the results from ordinary least-squares which would conclude that the birth-weight of infants born to black mothers are about 200 grams less on average than those born to white mothers.

The mother's weight gain entered the model as a linear and a quadratic term. The two far right panels of Figure 2.4 show the effect of a mother's weight gain. Based on the results of ordinary least-squares (the horizontal line), there is a very minor quadratic effect for mother's weight gain and, on average, a 1 pound increase in mother's weight, leads to an infant birth-weight increase of 10 grams. Using quantile regression, we are able to see a

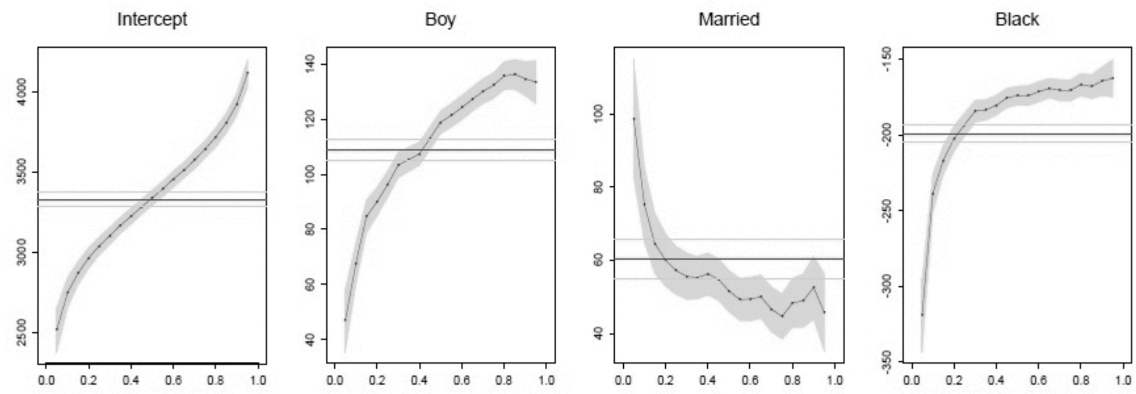

Figure 2.3: Quantile regression for birth.

more complete picture of the relationship between a mother's weight gain and infant birthweight. Figure 2.5 shows the marginal effect of mother's weight gain for all quantiles evaluated at four specific levels of mother's weight gain. These 4 levels are roughly the 10th, 25th, 75th, and 90th percentile of mother's weight gain.

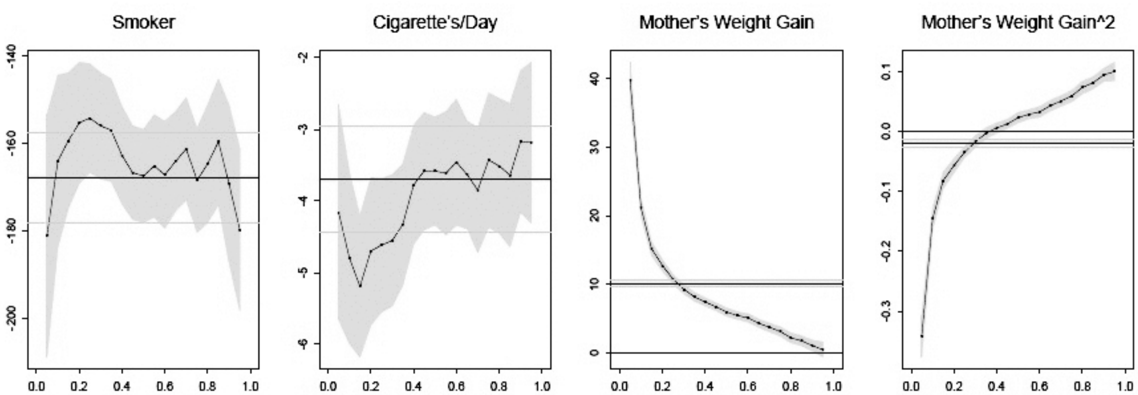

Figure 2.4: Quantile regression for birth (continued).

Conditional on low weight gain by the mother, the marginal effect of 1 pound increase of mother weight was about 30 grams at the lowest quantile and the effect diminished to only 5 grams at the higher quantiles (top-left panel of Figure 2.5). This relationship persisted at the slightly higher level of mother's weight gain but the marginal effect was not as pronounced (top-right panel of Figure 2.5). This declining marginal effects of mother's weight gain is minimal conditional on high weight gain by the mother (bottom two panels of Figure 2.5). For mothers who already gained about 40 pounds, each additional 1 pound of weight gain would only increase infant weight by about 5 to 10 grams. The ability to draw conclusions at a specific quantile of interest is the advantage of using quantile regression.

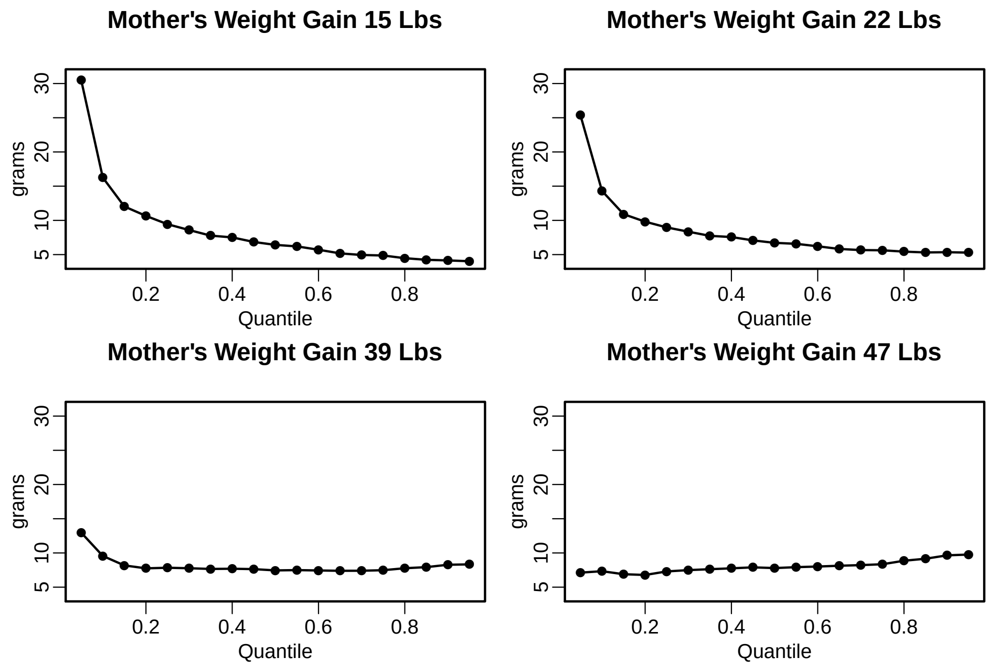

Figure 2.5: Effect of mother's weight gain.

Unlike the least squares method, quantile regression models are invariant to monotone transformations (Koenker 2005). Specifically, let $h(\cdot)$ be a monotone nondecreasing function on $\Re$, then for any random variable Y, $Q_{h(Y)}(\tau) = h(Q_Y(\tau))$; that is, the quantiles of the transformed random variable $h(Y)$ are simply the transformed quantiles of the original Y. This property is immediate from the elementary fact that, for any strictly monotone h, $P(Y \leq y) = P(h(Y) \leq h(y))$.

2.1.1 Estimation

Median regression, also known as L_1 regression, is an extension of the sample median when covariates are available. The solution for an observed sample is obtained by solving the following function,

$$\min_{\beta \in \Re^p} \sum_{i=1}^{n} |y_i - x_i'\beta|, \tag{2.1}$$

where $\{y_1, \cdots, y_n\}$ is the random sample and x_i is the covariates associate with random sample y_i where the first component of x_i is a constant one. Koenker and Bassett (1978) generalized the median regression to the τ^{th} quantile $(0 < \tau < 1)$ by simply replacing the absolute value in Equation (2.1) with the loss function, $\rho_\tau(\cdot)$, i.e.

$$\min_{\beta \in \Re^p} \sum_{i=1}^{n} \rho_\tau(y_i - x_i'\beta), \tag{2.2}$$

where the quantile loss function is defined as

$$\rho_\tau(u) = u(\tau - I(u < 0)) \tag{2.3}$$

The piecewise linear loss function, $\rho_\tau(\cdot)$ is illustrated in Figure 2.6.

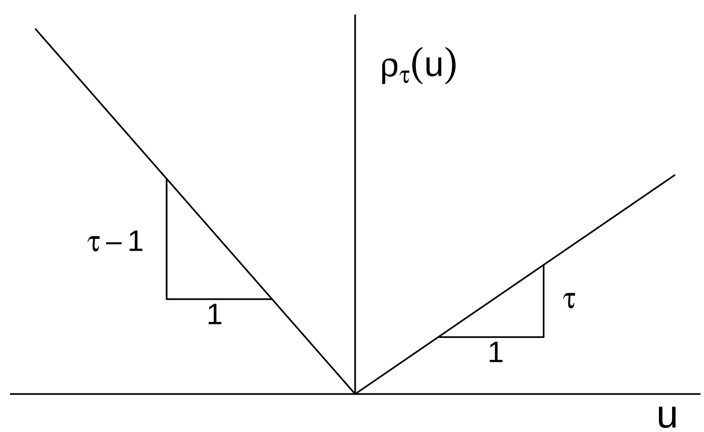

Figure 2.6: Quantile regression loss function

We now review several algorithms available to solve Equation (2.2) namely the simplex, interior point, and smoothing algorithms.

Simplex Algorithm Equation (2.2) is equivalent to

$$\min_{\boldsymbol{\beta}\in\Re^p}\left[\sum_{i\in\{i:y_i\geq \boldsymbol{x_i}\boldsymbol{\beta}\}}\tau|y_i-\boldsymbol{x_i'}\boldsymbol{\beta}|+\sum_{i\in\{i:y_i<\boldsymbol{x_i'}\boldsymbol{\beta}\}}(1-\tau)|y_i-\boldsymbol{x_i'}\boldsymbol{\beta}|\right]$$

$$=\min_{\boldsymbol{\beta},\boldsymbol{u},\boldsymbol{v}}\tau\mathbf{1_n'}\boldsymbol{u}+(1-\tau)\mathbf{1_n'}\boldsymbol{v} \tag{2.4}$$

$$\text{such that } \boldsymbol{y}-\boldsymbol{x'}\boldsymbol{\beta}=\boldsymbol{u}-\boldsymbol{v}, \boldsymbol{\beta}\in\Re^p, \text{ and } \boldsymbol{u}\geq 0, \boldsymbol{v}\geq 0,$$

where $\boldsymbol{y}=(y_1,\cdots,y_n)', \boldsymbol{x}=(x_1,\cdots,x_n)'$, $\mathbf{1_n}$ is a $(n\times 1)$ vector of the value 1, and $(\cdot)_+$ denotes the positive part. Taking this one step further, let $\boldsymbol{\phi}=(\boldsymbol{\beta})_+$, $\boldsymbol{\psi}=(-\boldsymbol{\beta})_+$, $\boldsymbol{B}=[\boldsymbol{x}\quad -\boldsymbol{x}\ \ \boldsymbol{I}\quad -\boldsymbol{I}]$, $\boldsymbol{\theta}=(\boldsymbol{\phi}',\ \ \boldsymbol{\psi}',\ \ \boldsymbol{u}',\ \ \boldsymbol{v}')'$, and $\boldsymbol{d}=(\mathbf{0}',\mathbf{0}',\tau\mathbf{1_n'},(1-\tau)\mathbf{1_n'})'$ where $\mathbf{0}'=(0,0,\cdots,0)_p$). We may reformulate the problem as a standard linear programming minimization problem with primal form:

$$\min_{\boldsymbol{\theta}}\boldsymbol{d}'\boldsymbol{\theta}\qquad \text{such that } \boldsymbol{B}\boldsymbol{\theta}=\boldsymbol{y},\ \boldsymbol{\theta}\geq 0.$$

Thus, the Equation (2.2) may be solved using the simplex method. The simplex method starts at a feasible vertex and travels from vertex to vertex along the edge of the polyhedral constraint set. The path is chosen by the steepest descent and the algorithm continues until arriving at the optimum.

In the special case when $\tau=0.5$, i.e. median regression, the simplex method algorithm developed by Barrodale and Roberts (1973) is often used for solving the optimization problem since it appears to be superior computationally to other algorithms. The Barrodale and Roberts algorithm is implemented in two stages. Stage 1 only chooses the columns in $\boldsymbol{x}$ or $-\boldsymbol{x}$ as pivotal column during the first p iterations. Stage 2 interchanges nonbasic columns with basic columns in $\boldsymbol{I}$ or $-\boldsymbol{I}$. The basic columns in $\boldsymbol{x}$ and $-\boldsymbol{x}$ are forced to remain in the basis during Stage 2. Stage 1 will be executed p times and Stage 2 will be executed until no suitable vector can enter or leave the basis. The Barrodale and

Roberts algorithm can be extended for any given quantile as described by Koenker and d'Orey (1987). The simplex method provides an extremely efficient solution to quantile regression when the dataset size is moderate, for example, less than 5000 observations and 50 covariates. However, the computational speed became unsatisfactory for large datasets.

Interior Point Algorithm To overcome the computational difficulty of the simplex method when the dataset is large, interior point algorithms for linear programming were applied to quantile regression. Instead of traveling along the exterior of the constraint set, Newton steps were taken within the interior of a deformed constraint set toward the boundary. Consider the canonical linear program

$$\min\{c'x \mid Ax = b, x \geq 0\} \tag{2.5}$$

and assume that there is a strictly feasible solution in the interior of the constraint set. One way to find the solution is to decrease $c'x$ while ensuring the boundary of the feasible set is not crossed. We can achieve this by augmenting the objective function by a logarithmic barrier term. Let

$$B(x, \mu) = c'x - \mu \sum log(x),$$

and we would minimize $B(x, \mu | Ax = b)$ while reducing μ to zero. The inequality constraints in Equation (2.5) is replaced by the penalized log barrier, thus minimizing $B(x, \mu | Ax = b)$ by taking the Newton steps

$$\min_q \left\{ c'q - \mu\, q' X^{-1}\, 1_n + \frac{1}{2}\, \mu\, q'\, X^{-2}\, q | A\, q = 0 \right\},$$

where $X = \operatorname{diag}(x) = \begin{pmatrix} x_1 & 0 & \cdots & 0 \\ 0 & x_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & x_p \end{pmatrix}$.

To use the logarithm barrier for quantile regression, we first rewrite the linear programming problem (Equation (2.4)) in its dual representation,

$$\max_{\lambda}\{y'\lambda \,|\, X'\lambda = 0, \lambda \in [\tau - 1, \tau]^n\},$$

or, by setting $a = \lambda + 1 - \tau$,

$$\max_{a}\{y'a \,|\, X'a = (1-\tau)X'1_n, a \in [0,1]^n\}. \tag{2.6}$$

Adding slack variables, s, such that $a + s = 1_n$, we have the barrier function

$$B(a, s, \mu) = y'a + \mu \sum_i (\log a_i + \log s_i),$$

and the Newton steps

$$\max_{\delta_a}\{y'\delta_a + \mu\delta_a'(A^{-1} - S^{-1})1_n - \frac{1}{2}\mu\delta_a'(A^{-2} + S^{-2})\delta_a\}$$

such that $X'\delta_a = 0$ where $A = \operatorname{diag}(a)$ and $S = \operatorname{diag}(s)$.

Since similar methods may be applied to both primal and dual formulations, attaching both formulations simultaneously actually improves the performance of the algorithm. Mehrotra (1992) implement the primal-dual log barrier approach successfully using the predictor-corrector step and it is extended for problems with free variables and problems with bounds on primal variables by Lustig et al. (1992). A detailed presentation of interior-point methods in quantile regression can be found in Portnoy et al. (1997).

Smoothing Algorithm When a dataset is large, another suitable estimation method is found through smoothing. The original non-differentiable objective function in Equation (2.2) can be approximated by a smooth Huber function (Huber 1973) to create the objective function,

$$\min_{\beta \in \Re^p} \sum_{i=1}^{n} H_{\gamma,\tau}(y_i - x_i'\beta), \tag{2.7}$$

where

$$H_{\gamma,\tau}(t) = \begin{cases} t(\tau-1) - \frac{1}{2}(\tau-1)^2\gamma & \text{if } t \leq (\tau-1)\gamma \\ t^2/(2\gamma) & \text{if } (\tau-1)\gamma \leq t \leq \tau\gamma \\ t\tau - \frac{1}{2}\tau^2\gamma & \text{if } t \geq \tau\gamma \end{cases}$$

and the γ is a positive real number referred to as the "threshold". $H_{\gamma,\tau}(\cdot)$ is a continuous differentiable function, as illustrated in Figure 2.7. The minimizer of Equation (2.7) is close to the minimizer of Equation (2.2) when γ is small and will produce the proper estimator before γ converge to zero. The smooth approximation method was originally developed by Madsen and Nielsen (1993) to solve linear L_1 estimation problems and was further extended by Chen (2007) to general quantile regression. The computational speed of the smoothed function is comparable to interior point method and is superior for a "fat" dataset, i.e. when the ratio of covariates to observations is greater than 0.05 and when $\boldsymbol{x}\boldsymbol{x}'$ is a non-sparse matrix (Chen 2007).

2.1.2 Inference

Methods to perform inference of quantile estimators can be separated into three types, namely direct estimation, inversion of a rank test, and resampling based methods. We will review the basic idea behind each method and compare their strengths and weaknesses.

Direct Estimation Consider the simplest case of

$$y_i = x_i'\beta + u_i,$$

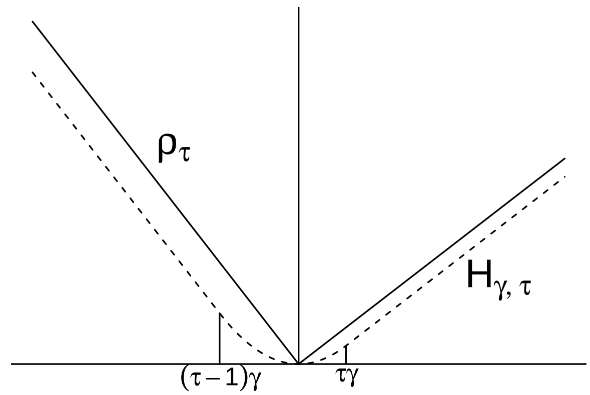

Figure 2.7: Huber function approximation of a quantile regression loss function. Dashed line is the Huber function approximation and the solid line is the quantile loss function.

where i indexes subject, y_i is the response, x_i is the covariate, and $\{u_i\}$ are iid F with density f and the density in a neighborhood of τ is greater than 0 (i.e. $f(F^{-1}(\tau)) > 0$). Under mild conditions, Koenker and Bassett (1978) showed

$$\sqrt{n}(\hat{\beta}(\tau) - \beta(\tau)) \xrightarrow{d} \mathcal{N}(0, w^2(\tau, F)D^{-1}),$$

where $\beta(\tau) = \beta + F^{-1}(\tau)e_1, e_1 = (1, 0, \cdots, 0)', w^2(\tau, F) = \tau(1-\tau)/f^2(F^{-1}(\tau))$, and $D = \lim_{n\to\infty} n^{-1} \sum_i x_i x_i'$.

When the error terms are non-iid; however, the asymptotic behavior of $\hat{\beta}(\tau)$ is more complicated and it takes on the sandwich form (Koenker and Machado 1999)

$$\sqrt{n}(\hat{\beta}(\tau) - \beta(\tau)) \xrightarrow{d} \mathcal{N}(0, \tau(1-\tau)H^{-1} J H^{-1}),$$

where $J(\tau) = \lim_{n\to\infty} n^{-1} \sum_i x_i x_i'$ and $H(\tau) = \lim_{n\to\infty} n^{-1} \sum_i x_i x_i' f_i(F_i^{-1}(\tau))$.

In either the iid or non-iid cases, the asymptotic variance of the quantile estimates depend upon the reciprocal of the density function evaluated at the quantile of interest termed the "sparsity function" (Tukey 1965) or "quantile-density function" (Parzen 1979). The dependence on the sparsity function should not be surprising since the precision would depend on how dense the observations are near the quantile of interest. In the iid case, the estimation of the sparsity function is well developed (Siddiqui 1960, Bofingeb 1975, Sheather and Maritz 1983, Welsh 1988, Hall and Sheather 1988). For the non-iid case, the estimation of H_n may be performed either using an extension of the methods used for the iid case (Hendricks and Koenker 1992) or by a method based on kernel density estimation (Powell 1991).

Inversion of a Rank Test Since direct estimation requires estimation of nuisance parameters (i.e. H), a method which provides a valid test without estimating H could be beneficial. The inversion of a rank test developed by Gutenbrunner et al. (1993) does avoid estimation of H_n.

Gutenbrunner and Jurecková (1992) observed that Hájek-Šidák rankscores (Hájek et al. 1967) may be viewed as a special case of a more general form for linear model. By inverting the test, confidence intervals can be efficiently computed for quantile estimators.

Consider the model $Q_\tau(Y|X_1, X_2) = X_1\beta_1 + X_2\beta_2$ and a test for the null hypothesis $H_0 : \beta_2 = \xi$ vs. $H_1 : \beta_2 \neq \xi$ where ξ is q-dimensional and τ is fixed. Under the null hypothesis and using the dual representation of Equation 2.6, the linear programming problem can be solved using

$$\hat{a}(\xi) = \max_a\{(Y - X_2\xi)'a \,|\, X_1'a = (1-\tau)X_1'1_n, a \in [0,1]^n\}.$$

The rank test statistic is defined as

$$T_n = S_n(\xi)'(\tau(1-\tau)\, q_n^2)^{-1} S_n(\xi) \to^d \chi_q^2,$$

where $q_n^2 = n^{-1}X_2'(I - X_1(X_1'X_1)^{-1}X_1')X_2$,

$$S_n = n^{-1/2}X_2'\,\hat{b}_n(\xi) \overset{d}{\sim} \mathcal{N}(0, \tau(1-\tau)q_n^2)$$

under the null with $\hat{b}_n(\xi) = \hat{a}(\xi) - (1-\tau)$. We can reject H_0 if $T_n(\xi) > \chi^2_{(q,1-\alpha)}$.

Confidence intervals for β_2 may be constructed by inverting the rank test. Different values of ξ can be tested and the collection of all ξ for which the null hypothesis is not rejected forms an approximate $(1-\alpha)$-th confidence interval for β_2. A more detailed development of this test can be found in Gutenbrunner et al. (1993).

The most important feature of this approach is that it is scale invariant and; therefore, avoids estimation of the sparsity function. The test can be carried out in conjunction with the simplex method (Koenker and d'Orey 1994).

Resampling Based Methods There are quite a few implementations of resampling methods for quantile regression inference. There are four basic types, namely, the x-y pair bootstrap, residual bootstrap, Markov chain marginal bootstrap (MCMB), and distinct resampling method devised by Parzen et al. (1994).

The x-y pair bootstrap resamples x-y pairs from the original dataset with replacement. The sampled data uses the original design points; therefore, it is able to accommodate some heteroscedasticity. A slight variation of the x-y pair bootstrap was introduced by Rao and Zhao (1992) and Chatterjee et al. (2005). Their methods also resample x-y pairs with replacement but then each of the bootstrapped observations is weighted by a randomly generated weight for estimation. Commonly used random variables for generating random weights are the exponential distribution with rate parameter equal to one and Poisson(1) since their mean and variance are both one.

The residual bootstrap resamples with replacement from the residual vector. The sampled residual vector is then added back to the fitted vector $X'\hat{\beta}(\tau)$ for estimation. It assumes that the error process is iid. A wild bootstrap procedure was proposed by Feng et al. (2011) for quantile regression. In their procedure, after the residuals are sampled, the absolute values of the residuals are multiplied by randomly generated weights before they are added back to the fitted vector for estimation.

The Markov chain marginal bootstrap (MCMB) was developed by He and Hu (2002) and it reduced the complexity of the bootstrap by resampling the marginal estimating equation at each bootstrap step. Due to this sampling scheme, only a one-dimensional equation is solved each time instead of the multi-dimensional equations. The resulting sequence of estimates, by construction, is a Markov chain. This method appears to perform well for data with heteroscedasticity in simulation studies.

The method developed by Parzen et al. (1994) is quite different from the bootstrap methods. It is based on a pivotal estimating function and is designed to handle data with heteroscedasticity. In practice, the procedure is carried out by augmenting the data with one additional observation for esitmation. The one additional observation is chosen such that the response variable, y_{n+1}, is an extremely large number so $I(y_{n+1} - x'_{n+1}\beta \leq 0)$ is always 0 and $x_{n+1} = n^{1/2}\,u/\tau$ where u is generated by a random vector U which is a weighted sum of independent and centered Bernoulli variables: $n^{-1/2}\sum_i x_i(\zeta_i - \tau)$ where $\{\zeta_i\}$ is a random observation from Bernoulli(τ).

2.2 Censored Quantile Regression

Quantile regression for data with "fixed " censoring times was first introduced in the econometrics discipline by Powell (1984; 1986). The term "fixed" here means that the censored values for the dependent variable are assumed to be known for all observations. One such data example is the pollutant concentration in the environment, where left censoring is typical due to detection limits of measuring instruments. It is also called the "Tobit " model after Tobin (1958) which can be written in the form

$$y_i = \max\{0, x_i'\beta + u_i\} \quad i = 1, \cdots, n,$$

where y_i is the dependent variable, x_i is the covariate for subject i, and $u_i \sim N(0, \sigma^2)$. The censored median regression estimator was defined in Powell (1984) to be the minimizer of

$$\frac{1}{n}\sum_i |y_i - \max\{0, x_i'\beta\}|.$$

Later on, Powell (1986) extended the "maximum score" estimator proposed by Manski (1975) to more general quantiles. It proposed the censored regression quantile estimator to be the minimizer of

$$\frac{1}{n}\sum_i \rho_\tau\left(y_i - \max\{0, x_i'\beta\}\right),$$

where $\rho_\tau(\cdot)$ is the quantile loss function as defined in (2.3). While this approach established an ingenious way to correct for "fixed" censoring, the objective function was not convex with respect to the parameters making it difficult to obtain a global minimizer. Several methods have been proposed to mitigate the related computational issues (Buchinsky and Hahn 1998, Chernozhukov and Hong 2002).

In much of survival analysis, however, censoring time is not always observed. To accommodate random censoring time, several methods were proposed in recent decades. The methods developed initially required stringent assumptions. A median regression model for random censoring was proposed by Ying et al. (1995) which required the censoring time to be independent of covariates. The model must be fit through minimization of discrete functions and may have multiple local minima. A simulated annealing algorithm is usually used to solve the minimization problem which can be computationally intensive. Honore et al. (2002) extended Powell's approach to incorporate random censoring, but it still required independence between the censoring time and covariates. Yang (1999) developed a median regression model which used weighted empirical survival and hazard functions when estimating. This method required more stringent demands for the error term, i.e.

the error terms have to be iid. For the remainder of this section, unless specified otherwise, T_i denotes the event time, C_i denotes the censoring time, and $Y_i = \min(T_i, C_i)$ where the index i refers to subject. Moreover, x_i denotes the covariate associated with subject i and $\delta_i = I(T_i \leq C_i)$ is the censoring indicator.

Under conditional independence, i.e. failure time and censoring time are independent conditional on covariates and without assuming stringent constraints on an error distribution, Portnoy (2003) innovatively proposed a recursively reweighted estimator. Specifically, Portnoy (2003) assumes all conditional quantiles are linear to estimate the quantile coefficients at a $\sqrt{n}$ rate. The algorithm generalizes the Kaplan-Meier estimator to the quantile regression setting and reduces to the Kaplan-Meier estimator when these is no censoring and only a single sample. Let $\{\tau_j^* : j = 1, 2, \cdots\}$ be the set of all breakpoints where the piecewise linear function changes the gradient. The algorithm (Portnoy 2003) proceeds as follows.

1. For the first breakpoint, τ_1^*, compute the first quantile using the (uncensored) quantile regression while pretending there is no censoring. For data that was censored, use the censored time as the response.

2. Define the weights, $w_i(\tau)$, for $\tau > \hat{\tau}_i$ as

$$w_i(\tau) = \frac{\tau - \hat{\tau}_i}{(1 - \hat{\tau}_i)},$$

   for censored observations. Suppose we already solved $\hat{\beta}(\tau)$ and weights $w_i(\tau)$ for all censored observations lying below the $\hat{\beta}(\tau_j^*)$ plane. For $\tau > \tau_j^*$ such that no additional censored point was crossed, $\beta(\tau)$ is the minimizer of

$$\sum_{i \notin K} \rho_\tau(Y_i - x_i'\beta) + \sum_{i \in K} \{w_i(\tau)\rho_\tau(C_i - x_i'\beta) + (1 - w_i(\tau)(Y_i^* - x_i'\beta))\},$$

where K denotes the indices for censored observations lying below $\hat{\beta}(\tau_j^*)$ and Y_i^* is any value large enough to exceed all $\{x_i'\beta : i \in K\}$.

3. Suppose there is at least one censored observation such that $C_i > x_i'\hat{\beta}(\tau_j^*)$ but $C_i < x_i'\hat{\beta}(\tau_{j+1}^*)$ then these observations need to be split,reweighted (as defined in step 2), then continue pivoting as in step 2.

4. The algorithm stops when either the next breakpoint is one or only censored observations remain.

The disadvantage of this method is that the quantile can not be computed until the entire lower quantile regression process is computed first (hence, assuming all lower quantiles are linear). The recursive scheme also complicated asymptotic inference.

To overcome inferential difficulties, Peng and Huang (2008) and Peng (2012) developed a new quantile regression method for survival data subject to conditionally independent censoring and used a martingale-based procedure which makes asymptotic inference more tractable. Unlike the method by (Portnoy 2003) which uses concepts from the Kaplan-Meier estimator, Peng and Huang (2008) linked their approach to the Nelson-Aalen estimator of the cumulative hazard function. Define $N(t) = I(Y \leq t, \delta = 1)$ and $N_i(t)$ is the sample analog of $N(t)$. Peng and Huang (2008) considered the estimating equation

$$n^{1/2} S_n(\beta, \tau) = 0, \tag{2.8}$$

where

$$S_n(\beta, \tau) = n^{-1} \sum_i x_i \left[N_i(e^{x_i'\beta(\tau)}) - \int_0^\tau I(y_i \geq e^{x_i'\beta(u)}) dH(u) \right]$$

and $H(x) = -\log(1 - x)$ for $0 \leq x < 1$.

$$\Lambda_T(t|X) = -\log\{1 - Pr(T \leq t|X)\}$$

is the cumulative hazard function of T_i conditional on X_i so $M_i(t) = N_i(t) - \Lambda_T(t \wedge T_i|x_i)$ is the martingale process associated with the counting process $N_i(t)$ and the conditional expectation of $M_i(t)$ equals zero ($E\{M_i(t)|x_i\} = 0$). Thus, we have

$$E\left\{n^{-1/2}\sum_i X_i\left[N_i(e^{X_i'\beta_0(\tau)}) - \Lambda_T[e^{X_i'\beta_0(\tau)} \wedge Y_i|x_i]\right]\right\} = 0,$$

where $\beta_0(\cdot)$ denotes the true value of $\beta(\cdot)$. Noting the fact that the quantile function is monotone in τ and $F_T[e^{x_i'\beta_0(u)}|x_i] = \tau$, we have

$$\Lambda_T[e^{x_i'\beta_0(\tau)} \wedge Y_i|x_i] = H(\tau) \wedge H\{F_T(Y_i|x_i)\} = \int_0^\tau I(Y_i \geq e^{x_i'\beta_0(u)})dH(u),$$

where $H(t)$ is defined as above. Combining these facts, one finds the estimating equation defined in Equation (2.8). Similar to Portnoy (2003), the method proposed by Peng and Huang (2008) still has the drawback that the entire lower quantile regression process needs to be computed before higher quantiles can be estimated. Both methods, Portnoy (2003) and Peng and Huang (2008), produced very similar estimates in small-sample simulation studies (Koenker 2008) and have been implemented in an `R` package (Koenker 2013).

Huang (2010) developed a new concept of quantile calculus while allowing for zero-density intervals and discontinuities in a distribution. The grid-free estimation procedure introduced by Huang (2010) circumvented grid dependency as in Portnoy (2003) and Peng and Huang (2008).

To avoid requiring that all lower quantiles are linear, Wang and Wang (2009) proposed a locally weighted method. Their approach assumes linearity at one prespecified quantile level of interest, thus relaxing the assumption of Portnoy (2003) and Peng and Huang (2008). The weighting scheme developed in Wang and Wang (2009) can be done in one single step making it much simpler than the recursive weighting scheme used in Portnoy

(2003). Specifically, estimators of Wang and Wang (2009) are the minimizer of the following weighted objective function,

$$n^{-1}\sum_{i=1}^{n}\left[w_i(F_0)\rho_\tau(Y_i - x_i'\beta) + \{1 - w_i(F_0)\}\rho_\tau(Y^* - x_i'\beta)\right],$$

where Y^* is a very large number such that $Y^* > x_i'\beta$,

$$w_i(F_0) = \begin{cases} 1 & \text{if } \delta_i = 1 \text{ or } F_0(C_i|x_i) > \tau \\ \frac{\tau - F_0(C_i|x_i)}{1 - F_0(c_i|x_i)} & \text{if } \delta_i = 0 \text{ and } F_0(C_i|x_i) < \tau \end{cases},$$

and $F_0(\cdot|x)$ is the cumulative distribution of survival time conditioned on the covariates. It is proposed that $F_0(\cdot|x)$ is estimated nonparametrically using the local Kaplan-Meier estimator. Since $F_0(\cdot|x)$ needs to be estimated nonparametrically using kernel estimations, their method suffers the curse of dimensionality and hence can only handle a small number of covariates.

Wey et al. (2014) developed a tree based approach to generate the weights used in Wang and Wang (2009). By avoiding the use of a kernel, their approach can generate weights for data with moderate to high dimensions including discrete covariates while assuming only local linearity.

Quantile regression models for failure time data with other censoring mechanisms have also been developed. For doubly censored failure time data where the outcome of interest is subject to both left censoring and right censoring, different methods were proposed by Lin et al. (2012) and Ji et al. (2012). A method developed by Lin et al. (2012) generalized the idea of Portnoy (2003) to the case of doubly censored data. Let L_i and R_i denote the left and right censoring time, respectively. $\tilde{Y}_i$ is defined as $\max(L_i, \min(T_i, R_i))$ and

$$\delta_i = \begin{cases} 1 & \text{if } L_i < T_i < R_i \text{ no censoring} \\ 2 & \text{if } T_i \geq R_i \text{ right censored} \\ 3 & \text{if } T_i \leq L_i \text{ left censored} \end{cases} .$$

The estimator in Lin et al. (2012) is the minimizer of

$$\sum_{\delta_i=1} \rho_\tau(T_i - x_i'\beta) + \sum_{\delta_i=2} \left\{ w_i^r(\tau)\rho_\tau(R_i - x_i'\beta) + (1 - w_i^r(\tau))\rho_\tau(\tilde{Y}_i - x_i'\beta) \right\}$$
$$+ \sum_{\delta_i=3} \left\{ w_i^l(\tau)\rho_\tau(L_i - x_i'\beta) + (1 - w_i^l(\tau))\rho_\tau(-\tilde{Y}_i - x_i'\beta) \right\},$$

where

$$w_i^r(\tau) = \frac{\tau - \tau_{R_i}}{1 - \tau_{R_i}}, \quad \text{if } \delta_i = 2 \text{ and } \tau > P(T_i < R_i | R_i, x_i) \text{ ; 1 otherwise}$$
$$w_i^l(\tau) = \frac{\tau_{L_i} - \tau}{\tau_{L_i}}, \quad \text{if } \delta_i = 3 \text{ and } \tau < P(T_i < L_i | L_i, x_i) \text{ ; 1 otherwise.}$$

Ji et al. (2012), on the other hand, generalized the method proposed by Peng and Huang (2008) to doubly censored data. It also considered an estimating equation as in Equation (2.8) but redefine $S_n(\beta, \tau)$ as

$$S_n(\beta, \tau) = n^{-1} \sum_i x_i \left\{ N_i[g(x_i'\beta(\tau))] - \int_0^\tau I[L_i < g(x_i'\beta(u)) \leq Y_i] \, dH(u) \right\},$$

where $g(\cdot)$ is a known monotone link function,$H(x) = -\log(1 - x)$ for $0 \leq x < 1$, and $N_i(t)$ is the counting process defined as $N_i(t) = I(Y_i \leq t, \delta_i = 1)$.

Zhou (2011) proposed a weighted method for randomly left truncated data where the weights are related to the weighting scheme proposed by He and Yang (2003). Noticeably, the estimation procedure reduces to the classical quantile regression when no truncation presents; however, it is not the case for most recursive methods. Shen (2014) extended the weighted method proposed by Zhou (2011) to left-truncated and right-censored data. Most recently, Sun et al. (2015) generalized the quantile regression models to accommodate recurrent events data.

The literature for quantile regression models on interval-censored data is lacking. To the best of our knowledge, the only median regression method available for interval-censored data is proposed by Kim et al. (2010), which extended the median regression developed by McKeague et al. (2001). Let T_{L_i} and T_{R_i} be the last visit time prior to event occurrence and first visit time after event occurred for subject i, respectively. Also let $\tilde{\delta}_i = I(T_{R_i} < \infty)$, which is an indicator function for the interval-censored observations. Define $0 = s_0 < s_1 < \cdots < s_q < s_{q+1} = \infty$ as the unique order points of T_{L_i} and T_{R_i} and $\alpha_{i,j} = I(T_{L_i} \leq s_j \leq T_{R_i})$ for the interval-censored cases where $j = 1, \cdots, q$. Kim et al. (2010) proposed that estimators be the root of

$$\frac{1}{n}\sum_{i=1}^{n} x_i \left\{ \tilde{\delta}_i \sum_{j=1}^{q} w_{ij} \left[I(s_j \geq x_i'\beta) - \frac{1}{2} \right] + (1-\tilde{\delta}_i) \left[I(T_{L_i} \geq x_i'\beta) + I(T_{L_i} < x_i'\beta)u_i - \frac{1}{2} \right] \right\} = 0,$$

where $w_{ij} = \alpha_{ij}\, f_{j|x}/(S_x(T_{L_i}) - S_x(T_{R_i}))$ for the interval-censored observations and $u_i = S_x(x_i'\beta)/S_x(T_{L_i})$ for the right-censored observations, and $S_x(t) = Pr(T > t|X = x)$. Under the discrete failure time assumption, $f_{l|x}$, can be estimated using a self-consistency algorithm (Turnbull 1976) and then be used to calculate $S_x(s_k) = 1 - \sum_{l \leq k} f_{l|x}$. The estimating procedure consists of estimating an initial value then iterating between estimation of $\hat{f}_x$, $\hat{S}_x$ for weight $(w_{i1}, \cdots, w_{iq})$ and u_i calculation; and the estimation of the parameter. There are a couple of drawbacks to this method. First, the asymptotic properties of the estimates were not established for the estimates; and second, it can only be applied when the covariates have a finite number of values.

2.3 Accelerated Failure Time Models for Interval-Censored Failure Time Data

The accelerated failure time (AFT) model relates the covariates linearly to the logarithm of the survival time,

$$\log(T) = X^T\beta + \epsilon,$$

where X is the p-dimensional covariate vector and ϵ is the error term which is independent of X. When the error distribution is left unspecified, the AFT model can be thought of as a semiparametric alternative to the Cox model or relative risk model. Since the non-parametric maximum likelihood estimator is not directly applicable to AFT models with interval-censored data, an inference procedure is more difficult.

Rabinowitz et al. (1995) proposed a class of score statistics that may be used for estimation and confidence procedures. Consider data from n subjects, indexed by i. Let T_i denote the log survival time and Z_i be a p-dimensional covariate for subject i. Rabinowitz et al. (1995) considered the linear regression model,

$$T_i = Z_i^T\beta + \epsilon_i,$$

where ϵ_i are independent and identically distributed residuals with distribution function F and density f. Let $X_{i,L}$ and $X_{i,U}$ be the last examination times preceding T_i and the first examination after T_i, respectively. Consider a function g with domain $[0,1]$, satisfying $g(0) = g(1) = 0$, let

$$\zeta_i(b) = \frac{g[F\{X_{i,U}(b)\}] - g[F\{X_{i,L}(b)\}]}{F\{X_{i,U}(b)\} - F\{X_{i,L}(b)\}} Z_i,$$

and let

$$\check{S}(b) = \sum_i^n \zeta_i(b).$$

Under the condition $g(0) = g(1) = 0$, $E\{\check{S}(b)\} = 0$. Unfortunately, $\check{S}(b)$ is not available for inference since the nuisance parameter F is unknown. Rabinowitz et al. (1995) proposed to estimate F then substitute the estimated F into $\check{S}(b)$. The estimate of β can be defined as a zero of $\check{S}(b)$ with the nuisance parameter F replaced by the estimated F.

Under the current status data setting, Murphy et al. (1999) and Shen (2000) developed likelihood-based methods. Consider the AFT model,

$$\log(T) = X^T\beta + \epsilon,$$

where T is the survival time and the error term ϵ has a density function F. Let C denote the observation time and $\delta \equiv I\{T \leq C\}$. Murphy et al. (1999) considered a penalized nonparametric maximum likelihood estimator in the AFT model under the current status setting. Consider the log likelihood

$$L_n(\beta, F) = \frac{1}{n}\sum_{i=1}^{n}\{\delta_i \log F(c_i - \beta x_i) + (1-\delta_i)\log[1 - F(c_i - \beta x_i)]\}.$$

The penalized likelihood is defined as

$$L_n(\beta, F) - \hat{\lambda}_n^2 J^2(F),$$

where the penalty $J(F)$ is defined as

$$J^2(F) = \int_D F''(u)^2 du,$$

the domain D is taken to be a finite interval which contains the support of $C - \beta X$ for every β, and the smoothing parameter $\hat{\lambda}_n^2$ determines the severity of the penalty. For asymptotics, the smoothing parameter, $\hat{\lambda}_n^2$, satisfy

$$\hat{\lambda}_n^2 = o_p\left(\frac{1}{n^{1/2}}\right), \quad \frac{1}{\hat{\lambda}_n} = O_p(n^{2/5}),$$

and may be data-dependent. The asymptotic properties of penalized maximum likelihood was established and a $\sqrt{n}$ convergence rate of the parameter estimates is possible under certain conditions.

Shen (2000) constructed a likelihood based on the random-sieve likelihood. Let $F(\epsilon_i(\theta))$ be the cdf with jump sizes $\{p_i\}_{i=1}^n$ at $\{\epsilon_i\}_{i=1}^n$ and let $G(\{p_i\}, \theta)$ be $(\sum_{i=1}^n x_i\epsilon_i(\theta)p_i, \sum_{i=1}^n \epsilon_i^2(\theta)p_i - \sigma^2)$ where σ^2 is the finite variance of ϵ. The profile random-sieve log-likelihood is defined as

$$\sum_{i=1}^n (\log[F(\epsilon_i(\theta))]^{\delta_i}) + \log[1 - F(\epsilon_i(\theta))]^{1-\delta_i}),$$

and the constraints are defined as

$$G(\{p_i\}, \theta) = 0.$$

The maximum random-sieve likelihood estimate can be obtained by maximizing the profile random-sieve log-likelihood over a random sieve

$$\mathcal{F}_n = \{(p_i) : G(\{p_i\}, \theta) = 0, \sum_{i=1}^n p_i \leq 1, 0 \leq p_i \leq 1\}.$$

The profile random-sieve log-likelihood for θ is then given by

$$\sup_{\{F \in \mathcal{F}_n\}} \sum_{i=1}^n [\delta_i \log F(\epsilon_i(\theta)) + (1 - \delta_i) \log(1 - F(\epsilon_i(\theta)))],$$

and estimate of θ maximizes the profile random-sieve log-likelihood.

Under a general interval-censored data setting, Betensky et al. (2001) studied a simple numerically efficient estimation procedure. The examination time and event time were assumed to be independent in Betensky et al. (2001). The examination times from the same individual were used as independent observations for estimation. When calculating the standard error of the estimates, the dependence between different measurements obtained from the same individual was then accounted for. Specifically, consider data from n subjects, indexed by i. Let T_i denote the survival time and Z_i be a p-dimensional covariate for subject i. Betensky et al. (2001) considered the linear regression model

$$\log(T_i) = Z_i^T \beta + \epsilon_i,$$

where ϵ_i are independent and identically distributed residuals independent of Z_i and X_i where X_i denote the subject's collection of examination times. Let $X_{i,L}$ and $X_{i,U}$ be the last examination times preceding T_i and the first examination after T_i, respectively. Let Y_{ij} be the indicator that the ith subject's event time precedes the jth examination time of subject i: $Y_{ij} = 1\{T_i \leq X_{ij}\}$, with X_{ij} being ancillary for F and β. Treating examination times from the same subject as independent observations, the conditional likelihood given the ancillaries was defined as

$$\prod_{i=1}^{n} \prod_{j=1}^{n_i} F\{X_{ij}(\beta)\}^{Y_{ij}} [1 - F\{X_{ij}(\beta)\}]^{1-Y_{ij}},$$

where $X_{ij}(\beta) \equiv \log(X_{ij}) - Z_i^T b$ for a p-dimensional vector b. Let $\hat{F}_b$ denote the nonparametric maximum likelihood estimator of F with $b = \beta$ then $\hat{F}_b$ can be calculated using the pool adjacent violators algorithm with Y_{ij} and $X_{ij}(b)$. Similar to Rabinowitz et al. (1995), Betensky et al. (2001) considered the scores,

$$S(b) = \sum_{i=1}^{n} \sum_{j=1}^{n_i} [Y_{ij} - \hat{F}_b\{X_{ij}(b)\}] Z_i.$$

To get the estimates of β, $S(b)$ can be computed for a fine grid of values of b and the estimators can be set to the b which made $S(b)$ closest to zero. The confidence interval for β can also be calculated while taking into account the correlation between the examination times within the same subject.

To overcome the numerical difficulty presented in previous methods and to include higher-dimensional covariates, Tian and Cai (2006) proposed to construct the estimator by inverting a Wald-type test for testing a null proportional hazards model. Consider an accelerated failure time model,

$$\log(T) = \beta^T Z + \epsilon.$$

Let C denote the observation time and $\delta \equiv I\{T \leq C\}$. Assume that the distribution of the residual $\epsilon = \log(T) - \beta^T Z$ is independent of the covariate Z which is equivalent to assuming that

$$\lambda_\epsilon(t|Z) = \lambda_0(t),$$

where $\lambda_\epsilon(\cdot|Z)$ is the hazard function of ϵ conditional on the covariate Z and $\lambda_0(\cdot)$ is some unknown baseline hazard function. One way to test this assumption is to fit the model

$$S_\epsilon(t|Z) = S_0(t)^{exp(\gamma_0^T Z)},$$

using residual data, $\{(\log(C_i) - \beta^T Z_i, \delta_i, Z_i) : i = 1, \cdots, n\}$ and then test the hypothesis $H_0 : \gamma_0 = 0$ based on an estimator of γ_0. Since the distribution of $\epsilon(\beta) = \log(T) - \beta^T Z$ is independent of Z if and only if β is at the true value, we can estimate β by solving the estimating equations,

$$\hat{\gamma}_n(\beta) = o_p(n^{-1/2}).$$

The estimation procedure can be carried out by 1) computing the nonparametric maximum likelihood estimators of β in a set of working proportional hazard models indexed by β then 2) finding the estimates which satisfy $\hat{\gamma}_n(\beta) = o_p(n^{-1/2})$. The first step can be solved using algorithms developed for nonparametric maximum likelihood estimators. When the covariate is one-dimensional, a grid search can be used to find the β. When the covariates are high-dimensional, a Markov chain Monte Carlo based procedure was developed to obtain the point estimator and a consistent estimator of its variance-covariance matrix simultaneously. Tian and Cai (2006) also extended the method to general interval-censored data.

CHAPTER 3: QUANTILE REGRESSION MODELS FOR CURRENT STATUS DATA

3.1 Introduction

Quantile regression (Koenker and Bassett 1978) is a robust estimation method for regression models which offers a powerful and natural approach to examine how covariates influence the location, scale, and shape of a response distribution. Unlike linear regression analysis, which focuses on the relationship between the conditional mean of the response variable and explanatory variables, quantile regression specifies changes in the conditional quantile as a parametric function of the explanatory variables. It has been applied in a wide range of fields including ecology, biology, economics, finance, and public health (Cade and Noon 2003, Koenker and Hallock 2001). Quantile regression for censored data was first introduced by Powell (Powell 1984; 1986), where the censored values for the dependent variable were assumed to be known for all observations (also known as the "Tobit" model). While this approach established an ingenious way to correct for censoring, the objective function was not convex over parameter values making global minimization difficult. Several methods have been proposed to mitigate related computational issues (Buchinsky and Hahn 1998, Chernozhukov and Hong 2002).

In most survival analysis, however, censoring time is not always observed. To accommodate a random censoring time, several methods were proposed over the past few decades. Early methods (Ying et al. 1995, Yang 1999, Honore et al. 2002) required stringent assumptions on the censoring time, i.e. the censoring time must be independent of covariates. Under conditional independence assumption where failure time and censoring time are independent conditional on covariates, Portnoy (2003) proposed a recursively reweighted estimator. Unfortunately, the quantile cannot be computed until the entire lower quantile regression process was computed first. The recursive scheme also complicated asymptotic inference. To overcome inferential difficulties, Peng and Huang (2008) and Peng (2012) developed a quantile regression method for survival data subject to conditionally independent censoring and used a martingale-based procedure which made asymptotic inference more tractable. However, the method developed by Peng and Huang (2008) still has the same drawback as in Portnoy (2003), namely, the entire lower quantile regression process must be computed first. Huang (2010) developed a new concept of quantile calculus while allowing for zero-density intervals and discontinuities in a distribution. The grid-free estimation procedure introduced by Huang (2010) circumvented grid dependency as in Portnoy (2003) and Peng and Huang (2008). To avoid the necessity of assuming that all lower quantiles were linear, Wang and Wang (2009) proposed a locally weighted method. Their approach assumed linearity at one prespecified quantile level of interest and thus relaxed the assumption of Portnoy (2003); however, their method suffered the curse of dimensionality and hence can only handle a small number of covariates.

Current status data arise extensively in epidemiological studies and clinical trials, especially in large-scale longitudinal studies where the event of interest, such as disease contraction, is not observed exactly but is only known to happen before or after an examination time. Many likelihood-based methods have been developed for current status data, such as proportional hazard models, proportional odds models, and additive hazard models (see Sun (2007) for a survey of different methods). Despite the fact that the development for censored quantile regression flourishes, the aforementioned methods were developed for right-censoring and are not suitable for current status data. To the best of our knowledge, the only method available for quantile regression models on interval-censored data was proposed by Kim et al. (2010) which was a generalization of the method proposed by McKeague et al. (2001). The proposed method can only be applied when the covariates took on a finite number of values since the method required estimation of the survival function conditional on covariates. The proposed method performed well in simulation studies , yet no theoretical justifications were offered. In this paper, we develop a new method for the conditional quantile regression model for current status data while allowing the censoring time to depend on the covariates.

The remaining paper is organized as follows. In Section 3.2, the proposed model is introduced and we establish estimation and inference procedures. Consistency and asymptotic distribution are presented in Section 3.3 with technical details deferred to Section 3.6. In Section 3.4, the small-sample performance is demonstrated via simulation studies and the application to data from the Voluntary HIV-1 Counseling and Testing Efficacy Study Group is given. Section 3.5 discusses the method presented herein.

3.2 The Method

3.2.1 Model and Data

Let T denote failure time and let X denote a $k \times 1$ covariate vector with the first component set as one. We consider a quantile regression model for the failure time,

$$Q_T(\tau \mid X) = X^T \beta(\tau), \qquad \tau \in (0,1), \tag{3.9}$$

where $Q_T(\tau \mid X)$ is the conditional quantile defined as

$$Q_T(\tau \mid X) = \inf\{t : \mathrm{pr}(T \leq t \mid X) \geq \tau\}$$

and the vector of unknown regression coefficients, $\beta(\tau)$, represents the covariate effects on the τth quantile of T which may depend on τ. Each element of $\beta(\tau)$ can be interpreted as an estimated difference in τth quantile by one unit change of the corresponding covariate while other variables in the model are held constant. Our interest lies in the estimation and inference on $\beta(\tau)$.

Let C denote observation time and define $\delta \equiv I(T \leq C)$ where $I(\cdot)$ is the indicator function. For the current status data, T is not observed and the observed data consist of n independent replicates of (C, X, δ), denoted by $\{(C_i, X_i, \delta_i)_{i=1,\cdots,n}\}$. It is assumed that T is conditionally independent of C given X. Since T is unobserved, we cannot directly estimate the conditional quantile function $Q_T(\tau \mid X)$ in Equation (3.9) making a standard quantile regression unsuitable for our problem.

The τth conditional quantile of a random variable Y conditional on X can be characterized as the solution to the expected loss minimization problem,

$$Z(\beta) = E\{E[\rho_\tau(Y - X^T\beta(\tau)) \mid X]\}, \tag{3.10}$$

where $\rho_\tau(u) = u[\tau - I(u < 0)]$. Furthermore, quantiles possess "equivariance to monotone transformations" (Koenker 2005) which means that we may analyze a transformation $h(Y)$ since the conditional quantile of $h(Y)$ is $h(X^T\beta(\tau))$ if $h(\cdot)$ is nondecreasing (Powell 1994). In current status data, we observe realizations of the transformed variable $\delta \equiv I(T \leq C)$ or, equivalently, $(1 - \delta) \equiv I(T > C)$ where the transformation is $h(T \mid C) = I(T > C)$ which is monotone nondecreasing. We apply the same transformation to the conditional quantile, $X^T\beta(\tau)$, and use the transformed conditional quantile, $I(X^T\beta(\tau) > C)$, in the subsequent analysis. The objective function in (3.10) is well-defined and is sufficient to identify the parameters of interests (Powell 1994). We can thus substitute $(1 - \delta)$ and $I(X^T\beta > C)$ in Equation (3.10) and get

$$Z(\beta) = E[E\{\rho_\tau[(1 - \delta) - I(X^T\beta(\tau) > C)] \mid X, C\}]. \tag{3.11}$$

Equation defined in (3.11) is used to identify $\beta(\tau)$ since it contains only the observable variables (C, X, δ). We can show that the derivative of $Z(\beta)$ with respect to β is zero at the true β (see Section 3.6 for details). Due to censoring, it is possible that not all $\beta(\tau)$ can be estimated using the observed data. We provide a sufficient condition to guarantee the identifiability for a fixed quantile in Section 3.3.1.

3.2.2 Parameter Estimation and Algorithm

To simplify notation, we use β instead of $\beta(\tau)$ henceforth. Assuming the formulation from Equation (3.11), the regression quantile estimator $\hat{\beta}_n$ (Koenker and Bassett 1978) is the minimizer of the objective function

$$\begin{aligned} Z_n(\beta) &= \sum_{i=1}^{n} \rho_\tau[(1-\delta_i) - I(X_i^T\beta > C_i)] \\ &= \sum_{i=1}^{n} \left[\tau I(\delta_i = 0) I(X_i^T\beta - C_i \leq 0) + (1-\tau) I(\delta_i = 1) I(X_i^T\beta - C_i > 0)\right] \\ &= \sum_{i=1}^{n} w_i\, I[y_i(X_i^T\beta - C_i) \leq 0] \end{aligned} \tag{3.12}$$

where

$$y_i = \begin{cases} 1 & \text{if } \delta_i = 0 \\ -1 & \text{if } \delta_i = 1 \end{cases}, \quad w_i = \begin{cases} \tau & \text{if } \delta_i = 0 \\ 1-\tau & \text{if } \delta_i = 1 \end{cases}.$$

The regression quantile estimator, $\hat{\beta}_n$, which minimizes Equation(3.12) is difficult to obtain by direct minimization since $Z_n(\beta)$ is neither convex nor continuous. To overcome this difficulty, we approximate $Z_n(\beta)$ as a difference of two hinge functions, where the approximation is controlled via a small constant, ϵ.

$$
\begin{aligned}
Z_{n,\epsilon}(\beta) &= \sum_{i=1}^{n} w_i \left\{ \frac{1}{\epsilon}\left[\frac{\epsilon}{2} - y_i(X_i^T\beta - C_i)\right]_+ - \frac{1}{\epsilon}\left[-\frac{\epsilon}{2} - y_i(X_i^T\beta - C_i)\right]_+ \right\} \\
&= \sum_{i=1}^{n} w_i \left[\frac{1}{2} - \frac{1}{\epsilon} y_i(X_i^T\beta - C_i)\right]_+ + \sum_{i=1}^{n}(-w_i)\left[-\frac{1}{2} - \frac{1}{\epsilon} y_i(X_i^T\beta - C_i)\right]_+ \qquad (3.13)
\end{aligned}
$$

where $\epsilon > 0$ and $[\cdot]_+$ denotes the positive part of the argument. We illustrate the approximation of a $0/1$ loss by the difference between two hinge functions in Figure 3.8.

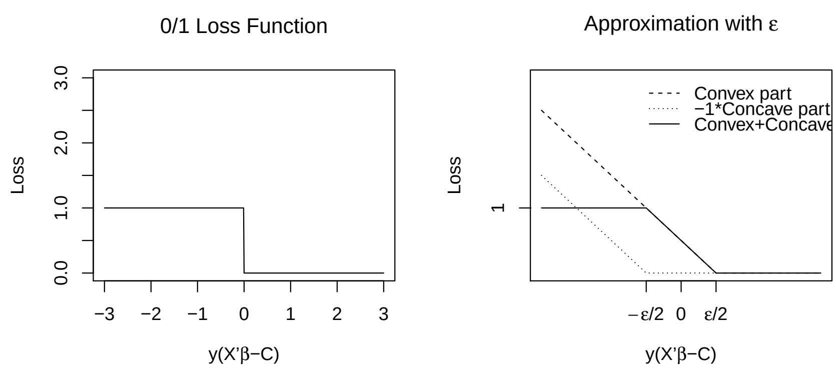

Figure 3.8: An illustration of using the difference between two hinge loss functions to approximate a $0/1$ loss. A smaller ϵ provides a closer approximation.

To mitigate computational difficulties, we utilize the concave-convex procedure proposed by Yuille and Rangarajan (2003). The concave-convex procedure relies on decomposing an objective function, $f(x)$, into a convex part, $f_{convex}(x)$, and a concave part, $f_{concave}(x)$ such that

$$f(x) = f_{convex}(x) + f_{concave}(x).$$

Optimization is carried out with an iterative procedure in which $f_{concave}(x)$ is linearized at the current solution $x^{(t)}$,

$$x^{(t+1)} = \arg\min_x \left[f_{convex}(x) + (x - x^{(t)}) f'_{concave}(x^{(t)}) \right],$$

making each iteration a convex optimization problem. The first value $x^{(0)}$ can be initialized with any reasonable guess.

To apply the concave-convex procedure to our optimization problem, we define the first term in Equation (3.13) as $f_{convex}(\beta)$ and the second term as $f_{concave}(\beta)$. The gradient of the concave part, $f_{concave}(\beta)$, is

$$\frac{\partial}{\partial \beta} f_{concave}(\beta) = \begin{cases} \sum_{i=1}^{n} w_i \left(\frac{1}{\epsilon} y_i X_i'\right) & \text{if } \frac{1}{2} + \frac{1}{\epsilon} y_i (X_i^T \beta - C_i) < 0 \\ 0 & \text{if } \frac{1}{2} + \frac{1}{\epsilon} y_i (X_i^T \beta - C_i) > 0 \end{cases}$$

Applying the concave-convex procedure to the above decomposition, we obtain

$$\begin{aligned} \beta^{(r+1)} = \arg\min_{\beta} \Bigg\{ & \sum_{i=1}^{n} w_i \left[\frac{1}{2} - \frac{1}{\epsilon} y_i (X_i^T \beta - C_i) \right]_+ \\ & + \sum_{i=1}^{n} w_i \frac{1}{\epsilon} y_i X_i' (\beta - \beta^{(r)}) \cdot I \left[\frac{1}{2} + \frac{1}{\epsilon} y_i (X_i^T \beta^{(r)} - C_i) < 0 \right] \Bigg\}, \end{aligned} \tag{3.14}$$

where $\beta^{(r)}$ denotes the estimated β at the rth iteration. The final form can be solved with a standard convex optimization algorithm with a decreasing sequence of $\epsilon = \{2^0, 2^{-1}, \cdots\}$. Specifically, the initial values for both the simulation studies and the real data example were generated using a coarse grid search. Given the initial value, we solve Equation (3.14) with $\epsilon = 2^0$. The solution with $\epsilon = 2^0$ is then used as the initial value to solve Equation (3.14) with $\epsilon = 2^{-1}$. This is repeated until the maximum relative change over all covariates is less than one percent. In this study, the `fminsearch` function from the optimization toolbox in MATLAB was used to solve for β. The `fminsearch` function performs unconstrained nonlinear optimization to find the minimum of a scalar function of several variables.

3.2.3 Inference

The confidence intervals for parameter estimates are obtained using a subsampling method since the bootstrap does not consistently estimate the asymptotic distribution for estimators with cube-root convergence (Abrevaya and Huang 2005). The subsampling method described below is from Politis et al. (1999). Subsampling can produce consistent estimated sampling distributions under extremely week assumptions even when the bootstrap fails and it can be used to obtain confidence intervals for parameter estimates. It should not be used to obtain standard errors; however, since our estimators are not normally distributed, even asymptotically (Section 3.3.3); therefore, there is no simple relation between the distribution of the estimators and standard errors (Horowitz 2010, page 108).

The justification for using the subsampling method in our study is discussed further in Section 3.3.3.

To obtain the confidence intervals of minimizer of Equation (3.13), $\hat{\beta}_{n,\epsilon}$, we produce subsamples $K_1, K_2, \cdots, K_{N_n}$ where K_j's are the $N_n \equiv \binom{n}{b}$ distinct subsets of $\{(C_i, X_i, \delta_i)_{i=1,\cdots,n}\}$ of size b. Let β_τ denote the true parameter values and $\hat{\beta}_{n,\epsilon,b,j}$ denote the estimated value produced by solving Equation (3.14) using the K_jth dataset.

Define

$$L_{n,b}(x) = N_n^{-1} \sum_{j=1}^{N_n} I[b^{1/3}(\hat{\beta}_{n,\epsilon,b,j} - \hat{\beta}_{n,\epsilon}) \leq x] \quad \text{and} \quad c_{n,b}(\gamma) = \inf\{x : L_{n,b}(x) \geq \gamma\}.$$

From Theorem 2.2.1 of Politis et al. (1999), for any $0 < \gamma < 1$, $P\left[n^{1/3}(\hat{\beta}_{n,\epsilon} - \beta_\tau) \leq c_{n,b}(\gamma)\right] \to \gamma$ under the condition that $b \to \infty$ as $n \to \infty$ and $b/n \to 0$. It follows that for any $0 < \alpha < 0.5$,

$$P\left[c_{n,b}\left(\frac{\alpha}{2}\right) < n^{1/3}(\hat{\beta}_{n,\epsilon} - \beta_\tau) \leq c_{n,b}\left(1 - \frac{\alpha}{2}\right)\right] \to 1 - \alpha$$

thus an asymptotic $1-\alpha$ level confidence interval for β_τ can be constructed with

$$\left[\hat{\beta}_{n,\epsilon} - n^{-1/3} c_{n,b}\left(1-\frac{\alpha}{2}\right),\, \hat{\beta}_{n,\epsilon} - n^{-1/3} c_{n,b}\left(\frac{\alpha}{2}\right)\right].$$

Symmetric confidence intervals can be obtained by modifying the above approach slightly. Define

$$\tilde{L}_{n,b}(x) = N_n^{-1}\sum_{j=1}^{N_n} I[b^{1/3}\,|\hat{\beta}_{n,\epsilon,b,j} - \hat{\beta}_{n,\epsilon}| \le x] \quad \text{and} \quad \tilde{c}_{n,b}(\gamma) = \inf\{x : \tilde{L}_{n,b}(x) \ge \gamma\}.$$

Again , if $b \to \infty$ as $n \to \infty$ and at the same time $b/n \to 0$, a symmetric confidence interval for β_τ can be constructed as

$$\left[\hat{\beta}_{n,\epsilon} - n^{-1/3}\tilde{c}_{n,b}(1-\alpha),\, \hat{\beta}_{n,\epsilon} + n^{-1/3}\tilde{c}_{n,b}(1-\alpha)\right]. \tag{3.15}$$

Symmetric confidence intervals are desirable because they often have nicer properties than the nonsymmetric version in finite samples (Banerjee and Wellner 2005). This fact was also observed in our simulation studies; hence, symmetric confidence intervals are recommended and used in this paper.

To avoid large scale computation issues, a stochastic approximation from Politis et al. (1999) is employed where B randomly chosen datasets from $\{1, 2, \cdots, N_n\}$ are used in the above calculation. Furthermore, the block size is chosen using the method implemented in Delgado et al. (2001) and Banerjee and Wellner (2005). Briefly, the algorithm for choosing block size is described below.

Step 1: Fix a selection of reasonable block sizes b between limits b_{low} and b_{up}.

Step 2: Draw M bootstrap samples from the actual dataset.

Step 3: For each bootstrap sample, construct a subsampling symmetric confidence interval with asymptotic coverage $1-\alpha$ for each block size b. Let $R_{m,b}$ be one if $\hat{\beta}_{n,\epsilon}$ was within the mth interval based on block size b and zero otherwise.

Step 4: Compute $\hat{h}(b) = M^{-1} \sum_{m=1}^{M} R_{m,b}$.

Step 5: Find the value $\tilde{b}$ that minimizes $|\hat{h}(b) - \alpha|$ and use $\tilde{b}$ as the block size when constructing confidence interval for the original data.

3.3 Asymptotic Properties

3.3.1 Identifiability

Prior to deriving the asymptotic properties of the proposed estimator, we will discuss a set of sufficient conditions for identifiability.

For a fixed quantile τ, let

$$Z_n(\beta) = \frac{1}{n} \sum_{i=1}^{n} \left[\tau I(\delta_i = 0) I(X_i^T \beta - C_i \leq 0) + (1 - \tau) I(\delta_i = 1) I(X_i^T \beta - C_i > 0) \right],$$

and

$$Z(\beta) = E[\tau I(\delta = 0) I(X'\beta - C \leq 0) + (1 - \tau) I(\delta = 1) I(X'\beta - C > 0)].$$

Let β_τ denote a minimizer of $Z(\cdot)$. The following conditions will be used in subsequent theorems.

Condition 1. *The support of f_X is not contained in any proper linear subspace of $\Re^k$.*

Condition 2. *For a fixed τ, with probability one, both the support of the conditional density of C given X, $f_{C|X}(\cdot)$, and the support of the conditional density of T given X, $f_{T|X}(\cdot)$, contain $X^T \beta_\tau$ in their interiors.*

Condition 1 is the typical full-rank condition.

Lemma 3.3.1. *Under Conditions 1 and 2, β_τ is identifiable, i.e. β_τ is the unique minimizer of $Z(\cdot)$.*

We prove Lemma 3.3.1 by showing $Z(\beta) - Z(\beta_\tau) > 0, \ \forall \beta \neq \beta_\tau$. A detailed proof is provided in Section 3.6. In our real data application, we suggest an empirical way to identify quantiles which are estimable.

3.3.2 Consistency

For a fixed quantile τ, let

$$Z_{n,\epsilon}(\beta) = \sum_{i=1}^{n} \Bigg(\tau I(\delta_i = 0) \left\{ I(X_i^T\beta - C_i \leq -\frac{\epsilon}{2}) + I(|X_i^T\beta - C_i| < \frac{\epsilon}{2}) \left[-\frac{1}{\epsilon}(X_i^T\beta - C_i - \frac{\epsilon}{2}) \right] \right\}$$
$$+ (1-\tau) I(\delta_i = 1) \left\{ I(X'\beta - C_i > \frac{\epsilon}{2}) + I(|X_i^T\beta - C_i| \leq \frac{\epsilon}{2}) \left[\frac{1}{\epsilon}(X_i^T\beta - C_i + \frac{\epsilon}{2}) \right] \right\} \Bigg)$$

We assume the following conditions for the consistency theorem.

Condition 3. *Let $\beta \in \mathcal{B}$ where $\mathcal{B}$ is a compact subset of $\Re^k$ which contains β_τ as an interior point.*

Condition 4. $M_T \equiv \sup_{T,X} f_{T|X}(T \mid X) < \infty$ *and* $M_C \equiv \sup_{C,X} f_{C|X}(C \mid X) < \infty$ *where $f_{C|X}$ and $f_{T|X}$ are the conditional density of C given X and T given X, respectively.*

Let $\hat{\beta}_{n,\epsilon}$ be the minimizer of $Z_{n,\epsilon}(\cdot)$ in $\mathcal{B}$.

Theorem 3.3.2. *Under Conditions 1–4, $\hat{\beta}_{n,\epsilon}$ converges to β_τ in probability as $n \to \infty$ and $\epsilon \to 0$.*

The proof follows by first showing that the collection of functions in $Z_n(\beta)$ is a VC-subgraph class and hence $Z_n(\beta)$ converge almost surely uniformly to $Z(\beta)$. In addition, $Z_{n,\epsilon}(\beta)$ converges almost surely uniformly to $Z_n(\beta)$ as $n \to \infty$ and $\epsilon \to 0$; thus we can conclude that $Z_{n,\epsilon}(\beta)$ converges almost surely uniformly to $Z(\beta)$. Next, we prove that $Z(\cdot)$ is continuous. Conditions 1 and 2 provide sufficient conditions for identifiability and hence, β_τ is the unique minimizer of $Z(\cdot)$. Since we assumed $\mathcal{B}$ is compact, we can then conclude that $\hat{\beta}_{n,\epsilon}$ converges to β_τ in probability by a standard argument for M-estimators (Theorem 2.1 of Newey and McFadden (1994)). A detailed proof is provided in Section 3.6.

3.3.3 Asymptotic Distribution

This section shows that $n^{1/3}(\hat{\beta}_{n,\epsilon} - \beta_\tau)$ converges to a nondegenerate distribution. The convergence rate is atypical because our objective function (3.12) is non-smooth and not everywhere differentiable; this is sometimes called the "sharp-edge effect" (Kim and Pollard 1990). We will make the following assumptions which guarantee the asymptotic distribution will be nondegenerate, namely

Condition 5. *The ϵ which is used in Equation (3.13) is $o(n^{-2/3})$.*

Condition 6. *The true distribution P of C, T and X is absolutely continuous with respect to Lebesgue measure.*

Condition 7. *X is bounded.*

Condition 8. *Let $V(\beta_\tau)_{i,j} = P_x\left[X_i X_j f_{C|X}(X'\beta_\tau \mid X) f_{T|X}(X'\beta_\tau \mid X)\right]$ and $V(\beta_\tau)$ is positive definite where X_i and X_j are elements of X.*

We may now proceed with the main result.

Theorem 3.3.3. *Under Conditions 1–8, the process $\left\{n^{2/3}\left[Z_n\left(\beta_\tau + sn^{-1/3}\right) - Z_n\left(\beta_\tau\right)\right] : s \in \Re^k\right\}$ converges in distribution to a Gaussian process $\left\{\Gamma(s) : s \in \Re^k\right\}$ with continuous sample paths, mean $s'V(\beta_\tau)s/2$, and covariance H, where V is the second order expansion of $Z(\beta)$ at β_τ, and*

$$H(s,r) = \lim_{\alpha\to\infty} \alpha P_{C,X}\left\{ \left[\tau - F_{T|X}(C \mid X)\right]^2 \left[I\left(\frac{X'r \vee X's}{\alpha} < C - X'\beta_\tau \le 0\right) + I\left(0 < C - X'\beta_\tau < \frac{X'r \wedge X's}{\alpha}\right)\right]\right\}$$

when it exists. Furthermore, $n^{1/3}(\hat{\beta}_{n,\epsilon} - \beta_\tau) \to_d \arg\inf \Gamma(s)$.

Theorem 2 follows by verifying the conditions of the main theorem from Kim and Pollard (1990). Provided that V is positive definite, we can conclude that $n^{1/3}(\hat{\beta}_{n,\epsilon} - \beta_\tau)$ converges to a nondegenerate distribution. A detailed proof is provided in Section 3.6.

Subsampling can produce consistent estimated sampling distributions for our estimator and it is an immediate consequence of Theorem 2.2.1 from Politis et al. (1999). In our study, we choose block size $b = N^\gamma$ where $\gamma = \{1/3, 1/2, 2/3, 3/4, 0.8, 5/6, 6/7, 0.9, 12/13, 0.95\}$ thus $b \to \infty$ and $b/N \to 0$ as $N \to \infty$. $n^{1/3}(\hat{\beta}_{n,\epsilon} - \beta_\tau)$ converges to a nondegenerate continuous distribution. All conditions in Theorem 2.2.1 from Politis et al. (1999) are met thus we can construct confidence intervals as stated in Section 3.2.3.

3.4 Numerical Studies

3.4.1 Simulation

Two simulation studies were carried out to test the finite sample performance of our estimator. We used conditional quantile functions which were linear in the covariate for each studies. In the first scenario, *Simulation 1*, the conditional quantile functions had identical linear coefficient and differed only in intercept. In the second scenario, *Simulation 2*, both the intercepts and covariate effects varied over the quantiles. *Simulation 1* represents a situation where the errors are independent and identically distributed and *Simulation 2* represents a situation where the errors are heteroscedastic.

In *Simulation 1*, the covariate is $X \equiv (1, X_1, X_2)'$ where $X_1 \sim$ Uniform $[0, 2]$ and $X_2 \sim$ Bernoulli(0.5). The unobserved failure times, T, were generated from the linear model, $T = 2 + 3\,X_1 + X_2 + 0.3\,U$. The observation times, C, were generated from the linear model, $C = 1.9 + 3.2\,X_1 + 0.8\,V$ when $X_2 = 0$ and $C = 3.1 + 2.8\,X_1 + 0.8\,V$ when $X_2 = 1$. Both U and V were generated from $N(0, 1)$. The proportion of events occurred prior to observation time, $\delta = 1$, was about 50%. The underlying 0.25 quantile is $Q_T(0.25|X) = 1.798 + 3\,X_1 + X_2$, the underlying 0.50 quantile is $Q_T(0.50|X) = 2 + 3\,X_1 + X_2$, and the underlying 0.75 quantile is $Q_T(0.75|X) = 2.202 + 3\,X_1 + X_2$. Since it is possible that T and/or C are negative, in a survival analysis context, we can treat T and C as the logarithm of survival time and logarithm of observation time, respectively. In *Simulation 2*, the covariate setup is the same as in *Simulation 1*. Unobserved failure times were generated from the linear model, $T = 2 + 3\,X_1 + X_2 + (0.2 + 0.5\,X_1)\,U$ and the observation times, C, were generated from the linear model $C = 1.8 + 3.2\,X_1 + 0.8\,X_2 + 0.8\,V$. Both U and V were generated from exponential distribution with rate equal to one. The proportion of events occurred prior to observation time, $\delta = 1$, was about 50%. The underlying 0.25 quantile is $Q_T(0.25|X) = 2.058 + 3.144\,X_1 + X_2$, the underlying 0.50 quantile is $Q_T(0.50|X) = 2.139 + 3.347\,X_1 + X_2$, and the underlying 0.75 quantile is $Q_T(0.75|X) = 2.277 + 3.693\,X_1 + X_2$. For each scenario, we reported the mean bias, mean squared error, and median absolute deviation based on 1000 simulations. Sample sizes were chosen to be $n = 200, 400,$ and 800 for each simulation setup. We interested in estimation of 0.25th, median, and 0.75th quantiles. Since the unobserved event time, T, and the observation time, C, were generated as a function of covariates and the error terms were generated from distributions which had positive density in the neighborhood of quantiles of interests, our simulation setup satisfy the identifiability conditions in Lemma 3.3.1.

For each simulated dataset, the procedure described at the end of Section 3.2.2 was used to estimate β. Symmetric confidence intervals as in Equation (3.15) were calculated based on a stochastic approximation with 500 subsamples. To decrease computational burden, the block size was determined via a pilot simulation in the same fashion as described in Banerjee and McKeague (2007). In a small scale simulation study, we examined the block size chosen by the algorithm described in Section 3.2.3 and by the pilot simulation method described in Banerjee and McKeague (2007). The block sizes chosen by either method produced similar average coverage which indicated the coverage presented in this section is a good representation of the coverage when confidence intervals are constructed using the algorithm described in Section 3.2.3. The optimal subsampling block size was determined from the following selected block sizes: $\{n^{1/3}, n^{1/2}, n^{2/3}, n^{3/4}, n^{0.8}, n^{5/6}, n^{6/7}, n^{0.9}, n^{12/13}, n^{0.95}\}$.

Table 4.5 and Table 3.2 summarize the results for *Simulation 1* and *Simulation 2* with sample size equal to 200, 400, and 800 at the 0.25th, median, and 0.75th quantiles. In the tables, "Truth" is the true parameter value; "Bias" is the mean bias of the estimates from all replicates; "MSE" is the mean squared error; "MAD" is the median absolute deviation of the estimates; "CP" is the average coverage from subsampling symmetric confidence intervals; and "Length" is the average confidence interval length. The tables show that the regression coefficient estimators have negligible bias.

In *Simulation 1*, the bias has a decreasing trend as the sample size increases for all quantiles and parameters. The mean squared errors and median absolute deviations decrease as the sample size increases for all quantiles and parameters. The subsampling confidence interval coverage is slightly lower than the nominal 95% level in smallest sample size (N=200) but the empirical coverage probability is close to 95% as the sample size increases. In *Simulation 2*, the bias for all quantiles is small for all sample sizes. There is a general decreasing trend for bias when the sample size increases. The mean squared errors and median absolute deviations decrease as the sample size increases for all quantiles and parameters. The average 95% confidence interval coverage rate is a bit low for the smallest sample size (N=200) but gets closer to the nominal 0.95 level as the sample size increases. In both scenarios, the median absolute deviations for sample size 800 is roughly 63% of the median absolute deviation for sample size 200 which is consistent with the cube-root rate.

The algorithm converge in all of our simulation studies. Non-convergence of the algorithm would be an indication that the data might not have sufficient information to support the estimation at the specified quantile. The computation time to estimate one quantile for each of the 100 simulated dataset ranged from 30 to 42 seconds for sample sizes

200 to 800 using a computer equipped with an Intel(R) Core(TM) i5-2500 CPU @3.30GHz 3.60 GHz CPU and 4.00 GB RAM. The computation times are similar for the two simulation scenario.

To illustrate the strength and limitation of the proposed method, an accelerated failure time (AFT) model with normally distributed errors was fit to the simulated datasets. Table 3.3 shows the results from the AFT models. When the error distribution in the AFT model is correctly specified, as in *Simulation 1*, the estimates have negligible bias. The true parameter values of the AFT model is the same as the true parameter value at the 0.5 quantile because the Normal distribution is symmetric; therefore, the conditional mean is the same as the conditional median. Since the error terms are correctly specified, the parametric method has higher efficiency than the proposed method which can be seen from the much smaller confidence interval length. When the error distribution is incorrectly specified, as in *Simulation 2*, the estimates are alarmingly biased. The coverage percentage is low for β_2 and is extremely low for both β_0 and β_1 even though the confidence intervals are narrow. Our method has a lower efficiency than the parametric method when the error distribution can be correctly specified in the parametric method. On the other hand, when the error distribution is incorrectly specified in the parametric method, our proposed method clearly outperforms the parametric method in terms of unbiased estimation and retaining proper coverage levels. The strength of our proposed method lies in the fact that it is a semiparametric method thus we do not need to know the true underlying distribution of the error terms in the AFT model.

Table 3.1: Simulation results for *Simulation 1*, based on 1000 simulation replicates.

N	τ	Parameter	Truth	Bias	MSE	MAD	CP	Length
200	0.25	β_0	1.798	0.026	0.036	0.117	0.921	0.679
		β_1	3.000	-0.010	0.022	0.097	0.941	0.542
		β_2	1.000	0.020	0.028	0.110	0.923	0.604
	0.50	β_0	2.000	0.002	0.032	0.120	0.931	0.633
		β_1	3.000	-0.006	0.018	0.088	0.946	0.500
		β_2	1.000	0.012	0.025	0.106	0.935	0.547
	0.75	β_0	2.202	-0.011	0.039	0.138	0.925	0.695
		β_1	3.000	-0.010	0.023	0.100	0.931	0.544
		β_2	1.000	0.010	0.029	0.109	0.927	0.602
400	0.25	β_0	1.798	0.007	0.020	0.097	0.942	0.530
		β_1	3.000	-0.002	0.012	0.074	0.958	0.415
		β_2	1.000	0.009	0.018	0.094	0.943	0.476
	0.50	β_0	2.000	-0.004	0.017	0.083	0.940	0.472
		β_1	3.000	0.002	0.010	0.066	0.944	0.372
		β_2	1.000	0.006	0.014	0.074	0.938	0.423
	0.75	β_0	2.202	-0.005	0.020	0.098	0.945	0.527
		β_1	3.000	-0.001	0.012	0.074	0.950	0.411
		β_2	1.000	-0.002	0.016	0.087	0.939	0.473
800	0.25	β_0	1.798	0.006	0.012	0.074	0.942	0.403
		β_1	3.000	-0.001	0.007	0.057	0.946	0.315
		β_2	1.000	0.002	0.010	0.068	0.941	0.365
	0.50	β_0	2.000	-0.003	0.010	0.065	0.939	0.367
		β_1	3.000	0.001	0.006	0.049	0.958	0.288
		β_2	1.000	0.004	0.009	0.062	0.944	0.338
	0.75	β_0	2.202	-0.005	0.012	0.073	0.937	0.406
		β_1	3.000	0.003	0.007	0.054	0.950	0.319
		β_2	1.000	-0.001	0.010	0.065	0.950	0.365

Truth is the true parameter value; Bias is mean of bias from 1000 replicates; MSE is mean squared error; MAD is median absolute deviation of the estimates; CP is the empirical coverage probabilities with a nominal level of 0.95 from subsampling symmetric confidence intervals with 500 subsamples; and Length is mean confidence interval length.

Table 3.2: Simulation results for *Simulation 2*, based on 1000 simulation replicates.

N	τ	Parameter	Truth	Bias	MSE	MAD	CP	Length
200	0.25	β_0	2.058	0.003	0.039	0.073	0.934	0.759
		β_1	3.144	0.010	0.020	0.087	0.956	0.563
		β_2	1.000	0.042	0.038	0.090	0.942	0.749
	0.50	β_0	2.139	0.027	0.030	0.109	0.938	0.662
		β_1	3.347	0.014	0.035	0.123	0.930	0.657
		β_2	1.000	0.021	0.040	0.126	0.937	0.745
	0.75	β_0	2.277	0.059	0.082	0.166	0.931	0.973
		β_1	3.693	-0.031	0.099	0.203	0.908	1.037
		β_2	1.000	0.006	0.110	0.209	0.939	1.284
400	0.25	β_0	2.058	0.020	0.012	0.053	0.953	0.439
		β_1	3.144	0.001	0.008	0.059	0.955	0.358
		β_2	1.000	0.003	0.013	0.063	0.957	0.457
	0.50	β_0	2.139	0.025	0.016	0.079	0.949	0.480
		β_1	3.347	0.011	0.019	0.096	0.935	0.495
		β_2	1.000	0.006	0.022	0.094	0.940	0.550
	0.75	β_0	2.277	0.039	0.045	0.128	0.952	0.768
		β_1	3.693	-0.016	0.059	0.158	0.934	0.843
		β_2	1.000	0.012	0.062	0.160	0.950	0.924
800	0.25	β_0	2.058	0.014	0.004	0.040	0.957	0.272
		β_1	3.144	0.001	0.004	0.041	0.954	0.253
		β_2	1.000	0.001	0.006	0.050	0.955	0.299
	0.50	β_0	2.139	0.017	0.009	0.059	0.954	0.350
		β_1	3.347	0.000	0.010	0.068	0.940	0.373
		β_2	1.000	0.003	0.012	0.070	0.950	0.414
	0.75	β_0	2.277	0.026	0.024	0.103	0.968	0.595
		β_1	3.693	-0.002	0.036	0.126	0.942	0.676
		β_2	1.000	0.009	0.037	0.124	0.948	0.707

Truth is the true parameter value; Bias is mean of bias from 1000 replicates; MSE is mean squared error; MAD is median absolute deviation of the estimates; CP is the empirical coverage probabilities with a nominal level of 0.95 from subsampling symmetric confidence intervals with 500 subsamples; and Length is mean confidence interval length.

Table 3.3: Results from accelerated failure time models with normal error, based on 1000 simulation replicates.

Simulation	N	Parameter	Truth	Bias	MSE	MAD	CP	Length
1	200	β_0	2	0.009	0.010	0.069	0.935	0.371
		β_1	3	-0.007	0.006	0.052	0.945	0.290
		β_2	1	0.001	0.008	0.057	0.933	0.332
	400	β_0	2	0.002	0.005	0.046	0.951	0.263
		β_1	3	-0.001	0.003	0.035	0.955	0.205
		β_2	1	-0.001	0.004	0.040	0.948	0.236
	800	β_0	2	-0.0004	0.002	0.033	0.954	0.187
		β_1	3	0.0001	0.001	0.024	0.948	0.146
		β_2	1	0.001	0.002	0.020	0.946	0.167
2	200	β_0	2.2	0.269	0.090	0.268	0.497	0.550
		β_1	3.5	0.326	0.119	0.323	0.149	0.426
		β_2	1	0.061	0.021	0.095	0.928	0.496
	400	β_0	2.2	0.281	0.088	0.283	0.185	0.389
		β_1	3.5	0.317	0.107	0.315	0.016	0.301
		β_2	1	0.056	0.011	0.071	0.913	0.351
	800	β_0	2.2	0.275	0.080	0.276	0.027	0.274
		β_1	3.5	0.318	0.105	0.318	0	0.212
		β_2	1	0.061	0.008	0.065	0.833	0.247

Truth is the true parameter value; Bias is mean of bias from 1000 replicates; MSE is mean squared error; MAD is median absolute deviation of the estimates; CP is the empirical coverage probabilities with a nominal level of 0.95; and Length is mean confidence interval length.

3.4.2 Application

We applied the proposed method to analyze "The Voluntary HIV-1 Counseling and Testing Efficacy Study Group" data. The detailed study design and outcome of the original clinical trials are described in Kamenga et al. (2000) and Coates et al. (2000), respectively. In this study, 3120 individuals and 586 couples were enrolled in Kenya, Tanzania, and Trinidad. Individual or couple participants were randomly assigned to HIV-1 voluntary counseling and testing (VCT) or basic health information (BHI) group. At the first follow-up (around 6 months after baseline), sexually transmitted diseases (STD) were diagnosed and treated. No further testing for sexually transmitted diseases was performed after the first follow-up.

To remove the correlation effects in couples, only females (N=2172) were included in our analysis. Furthermore, we excluded females who did not have the first follow-up (N=377), who had STD symptoms at baseline (N=876), and who had missing STD outcomes at their first follow-up (N=16). The final analysis included 903 females aged 17-66 (median=26) of which 48.5% (N=438) were in the VCT group.

The first follow-up was between 121 and 582 days (median=198 days) after the baseline visit. Histograms for the numbers of days between the baseline visit and first follow-up for the VCT and BHI groups are shown in Figure 3.9. There does not appear to be a difference in follow-up time between the VCT and BHI groups. In our analysis, the outcome "Any STD" is defined as any positive lab results or any self-reported STD symptom presented at the first follow-up. The STD tested were *Trichomonas vaginalis*, *Neisseria gonorrhoeae*, *Chlamydia trachomatis*, and syphilis. Self-reported STD symptoms included pain or burning around vagina when urinating, non-traumatic sores or boils around vagina, itching around vagina, abnormal vaginal discharge, and pain in the bottom of stomach (not related to a menstrual period or using an IUD). Among 903 females, 333 (26.9%) contracted "Any STD" by first follow-up (141 from the VCT group and 192 from the BHI group). The (unobserved) failure time of interest was the number of days to "Any STD" contraction after the baseline visit. The analysis examined the effect of VCT versus BHI and the effect of participants' age on the quantiles of time to "Any STD."

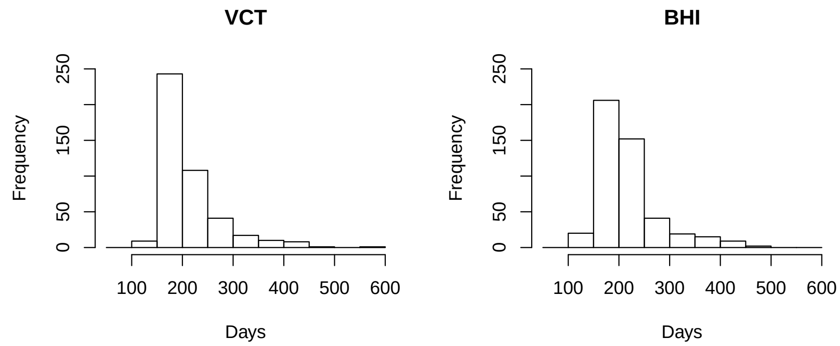

Figure 3.9: Days between baseline visit and first follow-up for 903 females in voluntary counseling and testing (VCT) and basic health information (BHI).

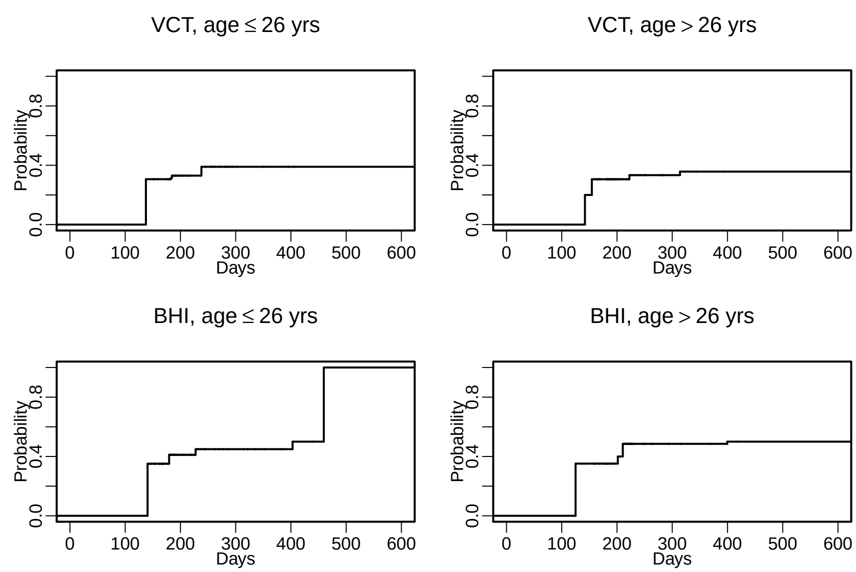

Figure 3.10: Nonparametric maximum likelihood estimator (NPMLE) of the (unobserved) failure time distribution function stratified by voluntary counseling and testing (VCT) vs. basic health information (BHI) and age (below or above median age).

Before applying our proposed method on the data, nonparametric maximum likelihood estimator (NPMLE) of the (unobserved) failure time distribution function was carried out for the data (Wellner and Zhan 1997, Gentleman and Vandal 2011). NPMLE was used to determine whether certain quantiles can be reasonably estimated based on the data available. The NPMLE of distribution function stratified by VCT versus BHI and by age (below or above median age) is shown in Figure 3.10.

Based on the NPMLE in Figure 3.10, it is clear that the data may provide enough information for estimation only for quantiles less than 0.4th for VCT group; thus, we focused only on 0.05 to 0.35 quantiles per 0.05 increments when performing data analysis. We centered the age at 30 then divided it by 5. We also divided the time between baseline visit and first follow-up by 30.5 to convert the time form days to months. Our proposed model was fitted for the lower quantiles and symmetric confidence intervals were constructed by subsampling where the block size was chosen based on the algorithm presented in Section 3.2.3. The results are summarized in Table 3.4 and Figure 3.11.

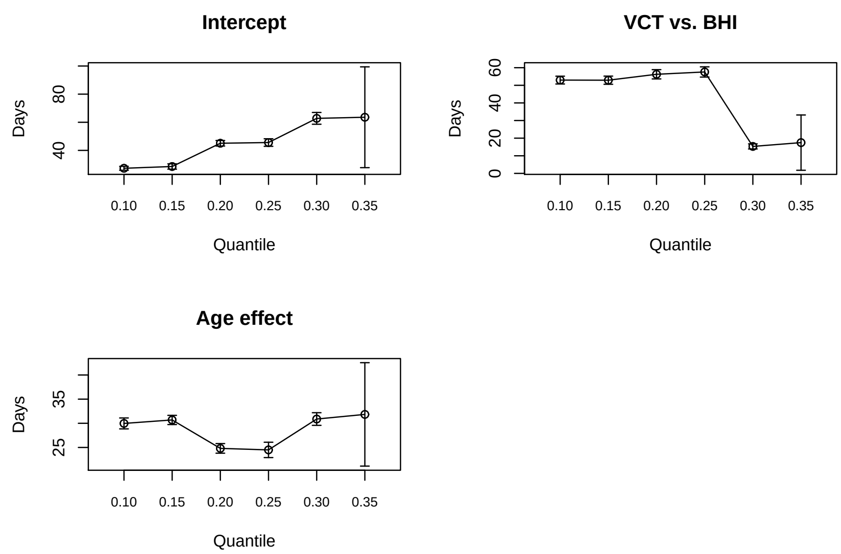

Figure 3.11: HIV data: effect on delaying STD contraction. VCT stands for voluntary counseling and testing and BHI stands for basic health information. The vertical bars are symmetric confidence interval constructed using a subsampling method.

The results indicated that VCT had statistically significant delay effect in STD contraction compared to BHI at 0.05, 0.15, 0.20, and 0.35 quantiles. Older age also showed

Table 3.4: Results of analyzing "The Voluntary HIV-1 Counseling and Testing Efficacy Study Group" data, effect on delaying STD contraction in months.

Quantile	Parameter	Estimate	Lower C.I.	Upper C.I.
0.05	Intercept	3.103	-1.129	7.334
	VCT vs. BHI	1.538	0.151	2.924
	(Age – 30)/5	0.328	-0.995	1.651
0.10	Intercept	3.130	-1.753	8.012
	VCT vs. BHI	1.509	-0.019	3.037
	(Age – 30)/5	0.328	-1.127	1.783
0.15	Intercept	1.295	-2.192	4.781
	VCT vs. BHI	1.616	0.108	3.125
	(Age – 30)/5	0.948	-0.019	1.915
0.20	Intercept	1.303	-1.090	3.696
	VCT vs. BHI	1.697	0.174	3.221
	(Age – 30)/5	0.933	0.308	1.558
0.25	Intercept	1.305	-2.053	4.664
	VCT vs. BHI	1.639	-2.017	5.295
	(Age – 30)/5	0.937	-1.211	3.085
0.30	Intercept	1.215	-1.613	4.042
	VCT vs. BHI	0.424	-5.817	6.665
	(Age – 30)/5	2.856	0.697	5.015
0.35	Intercept	2.563	-0.258	5.383
	VCT vs. BHI	6.774	2.363	11.184
	(Age – 30)/5	2.203	0.490	3.915

VCT stands for voluntary counseling and testing; BHI stands for basic health information.

a significant delay in STD contraction at 0.20, 0.30, and 0.35 quantile. Specifically, compared to BHI, participants in VCT delayed STD contraction by about 1.5-1.7 months for 0.05 to 0.20 quantiles and by about 6 months at 0.35 quantiles. Each 5 years increase in age delayed STD contraction by around 0.5 month at the lower quantiles examined to around 2.5 months at higher quantiles albeit the effects were only significant at some quantiles. The effect of VCT in delaying STD contraction had a pretty constant effect at lower quantiles among quantiles examined and the effect changed more dramatically at higher quantiles. The dramatic changes at the higher quantile may suggest some identifiability issue for the higher quantile. In theory, the intercept term should be non-decreasing but our point estimate for the intercept did not have such a pattern. From Figure 3.10 we can see that from 0 to 0.4 quantile, the NPMLE only have one or two jumps which indicate that there is very limited information available to distinguish different quantiles if they are within the same jump. Thus, our estimated intercepts did not have a non-decreasing pattern and they all had wide confidence intervals. Quantile regression models provided two unique advantages in analyzing the HIV data. First, compared to logistic regression which is the analysis method of choice for the original clinical trial (Coates et al. 2000), quantile regression provided a more intuitive interpretation of the VCT effect. Rather than presenting an odds ratio, quantile regression allows us to interpret the VCT effect as days in delaying STD contraction. Second, quantile regression allowed us to examine the VCT effect in different quantiles. Specifically, it might be of interest to know the effect of VCT in the lower quantiles as we might hope VCT has a greater delaying effect in lower quantiles. Based on our results, VCT indeed have a statistical significant effect in most lower quantiles examined; thus, it indicated that VCT was effective to reduce unprotected sexual intercourse in most of the lower quantiles. For example, 0.05 quantile of time to STD contraction for 30 years old females in VCT group were 1.5 months higher than females in BHI group albeit it had no statistical significant effect in changing the odd of STD contraction (Coates et al. 2000).

3.5 Discussion

To solve the non-convex objective function in Equation (3.12), we used the difference between two convex hinge functions Equation (3.13) to approximate the objective function. One practical issue is how to choose a good initial value. Currently, we used a coarse grid search to generate the initial value. A grid search can be done in low dimension data but it is not practical when the data is high dimensional. Further work is needed to investigate a practical method to produce reasonable initial value for high dimensional data.

In real data analysis, it may be the case that not all quantiles are estimable. It is not due to the estimation procedure but the sparse data structure. Consider a situation where a disease requires a long incubation period, if the observation times are all concentrated in a short period, most subjects would not have developed any symptoms yet. We will have most people at $\delta = 0$ at the end of the observation period. In this case, the higher tail quantiles will not be estimable because we simply do not have enough information. We recommend to obtain a NPMLE of the cumulative density function stratified by covariates as we did for the HIV data. The NPMLE results can provide useful information about what quantiles can be reasonably estimated.

The method proposed in this paper can be easily extended to type II interval-censored data. For example, suppose the event occurred in interval $(L, R]$, we can simply treat this as 2 records in current status data format. The first record would have $C = L,$ and $\delta = 0$ and the second record would have $C = R,$ and $\delta = 1$ then the same optimization routine can be carried out for estimation. Extension to right censored data is also possible. Intuitively, the non-censored observations can be treated as the event occurred within a very small interval and right-censored observation can be treated as current status with C equals the censoring time and $\delta = 0$. This extension will not require the "global linearity assumption" which is commonly assumed in existing quantile regression models for right-censoring data (Portnoy 2003, Peng and Huang 2008). Furthermore, models with varying coefficient or nonparametric quantile regression model may be useful for practical purposes which warrant future investigation.

3.6 Proof of lemma and theorems

Proof of Lemma 3.3.1.

$$
\begin{aligned}
Z(\beta) - Z(\beta_\tau) = P_{C,X} \Big[& \Big\{ \tau I(\delta = 0) I(X^T\beta \le C) + (1-\tau) I(\delta = 1) I(X^T\beta > C) \Big\} \\
& - \Big\{ \tau I(\delta = 0) I(X^T\beta_\tau \le C) + (1-\tau) I(\delta = 1) I(X^T\beta_\tau > C) \Big\} \Big] \\
= P_{C,X} & \Big[\{\tau I(\delta = 0) - (1-\tau) I(\delta = 1)\} \Big\{ I(X^T\beta \le C < X^T\beta_\tau) - I(X^T\beta_\tau \le C < X^T\beta) \Big\} \Big] \\
= P_{C,X} & \Big[\{\tau - I(\delta = 1)\} \Big\{ I(X^T\beta \le C < X^T\beta_\tau) - I(X^T\beta_\tau \le C < X^T\beta) \Big\} \Big] \\
= P_{C,X} & \Big[\{\tau - F_{T|X}(C|X)\} \Big\{ I(X^T\beta \le C < X^T\beta_\tau) - I(X^T\beta_\tau \le C < X^T\beta) \Big\} \Big] \\
= \int_X \int_{C|X} & \{\tau - F_{T|X}(C \mid X)\} \\
& \Big\{ I(X^T\beta \le C < X^T\beta_\tau) - I(X^T\beta_\tau \le C < X^T\beta) \Big\} \, dP_{C|X} \, dP_X \\
= \int_X \int_{(X^T\beta \le C < X^T\beta_\tau) \bigcup (X^T\beta_\tau \le C < X^T\beta)} & \left| \int_{X^T\beta_\tau}^{c} f_{T|X}(t) \, f_{C|X}(c) \, dt \right| dc \, dP_X
\end{aligned}
$$

Condition 2 insures that the integrand of the inner integral is strictly positive. We can then apply Theorem 1.6.6 (b) from Ash and Doléans-Dade (2000) which states "Let h be Borel measurable. If $h \ge 0$ and $\int_\Omega h d\mu = 0$, then $h = 0$ a.e.". Since the integrand, $f_{T|X}(t) \, f_{C|X}(c)$, is positive, we can conclude that the inner most integral is positive using a contrapositive argument. Applying the same theorem two more times, along with the full rank condition in Condition 1, we can then conclude that, $Z(\beta) - Z(\beta_\tau) > 0$ for all $\beta \ne \beta_\tau$ and hence, β_τ is identifiable. $\square$

Remark It is true that the derivative of $Z(\beta)$ with respect to β is zero at the true β. $Z(\beta)$ is defined as

$$
\begin{aligned}
Z(\beta) &= E \Big\{ \tau I(\delta = 0) I(X^T\beta - C \le 0) + (1-\tau) I(\delta = 1) I(X^T\beta - C > 0) \Big\} \\
&= \int_X \left(\int_{X^T\beta}^{\infty} \tau \, dF_{C|X} + \int_0^{X^T\beta} F_{T|X}(c) \, dF_{C|X} - \tau \int_0^{\infty} F_{T|X}(c) \, dF_{C|X} \right) dF_X .
\end{aligned}
$$

We can take the derivative of $Z(\beta)$ with respect to β as,

$$\frac{\partial}{\partial\beta}Z(\beta) = \int_X \left\{-\tau f_{C|X}(X^T\beta)X + F_{T|X}(X^T\beta)\, f_{C|X}(X^T\beta)X\right\} dF_X.$$

By definition, $F_{T|X}(X^T\beta_\tau) = \tau$, so it is immediate that $\frac{\partial}{\partial\beta}Z(\beta)\mid_{\beta=\beta_\tau} = 0$.

Proof of Theorem 3.3.2. We shall prove this theorem by showing

$$\sup_{\beta\in\mathcal{B}} |Z_{n,\epsilon}(\beta) - Z(\beta)| \to 0 \quad \text{almost surely} \tag{3.16}$$

and then showing $Z(\beta)$ is continuous. By Lemma 3.3.1, β_τ is the unique minimizer of $Z(\cdot)$ and since $\mathcal{B}$ is assumed, we can use Theorem 2.1 of Newey and McFadden (1994) to conclude that $\hat\beta_{n,\epsilon} \to \beta_\tau$ in probability.

We can show Equation (4.33) is true by proving

$$\sup_{\beta\in\mathcal{B}} |Z_n(\beta) - Z(\beta)| \to 0 \quad \text{almost surely}, \qquad \sup_{\beta\in\mathcal{B}} |Z_{n,\epsilon}(\beta) - Z_n(\beta)| \to 0 \quad \text{almost surely}$$

since

$$\sup_{\beta\in\mathcal{B}} |Z_{n,\epsilon}(\beta) - Z(\beta)| \le \sup_{\beta\in\mathcal{B}} |Z_{n,\epsilon}(\beta) - Z_n(\beta)| + \sup_{\beta\in\mathcal{B}} |Z_n(\beta) - Z(\beta)| \tag{3.17}$$

The class of indicator functions $I(\delta = 0)$, $I(\delta = 1)$, $\mathcal{I}_1 \equiv \{I(X^T\beta - C \le 0) : \beta \in \mathcal{B}\}$, and $\mathcal{I}_2 \equiv \{I(X^T\beta - C > 0) : \beta \in \mathcal{B}\}$ are examples of Vapnik-Červonenkis (VC)-subgraph classes. τ and $1-\tau$ are fixed functions and thus by Lemma 2.6.18 (i) and (vi) of van der Vaart and Wellner (1996), the classes $\tau\, I(\delta = 0)\,\mathcal{I}_1$ and $(1-\tau)\, I(\delta = 1)\,\mathcal{I}_2$ are also VC-subgraph classes. Finally, (v) of the same lemma gives that $\mathcal{Z} \equiv \{Z_n(\beta) : \beta \in \mathcal{B}\}$ is a VC-subgraph class. Since $\mathcal{Z}$ is a VC-subgraph class, it is also a Glivenko-Cantelli class; hence, $\sup_{\beta\in\mathcal{B}} |Z_n(\beta) - Z(\beta)| \to 0$ almost surely.

Since $I(|X_i^T\beta - C_i| \le \epsilon/2)$ is a VC class of functions, $\mathbf{P}_n\{I(|X_i^T\beta - C_i| \le \epsilon/2)\}$ converges to $P\{I(|X_i^T\beta - C_i| \le \epsilon/2)\}$ uniformly over $\mathcal{B}$ where $\mathbf{P}_n$ is the empirical measure and P is the true underlying measure. Thus, we have

$$
\begin{aligned}
\sup_{\beta\in\mathcal{B}} |Z_{n,\epsilon}(\beta) - Z_n(\beta)| &\leq \sup_{\beta\in\mathcal{B}} \mathbf{P}_n\{I(|X_i^T\beta - C_i| \leq \epsilon/2)\} \\
&\underset{n\uparrow\infty}{\longrightarrow} \sup_{\beta\in\mathcal{B}} P\{I(|X^T\beta - C| \leq \epsilon/2)\} \\
&= \sup_{\beta\in\mathcal{B}} P_X P_{C|X}(|X^T\beta - C| \leq \epsilon/2 \mid X) \\
&\leq P_X(\epsilon M_C \mid X) = \epsilon M_C
\end{aligned} \tag{3.18}
$$

By Condition 4, Equation (4.35) is bounded and converges to 0 as $\epsilon \to 0$, thus we can conclude that $\sup_{\beta\in\mathcal{B}} |Z_{n,\epsilon}(\beta) - Z_n(\beta)| \to 0$ almost surely as $n \to \infty$ and $\epsilon \to 0$.

Since each term on the right hand side of Equation (4.34) converges to 0 almost surely, we can conclude that $\sup_{\beta\in\mathcal{B}} |Z_{n,\epsilon}(\beta) - Z(\beta)| \to 0$ almost surely as $n \to \infty$ and $\epsilon \to 0$.

To show that $Z(\cdot)$ is continuous, we re-express $Z(\cdot)$ as

$$
\begin{aligned}
Z(\beta) &= E\left\{\tau I(\delta = 0)I(X^T\beta - C \leq 0) + (1-\tau)I(\delta = 1)I(X^T\beta - C > 0)\right\} \\
&= \int_X \left[\tau \int_{X^T\beta}^{\infty} \left\{1 - F_{T|X}(c)\right\} dF_{C|X} + (1-\tau)\int_0^{X^T\beta} F_{T|X}(c) dF_{C|X}\right] dF_X \\
&= \int_X \left(\int_{X^T\beta}^{\infty} \tau\, dF_{C|X} + \int_0^{X^T\beta} F_{T|X}(c)\, dF_{C|X} - \tau \int_0^{\infty} F_{T|X}(c)\, dF_{C|X}\right) dF_X.
\end{aligned} \tag{3.19}
$$

Only the first two inner integrals are functions of β. Under Condition 4, both of these inner integrals are bounded and continuous with respect to β; therefore, $Z(\cdot)$ is continuous. □

Proof of Theorem 3.3.3. Before proceeding with the proof, we will state the main theorem from Kim and Pollard (1990). The theorem concerns estimators defined by minimization of process $\mathbf{P}_n g(\cdot, \theta) = \frac{1}{n}\sum_{i\leq n} g(\xi_i, \theta)$, where $\{\xi_i\}$ is a sequence of independent observations taken from a distribution P and $\{g(\cdot, \theta) : \theta \in \Theta\}$ is a class of functions indexed by a subset

Θ of $\Re^d$. $\mathbf{P}_n$ denotes the expectation with respect to the empirical process. The envelope $G_R(\cdot)$ is defined as the supremum of $|g(\cdot,\theta)|$ over the class $\mathbf{g}_R = \{g(\cdot,\theta) : |\theta - \theta_o| \le R\}$.

Kim and Pollard (1990) *Let* $\{\theta_n\}$ *be a sequence of estimators for which*

(i) $\mathbf{P}_n g(\cdot\,,\theta_n) \le \inf_{\theta\in\Theta} \mathbf{P}_n g(\cdot,\theta) + o_p(n^{-2/3})$.

Suppose

(ii) θ_n *converges in probability to the unique* θ_0 *that minimizes* $Pg(\cdot,\theta)$*;*

(iii) θ_0 *is an interior point of* Θ.

Let the functions be standardized so that $g(\cdot,\theta_0) \equiv 0$. *If the classes* $\mathbf{g}_R$, *for* R *near 0, are uniformly manageable for the envelopes* G_R *and satisfy*

(iv) $Pg(\cdot,\theta)$ *is twice differentiable with second derivative matrix* V *at* θ_0*;*

(v) $H(s,t) = \lim_{\alpha\to\infty} \alpha P g(\cdot,\theta_0 + s/\alpha) g(\cdot,\theta_0 + t/\alpha)$ *exists for each* s, t *in* $\Re^d$ *and*

$$\lim_{\alpha\to\infty} \alpha P g(\cdot,\theta_0 + t/\alpha)^2 I\left\{|g(\cdot,\theta_0 + t/\alpha)| > \epsilon\alpha\right\} = 0$$

for each $\epsilon > 0$ *and* t *in* $\Re^d$*;*

(vi) $PG_R^2 = O(R)$ *as* $R \to 0$ *and for each* $\epsilon > 0$ *there is a constant* K *such that* $PG_R^2 I\{G_R > K\} < \epsilon R$ *for* R *near 0;*

(vii) $P|g(\cdot,\theta_1) - g(\cdot,\theta_2)| = O(|\theta_1 - \theta_2|)$ *near* θ_0*;*

then the process $n^{2/3}\mathbf{P}_n g(\cdot,\theta_0 + tn^{-1/3})$ *converges in distribution to a Gaussian process* $Z(t)$ *with continuous sample paths, expected value* $t^T V t/2$ *and covariance kernel* H.

If V *is positive definite and if* Z *has nondegenerate increments, then* $n^{1/3}(\theta_n - \theta_0)$ *converges in distribution to the (almost surely unique) random vector that minimizes* $Z(t)$.

Now we proceed with our proof of theorem 2. It will be convenient to define a version of the original objective function centered at the true value β_τ,

$$\begin{aligned}
g(\beta) = & \left\{\tau I(\delta=0) I(X^T\beta \le C) + (1-\tau) I(\delta=1) I(X^T\beta > C)\right\} \\
& \quad - \left\{\tau I(\delta=0) I(X^T\beta_\tau \le C) + (1-\tau) I(\delta=1) I(X^T\beta_\tau > C)\right\} \\
& = \{\tau I(\delta=0) - (1-\tau) I(\delta=1)\} \left\{I(X^T\beta \le C < X^T\beta_\tau) - I(X^T\beta_\tau \le C < X^T\beta)\right\} \\
& = \{\tau - I(\delta=1)\} \left\{I(X^T\beta \le C < X^T\beta_\tau) - I(X^T\beta_\tau \le C < X^T\beta)\right\}.
\end{aligned}$$

Under the true distribution P, we have$P(g(\beta)) = Z(\beta) - Z(\beta_\tau)$. The minimum value of $P(g(\cdot))$ is then obtained at the $\arg\min$ of $Z(\cdot)$ and $P(g(\beta_\tau)) = 0$. The estimator we use here is $\hat{\beta}_{n,\epsilon}$ which is the minimizer of $Z_{n,\epsilon}(\cdot)$ defined as in (3.13).

The first condition of the main theorem from Kim and Pollard (1990) is satisfied under Condition 5. By the definition of $\hat{\beta}_{n,\epsilon}$, we have $Z_{n,\epsilon}(\hat{\beta}_{n,\epsilon}) \le \inf_\beta Z_{n,\epsilon}(\beta) + o_p(n^{-2/3})$. We also have $Z_n(\hat{\beta}_{n,\epsilon}) - \inf_\beta Z_n(\beta) \ge 0$ by definition.

$$\begin{aligned}
Z_n(\hat{\beta}_{n,\epsilon}) - \inf_\beta Z_n(\beta) & \le Z_{n,\epsilon}(\hat{\beta}_{n,\epsilon}) + \mathbf{P}_n\{I(|C - X^T\beta_{n,\epsilon}| < \epsilon/2)\} \\
& \qquad - \inf_{\beta\in\beta}\left[Z_{n,\epsilon}(\beta) - \mathbf{P}_n\{I(|C - X^T\beta| < \epsilon/2)\}\right] \\
& \le Z_{n,\epsilon}(\hat{\beta}_{n,\epsilon}) - \inf_\beta Z_{n,\epsilon}(\beta) + 2\sup_\beta \mathbf{P}_n\{I(|C - X^T\beta| < \epsilon/2)\} \\
& = Z_{n,\epsilon}(\hat{\beta}_{n,\epsilon}) - \inf_\beta Z_{n,\epsilon}(\beta) + 2\sup_\beta P(|C - X^T\beta| < \epsilon/2) + o_p(\epsilon\, n^{-1/2}) \\
& \le Z_{n,\epsilon}(\hat{\beta}_{n,\epsilon}) - \inf_\beta Z_{n,\epsilon}(\beta) + 2\epsilon M_C + o_p(n^{-7/6}) = o_p(n^{-2/3})
\end{aligned}$$

Therefore, $Z_n(\hat{\beta}_{n,\epsilon}) \le \inf_\beta Z_n(\beta) + o_p(n^{-2/3})$ which satisfied the first condition. The second condition, $\hat{\beta}_{n,\epsilon} \to \beta_\tau$ in probability, has been verified in Theorem 3.3.2. The third condition is satisfied by assuming Condition 2.

The remaining four conditions of the theorem deal with the nature of expectation of g

under the measure P. $P(g)$ may be expressed as

$$
\begin{aligned}
P(g(\beta)) = \; & P\left[\{\tau - I(\delta = 1)\}\left\{I(X^T\beta \le C < X^T\beta_\tau) - I(X^T\beta_\tau \le C < X^T\beta)\right\}\right] \\
= \; & P_{C,X}\left\{P\left[\{\tau - I(\delta = 1)\} \mid C, X\right]\left\{I(X^T\beta \le C < X^T\beta_\tau) - I(X^T\beta_\tau \le C < X^T\beta)\right\}\right\} \\
= \; & P_{C,X}\left[\left\{\tau - F_{T|X}(C \mid X)\right\}\left\{I(X^T\beta \le C < X^T\beta_\tau) - I(X^T\beta_\tau \le C < X^T\beta)\right\}\right]
\end{aligned}
$$

where $F_{T|X}(\cdot \mid \cdot)$ is the conditional distribution of T given X. This expectation is dominated by:

$$
\begin{aligned}
P|g(\beta)| = P & \left|\left\{\tau - F_{T|X}(C \mid X)\right\}\left\{I(X^T\beta \le C < X^T\beta_\tau) - I(X^T\beta_\tau \le C < X^T\beta)\right\}\right| \\
& \le P\left\{I(X^T\beta \le C < X^T\beta_\tau) + I(X^T\beta_\tau \le C < X^T\beta)\right\} \le 1.
\end{aligned}
$$

Since P is absolutely continuous with respect to Lebesgue measure, for any sequence $d_n \to 0$, the dominated convergence theorem tells us $P(g(\beta + d_n)) \to P(g(\beta))$. In other words, $P(g(\beta))$ is continuous with respect to β.

We may expand $P(g(\beta))$ with a Taylor expansion. The first derivative is found by interchanging integration (expectation) and differentiation to find

$$
\begin{aligned}
\frac{\partial}{\partial\beta_i}P(g(\beta)) &= \int_X \left\{\tau - F_{T|X}(X^T\beta \mid X)\right\} X_i\, f_{C|X}(X^T\beta \mid X) \\
&\qquad\qquad \{-I(X^T\beta < X^T\beta_\tau) - I(X^T\beta_\tau < X^T\beta)\}\, dP_X \\
&= \int_X \left\{F_{T|X}(X^T\beta \mid X) - \tau\right\} X_i\, f_{C|X}(X^T\beta \mid X)\, dP_X,
\end{aligned}
$$

where $f_{C|X}$ is the density of the observation time C conditioned on X and X_i is an element of $\mathbf{X_i}$. Evaluated at β_τ, the term $F_{T|X}(\beta_\tau X \mid X) - \tau$ equal to zero by definition of the τth quantile, making the derivative equal zero as would be expected for an extrema.

Taking one step further, the second derivative would be

$$
\begin{aligned}
V(\beta)_{i,j} &= \frac{\partial^2}{\partial\beta_i\partial\beta_j} P(g(\beta)) \\
&= \int_X \left(F_{T|X}(X^T\beta \mid X) - \tau\right) X_i X_j \frac{\partial}{\partial C} f_{C|X}(X^T\beta \mid X)\, dP_X \\
&\quad + \int_X X_i X_j f_{C|X}(X^T\beta \mid X) f_{T|X}(X^T\beta \mid X)\, dP_X.
\end{aligned}
$$

At β_τ, the first integral vanishes and only the second remains taking the form

$$
V(\beta_\tau)_{i,j} = \int_X \left\{ X_i X_j f_{C|X}(X^T\beta_\tau \mid X) f_{T|X}(X^T\beta_\tau \mid X) \right\} dP_X
$$

As the entries are dominated by

$$
|V(\beta_\tau)_{i,j}| \leq M_C M_T M_{|X|}^2,
$$

where $M_{|X|}$ is the bound over all $|X_i|$ and M_C and M_T are defined in Condition 4. $V(\beta_\tau)_{i,j}$ will be well defined verifying the fourth condition of the theorem. Writing

$$
V(\beta_\tau) = P_X(XX^T h(X)), \text{ with } h(X) = f_{C|X}(X^T\beta \mid X) f_{T|X}(X^T\beta \mid X) \geq 0,
$$

show that $V(\beta_\tau)$ would be a symmetric positive semi-definite matrix since it is a positive mixture of the positive semi-definite terms XX^T. A sufficient condition for $V(\beta_\tau)$ to be positive definite is that the Lebesgue measure of the set $\{X : f_{C|X}(X^T\beta_\tau \mid X) f_{T|X}(X^T\beta_\tau \mid X) f_X(X) > 0\}$ is greater than zero.

To control asymptotic covariance of $Z(s)$, let

$$
\begin{aligned}
H(s,r) &= \lim_{\alpha\to\infty} \alpha P\{g(\beta_\tau + r/\alpha) g(\beta_\tau + s/\alpha)\} \\
&= \lim_{\alpha\to\infty} \alpha P_{C,X}\Bigg\{ \{\tau - F_{T|X}(C \mid X)\}^2 \\
&\qquad \Big[I\{X^{T}(\beta_\tau + r/\alpha) \le C < X^{T}\beta_\tau\} - I\{X^{T}\beta_\tau \le C < X^{T}(\beta_\tau + r/\alpha)\}\Big] \\
&\qquad \Big[I\{X^{T}(\beta_\tau + s/\alpha) \le C < X^{T}\beta_\tau\} - I\{X^{T}\beta_\tau \le C < X^{T}(\beta_\tau + s/\alpha)\}\Big] \Bigg\} \\
&= \lim_{\alpha\to\infty} \alpha P_{C,X}\Bigg[\{\tau - F_{T|X}(C \mid X)\}^2 \\
&\qquad \left\{ I\left(\frac{X^{T}r \vee X^{T}s}{\alpha} \le C - X^{T}\beta_\tau < 0\right) + I\left(0 \le C - X^{T}\beta_\tau < \frac{X^{T}r \wedge X^{T}s}{\alpha}\right)\right\}\Bigg],
\end{aligned}
$$

where $\vee$ and $\wedge$ denote maximum and minimum, respectively. Using Condition 4 and 7, we have

$$
\begin{aligned}
&P_{C,X}\Bigg[\{\tau - F_{T|X}(C \mid X)\}^2 \left\{ I\left(\frac{X^{T}r \vee X^{T}s}{\alpha} \le C - X^{T}\beta_\tau < 0\right) + \right. \\
&\qquad\qquad \left. I\left(0 \le C - X^{T}\beta_\tau < \frac{X^{T}r \wedge X^{T}s}{\alpha}\right)\right\}\Bigg] \\
&\quad \le \int_X M_C \frac{|X^{T}r| \wedge |X^{T}s|}{\alpha}\, dP_X \le M_C \frac{(\|s\|_1 + \|r\|_1) M_{|X|}}{\alpha} = O(\alpha^{-1}),
\end{aligned}
$$

hence, along with Condition 6, $H(s,r)$ is well defined by the dominated convergence theorem satisfying the fifth condition.

Let G_R be the envelope of $\mathbf{g}_R \equiv \{g(\beta) : \|\beta - \beta_\tau\|_\infty < R \le \tilde{\epsilon}\}$, i.e.,

$$
G_R = |\tau - I(\delta = 1)|\, I(|C - X^{T}\beta_\tau| < R \max |X_j|) \le I(|C - X^{T}\beta_\tau| < R \max |X_j|)
$$

A sufficient condition for the class $\mathbf{g}_R$ to be uniformly manageable is that its envelope function G_R is uniformly square integrable given that $\{g(\beta)\}$ is VC-subgraph (Mohammadi and Van De Geer 2005). Since G_R is bounded by one, it is uniformly square integrable for R close to zero. Together with the fact that $\{g(\beta)\}$ is VC-subgraph, we conclude that $\mathbf{g}_R$ is uniformly manageable. Then

$$\begin{aligned} P(G_R^2) &\leq \int_X \int_{\Re} I(|C - X^T \beta_\tau| < R \max |X_j|)\, dP_{C|X}\, dP_X \\ &\leq \int_X 2M_C R \max(|X_j|)\, dP_X \leq R(2M_C M_{|X|}) = O(R). \end{aligned}$$

For any $\epsilon > 0$, we can use $K = 2$, then $E\{G_R^2 I(G_R > K)\} = 0 < \epsilon R$ since G_R is less than K everywhere. Combining these two traits satisfying the sixth condition of the theorem.

The final condition is verified by letting $G_{R,\beta}$ be the envelope of $\{g(\tilde{\beta}) - g(\beta) : \|\tilde{\beta} - \beta\|_\infty < R\}$, i.e.,

$$G_{R,\beta} \;=\; |\tau - I(\delta = 1)|\, I(|C - X^T \beta| < R \max |X_j|) \;\leq\; I(|C - X^T \beta| < R \max |X_j|).$$

Using the same integration inequalities as used in the preceding for G_R we find that $P|g(\tilde{\beta}) - g(\beta)| = O(\|\tilde{\beta} - \beta\|_\infty) = O(\|\tilde{\beta} - \beta\|_1)$ over all $\beta, \tilde{\beta}$ in an $\tilde{\epsilon}$ neighborhood of β_τ since $\|\cdot\|_\infty$ and $\|\cdot\|_1$ are equivalent metrics.

As the seven conditions are satisfied, the conclusion of the main theorem in Kim and Pollard (1990) follows which in turn proved Theorem 2. $\square$

CHAPTER 4: QUANTILE REGRESSION MODELS FOR CASE II INTERVAL-CENSORED FAILURE TIME DATA

4.1 Introduction

Quantile regression (Koenker and Bassett 1978) describes how covariates influence the location, scale, and shape of a response distribution. It specifies changes in the conditional quantile as a parametric function of explanatory variables and it is a more stable candidate to describe central tendency than the conditional mean model when the data is heteroscedastic. Quantile regression has been extended to handle censored data in survival analysis, providing a useful alternative to traditional Cox proportional hazards models since it relaxes the proportionality constraint and is robust when heterogeneity is present in the data.

Early methods (Ying et al. 1995, Yang 1999, Honore et al. 2002) required that the censoring time must be independent of covariates which is too restrictive for many real data applications. Conditional independence is a weaker condition assuming that failure time and censoring time are independent conditional on covariates. Combining conditional independence and without assuming constraints on an error distribution, Portnoy (2003) proposed a recursively reweighted estimator. Unfortunately, the quantile can not be computed until all lower quantile regression estimators are computed first. The recursive scheme also complicates asymptotic inference. To overcome inferential difficulties, Peng and Huang (2008) developed a quantile regression method for survival data subject to conditionally independent censoring and used a martingale-based procedure which makes asymptotic inference more tractable. However, it still has the same drawback as Portnoy (2003), namely, the entire lower quantile regression process must be computed first. To avoid the need to assume that all lower quantiles are linear, Wang and Wang (2009) proposed a locally weighted method. Unfortunately, their method suffers from the curse of dimensionality and hence can only handle a very limited number of covariates. Wey et al. (2014) and Leng and Tong (2013) proposed some alternatives to Wang and Wang (2009), which allow conditional independence assumption, yet they still suffer from the curse of dimensionality as in Wang and Wang (2009).

Interval-censored data arise extensively in epidemiological studies and clinical trials, especially in large-scale longitudinal studies where the event of interest, such as disease onset, can not be observed precisely, and is only known to occur between two examination times. Despite the fact that the development for censored quantile regression flourishes, the aforementioned methods were developed for right-censoring; hence, are not suitable for interval-censoring. To the best of our knowledge, the only method available for quantile regression models on interval-censored data was proposed by Kim et al. (2010) which is a generalization of the method developed by McKeague et al. (2001). The proposed method can only be applied when the covariates take on a finite number of values since the method requires estimation of the survival function conditional on covariates. The proposed method performed well in simulation studies, yet no theoretical justifications were offered. In this paper, we develop a new method for the conditional quantile regression model for Case II interval-censored data while allowing the censoring time to depend on the covariates without requiring the global linearity assumption.

The remaining paper is organized as follows. In Section 4.2, the proposed model is introduced and we establish estimation and inference procedures. Consistency and asymptotic distributions are presented in Section 4.3 with technical details deferred to Section 4.7. In Section 4.4 and Section 4.5, small-sample performance is demonstrated via simulation studies and an application to data from the Atherosclerosis Risk in Communities Study is given. Section 4.6 provides further discussion of the method presented herein.

4.2 Model and Inference Procedure

4.2.1 Models and Data

Let T denote failure time and let X denote a $k \times 1$ covariate vector with the first component set as one. We consider a quantile regression model for the failure time,

$$Q_T(\tau|X) = X^T\beta(\tau), \qquad \tau \in (0,1), \tag{4.20}$$

where $Q_T(\tau|X)$ is the conditional quantile defined as $Q_T(\tau|X) = \inf\{t : \mathrm{pr}(T \leq t|X) \geq \tau\}$ and the vector of unknown regression coefficients, $\beta(\tau)$, represents the covariate effects on the τth quantile of T which may depend on τ. Our interest lies in the estimation and inference on $\beta(\tau)$.

Let (L, R) be two random observation times satisfying $L < R$ with probability 1. Define $\delta_1 \equiv I(T \leq L)$ and $\delta_2 \equiv I(L < T \leq R)$ where $I(\cdot)$ is the indicator function. It is assumed that T is conditionally independent of L and R given $\boldsymbol{X}$. For Case II interval-censored data, T is not observed and we instead observe n independent replicates of $(L, R, X, \delta_1, \delta_2)$ denoted by $\{(L_i, R_i, \mathbf{X_i}, \delta_{i1}, \delta_{i2})_{i=1,2,\cdots,n}\}$.

The τth conditional quantile of T conditional on X can be characterized as the solution to the expected loss minimization problem (Powell 1994),

$$Z(\beta) = E\{\rho_\tau(T - X^T\beta(\tau))|X\}, \tag{4.21}$$

where $\rho_\tau(u) = u\{\tau - I(u < 0)\}$. Furthermore, quantiles possess "equivariance to monotone transformations" (Powell 1994) which means that we may analyze a transformation $h(T)$ since the conditional quantile of $h(T)$ is $h(X^T\beta(\tau))$ if $h(\cdot)$ is nondecreasing (Powell 1994). Consider the transformations $h_1(T|L,R) = I(T > L)$ and $h_2(T|L,R) = I(T > R)$ which are monotone nondecreasing, we can then apply the same transformation to the conditional quantile, $X^T\beta(\tau)$ and use the transformed conditional quantile, $I(X^T\beta(\tau) > L)$ and $I(X^T\beta(\tau) > R)$, in the subsequent analysis. The objective function in Equation (4.21) is well-defined irrespective of the existence of moments of the data and is sufficient to identify the parameters of interests (Powell 1994). We can thus substitute $(1-\delta_1)$, $I(X^T\beta > L)$, $(1-\delta_1-\delta_2)$, and $I(X^T\beta > R)$ in Equation (4.21) and get

$$Z^1(\beta) = E[\rho_\tau\{(1-\delta_1) - I(X^T\beta(\tau) > L)\}|X, L, R] \quad \text{and} \tag{4.22}$$

$$Z^2(\beta) = E[\rho_\tau\{(1-\delta_1-\delta_2) - I(X^T\beta(\tau) > R)\}|X, L, R]. \tag{4.23}$$

We can then estimate $\beta(\tau)$ using Equations (4.22) and (4.23) simultaneously since they contain only the observable variables $(L, R, X, \delta_1, \delta_2)$. Due to censoring, it is possible that not all $\beta(\tau)$ can be estimated using the observed data. We provide a sufficient condition to guarantee the identifiability for a fixed quantile in Section 4.3.1.

4.2.2 Parameter Estimation and Algorithm

We will suppress τ in $\beta(\tau)$ for notational simplicity. Assuming the formulation from Equation (4.22), the regression quantile estimator $\hat{\beta}_n$ (Koenker and Bassett 1978) is the minimizer of the objective function

$$\begin{aligned}
Z_n^1(\beta) &= \sum_{i=1}^n \rho_\tau\{(1-\delta_{i1}) - I(X_i^T\beta > L_i)\} \\
&= \sum_{i=1}^n \left\{\tau I(\delta_{i1}=0)I(X_i^T\beta - L_i \le 0) + (1-\tau)I(\delta_{i1}=1)I(X_i^T\beta - L_i > 0)\right\} \\
&= \sum_{i=1}^n w_{i1}\, I\{y_{i1}(X_i^T\beta - L_i) < 0\},
\end{aligned} \tag{4.24}$$

where

$$y_{i1} = \begin{cases} 1 & \text{if} \quad \delta_{i1} = 0 \\ -1 & \text{if} \quad \delta_{i1} = 1 \end{cases} \quad \text{and} \quad w_{i1} = \begin{cases} \tau & \text{if} \quad \delta_{i1} = 0 \\ 1-\tau & \text{if} \quad \delta_{i1} = 1 \end{cases}.$$

Similarly, assuming the formulation from Equation (4.23), the regression quantile estimator $\hat{\beta}_n$ is also the minimizer of the objective function

$$Z_n^2(\beta) = \sum_{i=1}^{n} \rho_\tau\{(1-\delta_{i1}-\delta_{i2}) - I(X_i^T\beta > R_i)\} = \sum_{i=1}^{n} w_{i2}\, I\{y_{i2}(X_i^T\beta - R_i) < 0\}, \quad (4.25)$$

where

$$y_{i2} = \begin{cases} 1 & \text{if} \quad \delta_{i1} = 0 \text{ and } \delta_{i2} = 0 \\ -1 & \text{if} \quad \delta_{i1} = 1 \text{ or } \delta_{i2} = 1 \end{cases} \quad \text{and} \quad w_{i2} = \begin{cases} \tau & \text{if} \quad \delta_{i1} = 0 \text{ and } \delta_{i2} = 0 \\ 1-\tau & \text{if} \quad \delta_{i1} = 1 \text{ or } \delta_{i2} = 1 \end{cases}.$$

To produce an estimate $\hat{\beta}_n$ using information from both observation times, we combine Equation (4.24) and (4.25) to from the objective function,

$$Z_n(\beta) = \sum_{i=1}^{n} w_{i1}\, I\{y_{i1}(X_i^T\beta - L_i) < 0\} + \sum_{i=1}^{n} w_{i2}\, I\{y_{i2}(X_i^T\beta - R_i) < 0\}. \quad (4.26)$$

The regression quantile estimator, $\hat{\beta}_n$, which minimizes Equation (4.26) is difficult to obtain by direct minimization since $Z_n(\beta)$ is neither convex nor continuous. To overcome this difficulty, we approximate $Z_n(\beta)$ as a difference of two hinge functions, where the approximation is controlled by a small constant (see Figure 4.12). Specifically, the approximation

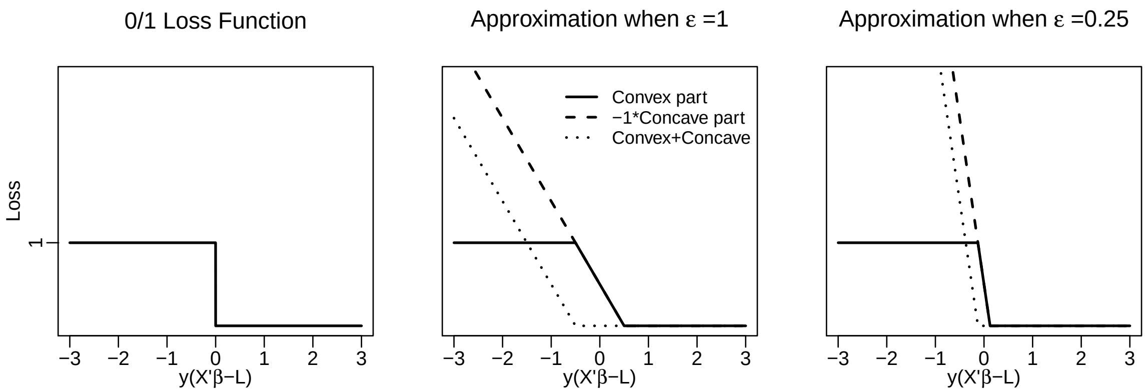

Figure 4.12: An illustration of using the difference between two hinge loss functions to approximate a 0/1 loss.

is defined as the following.

$$
\begin{aligned}
Z_{n,\epsilon}(\beta) &= \sum_{i=1}^{n} w_{i1}\left[\frac{1}{\epsilon}\left\{\frac{\epsilon}{2} - y_{i1}(X_i^T\beta - L_i)\right\}_+ - \frac{1}{\epsilon}\left\{-\frac{\epsilon}{2} - y_{i1}(X_i^T\beta - L_i)\right\}_+\right] \\
&\quad + \sum_{i=1}^{n} w_{i2}\left[\frac{1}{\epsilon}\left\{\frac{\epsilon}{2} - y_{i2}(X_i^T\beta - R_i)\right\}_+ - \frac{1}{\epsilon}\left\{-\frac{\epsilon}{2} - y_{i2}(X_i^T\beta - R_i)\right\}_+\right] \\
&= \sum_{i=1}^{n}\left[w_{i1}\left\{\frac{1}{2} - \frac{1}{\epsilon}y_{i1}(X_i^T\beta - L_i)\right\}_+ + w_{i2}\left\{\frac{1}{2} - \frac{1}{\epsilon}y_{i2}(X_i^T\beta - R_i)\right\}_+\right] \\
&\quad + \sum_{i=1}^{n}\left[(-w_{i1})\left\{-\frac{1}{2} - \frac{1}{\epsilon}y_{i1}(X_i^T\beta - L_i)\right\}_+ + (-w_{i2})\left\{-\frac{1}{2} - \frac{1}{\epsilon}y_{i2}(X_i^T\beta - R_i)\right\}_+\right]
\end{aligned}
\tag{4.27}
$$

where $\epsilon > 0$ and $\{x\}_+ = \max(x, 0)$.

To mitigate computational difficulties, we utilize the concave-convex procedure proposed by Yuille and Rangarajan (2003). The concave-convex procedure relies on decomposing an objective function, $f(x)$, into a convex part, $f_{convex}(x)$, and a concave part, $f_{concave}(x)$ such that $f(x) = f_{convex}(x) + f_{concave}(x)$. Optimization is carried out with an iterative procedure in which $f_{concave}(x)$ is linearized at the current solution $x^{(t)}$,

$x^{(t+1)} = \arg\min_x \left\{f_{convex}(x) + (x - x^{(t)})f'_{concave}(x^{(t)})\right\}$, making each iteration a convex optimization problem. The first value $x^{(0)}$ can be initialized with any reasonable guess.

To apply concave-convex procedure to our optimization problem, we define the first term in Equation (4.27) as $f_{convex}(\beta)$ and the second term as $f_{concave}(\beta)$. The gradient of the concave part, $f_{concave}(\beta)$, is

$$
\begin{aligned}
\frac{\partial}{\partial \beta(\tau)} f_{concave}(\beta) = & \sum_{i=1}^{n} w_{i1} \left(\frac{1}{\epsilon} y_{i1} X_i^T \right) \cdot I \left\{ \frac{1}{2} + \frac{1}{\epsilon} y_{i1} (X_i^T \beta - L_i) < 0 \right\} \\
& + \sum_{i=1}^{n} w_{i2} \left(\frac{1}{\epsilon} y_{i2} X_i^T \right) \cdot I \left\{ \frac{1}{2} + \frac{1}{\epsilon} y_{i2} (X_i^T \beta - R_i) < 0 \right\}
\end{aligned}
$$

Applying the concave-convex procedure to the above decomposition, we obtain

$$
\begin{aligned}
\beta^{(r+1)} = \arg\min_{\beta} \Bigg\{ & \sum_{i=1}^{n} \left[w_{i1} \left\{ \frac{1}{2} - \frac{1}{\epsilon} y_{i1} (X_i^T \beta - L_i) \right\}_+ + w_{i2} \left\{ \frac{1}{2} - \frac{1}{\epsilon} y_{i2} (X_i^T \beta - R_i) \right\}_+ \right] \\
& + \sum_{i=1}^{n} w_{i1} \left\{ \frac{1}{\epsilon} y_{i1} X_i^T (\beta - \beta^{(r)}) \right\} \cdot I \left\{ \frac{1}{2} + \frac{1}{\epsilon} y_{i1} (X_i^T \beta - L_i) < 0 \right\} \\
& + \sum_{i=1}^{n} w_{i2} \left\{ \frac{1}{\epsilon} y_{i2} X_i^T (\beta - \beta^{(r)}) \right\} \cdot I \left\{ \frac{1}{2} + \frac{1}{\epsilon} y_{i2} (X_i^T \beta - R_i) < 0 \right\} \Bigg\}, \qquad (4.28)
\end{aligned}
$$

where $\beta^{(r)}$ denotes the estimated $\beta(\tau)$ at the rth iteration.

The final form can be solved with a standard convex optimization algorithm with a decreasing sequence of $\epsilon = \{2^0, 2^{-1}, \cdots\}$. Specifically, given the initial value, we solved Equation (4.28) with $\epsilon = 2^0$ using the optimization toolbox in MATLAB. The solution with $\epsilon = 2^0$ was then used as the initial value to solve Equation (4.28) with $\epsilon = 2^{-1}$. This was repeated until the maximum relative change over all covariates was less than one percent.

Since the continuous approximation to the object function is not convex, a good initial value is necessary to circumvent the multitude of local minima. Observe that the objective function in Equation (4.26) bears a resemblance to that of of the support vector machine (SVM), i.e. the loss is added when $X_i^T \beta$ is on the "wrong side" of the observation times L_i or R_i. We thus use a weighted SVM to produce the initial value since it is a quadratic programming problem which is easy to compute. We first stack the observations such that

$\tilde{\boldsymbol{y}} = [y_{11}, y_{21}, \cdots y_{n1}, y_{12}, y_{22}, \cdots, y_{n2}]'$, $\tilde{\boldsymbol{C}} = [L_1, L_2, \cdots, L_n, R_1, R_2, \cdots, R_n]'$,

$\tilde{\boldsymbol{w}} = [w_{11}, w_{21}, \cdots w_{n1}, w_{12}, w_{22}, \cdots, w_{n2}]'$, and $\tilde{X}_i = \tilde{X}_{n+i} = X_i$ for $i = 1, \cdots, n$. We can then formulate a weighted-SVM as the solution to

$$\min_{\beta} \frac{1}{2}\|\beta\|^2 + d\sum_{j=1}^{2n} \tilde{w}_j \xi_j \qquad \text{subject to} \begin{cases} \tilde{y}_j(\tilde{X}_j'\beta - \tilde{C}_j) \geq 1 - \xi_j & \forall j \\ \xi_j \geq 0 \end{cases}, \tag{4.29}$$

where d is the "cost" parameter and can be selected using cross-validation. Equation (4.29) can be solved using a standard quadratic programming solver. The solution to Equation (4.29) can be used as the initial value in our proposed algorithm. One should note that, in our formulation, the intercept term is included in the $\|\beta\|^2$ term which differs from the typical SVM. The initial value used in both simulation studies and real data analysis was produced using this method.

4.2.3 Inference

The confidence intervals for parameter estimates are obtained using a subsampling method since bootstrap does not consistently estimate the asymptotic distribution for estimators with cube-root convergence (Abrevaya and Huang 2005). The subsampling method described below is from Politis et al. (1999). Subsampling can produce consistent estimated sampling distributions under extremely weak assumptions even when the bootstrap fails. The justification for using the subsampling method in our study is discussed further in Section 4.3.3.

To obtain the confidence intervals for the minimizer of (4.27), $\hat{\beta}_{n,\epsilon}$, we produce subsamples $K_1, K_2, \cdots, K_{N_n}$ where K_j's are the $N_n \equiv \binom{n}{b}$ distinct subsets of $\{(L_i, R_i, X_i, \delta_{i1}, \delta_{i2})_{i=1,\cdots,n}\}$ of size b. Let β_τ denote the true parameter values and $\hat{\beta}_{n,\epsilon,b,j}$ denote the estimated value produced by solving (4.28) using the K_jth dataset.

Define

$$L_{n,b}(x) = N_n^{-1} \sum_{j=1}^{N_n} I\{b^{1/3}(\hat{\beta}_{n,\epsilon,b,j} - \hat{\beta}_{n,\epsilon}) \leq x\} \quad \text{and} \quad c_{n,b}(\gamma) = \inf\{x : L_{n,b}(x) \geq \gamma\}.$$

From Theorem 3.3.3 given later, for any $0 < \gamma < 1$, $P\left(n^{1/3}(\hat{\beta}_{n,\epsilon} - \beta_\tau) \leq c_{n,b}(\gamma)\right) \to \gamma$ under the condition that $b \to \infty$ as $n \to \infty$ and $b/n \to 0$. It follows that for any $0 < \alpha < 0.5$, $P\left(c_{n,b}\left(\frac{\alpha}{2}\right) < n^{1/3}(\hat{\beta}_{n,\epsilon} - \beta_\tau) \leq c_{n,b}\left(1 - \frac{\alpha}{2}\right)\right) \to 1 - \alpha$ thus an asymptotic $1 - \alpha$ level confidence interval for β_τ can be constructed with

$$\left[\hat{\beta}_{n,\epsilon} - n^{-1/3} c_{n,b}\left(1 - \frac{\alpha}{2}\right), \hat{\beta}_{n,\epsilon} - n^{-1/3} c_{n,b}\left(\frac{\alpha}{2}\right)\right].$$

Symmetric confidence intervals can be obtained by modifying the above approach slightly. Define

$$\tilde{L}_{n,b}(x) = N_n^{-1} \sum_{i=j}^{N_n} I\{b^{1/3}\,|\hat{\beta}_{n,\epsilon,b,j} - \hat{\beta}| \leq x\} \quad \text{and} \quad \tilde{c}_{n,b}(\gamma) = \inf\{x : \tilde{L}_{n,b}(x) \geq \gamma\}.$$

Again, if $b \to \infty$ as $n \to \infty$ and $b/n \to 0$, a symmetric confidence interval for $\hat{\beta}$ can be constructed as

$$\left[\hat{\beta}_{n,\epsilon} - n^{-1/3}\tilde{c}_{n,b}(1-\alpha), \, \hat{\beta}_{n,\epsilon} + n^{-1/3}\tilde{c}_{n,b}(1-\alpha)\right]. \tag{4.30}$$

Symmetric confidence intervals are desirable because they often have nicer properties than the nonsymmetric version in finite samples (Banerjee and Wellner 2005). This fact was also observed in our simulation studies; hence, symmetric confidence intervals are recommended and used in this paper.

To avoid large scale computation issues, a stochastic approximation from Politis et al. (1999) is employed where B randomly chosen datasets from $\{1, 2, \cdots, N_n\}$ are used in the above calculation. Furthermore, the block size is chosen using the method implemented in Delgado et al. (2001).

4.3 Asymptotic Properties

4.3.1 Identifiability

We first provide a set of sufficient conditions for identifiability.

For a fixed quantile τ, let $Z_n(\beta)$ be the summation of Equation (4.24) and Equation (4.25), that is,

$$\begin{aligned}
Z_n(\beta) = \frac{1}{n}\sum_{i=1}^{n}[&\tau I(T_i > L_i)I(X_i^T\beta - L_i \le 0) + (1-\tau)I(T_i \le L_i)I(X_i^T\beta - L_i > 0) \\
&+ \tau I(T_i > R_i)I(X_i^T\beta - R_i \le 0) + (1-\tau)I(T_i \le R_i)I(X_i^T\beta - R_i > 0)] \\
= \frac{1}{n}\sum_{i=1}^{n}[&\tau\{I(T_i > L_i)I(X_i^T\beta - L_i \le 0) + I(T_i > R_i)I(X_i^T\beta - R_i \le 0)\} \\
&+ (1-\tau)\{I(T_i \le L_i)I(X_i^T\beta - L_i > 0) + I(T_i \le R_i)I(X_i^T\beta - R_i > 0)\}]
\end{aligned} \tag{4.31}$$

and

$$\begin{aligned}
Z(\beta) = E[&\tau\{I(T > L)I(X^T\beta - L \le 0) + I(T > R)I(X^T\beta - R \le 0)\} \\
&+ (1-\tau)\{I(T \le L)I(X^T\beta - L > 0) + I(T \le R)I(X^T\beta - R > 0)\}].
\end{aligned} \tag{4.32}$$

Let β_τ denote a minimizer of $Z(\cdot)$. We assume the following conditions.

Condition 9. *The support of f_X is not contained in any proper linear subspace of $\Re^k$.*

Condition 10. *For a fixed τ, with probability one, the support of the conditional densities of L and R given X, $f_{L|X}$ and $f_{R|X}$, and the support of the conditional density of T given X, $f_{T|X}$, contain $X^T\beta_\tau$ in their interiors.*

Lemma 4.3.1. *Under Conditions 9 and 10, β_τ is the unique minimizer of $Z(\beta)$.*

We prove Lemma 4.3.1 by showing $Z(\beta) - Z(\beta_\tau) > 0$, for any $\beta \ne \beta_\tau$. A detailed proof is provided in Section 4.7.

4.3.2 Consistency

Let $\hat{\beta}_{n,\epsilon}$ be the minimizer of $Z_{n,\epsilon}(\cdot)$ in $\mathcal{B}$ and for a fixed quantile τ, let

$$
\begin{aligned}
&Z_{n,\epsilon}(\beta) = \\
&\quad \sum_{i=1}^{n} \Bigg(\tau I(T_i > L_i) \left\{ I(X_i^T\beta - L_i \leq -\frac{\epsilon}{2}) + I(|X_i^T\beta - L_i| < \frac{\epsilon}{2}) \left[-\frac{1}{\epsilon}(X_i^T\beta - L_i - \frac{\epsilon}{2}) \right] \right\} \\
&\quad + (1-\tau) I(T_i \leq L_i) \left\{ I(X^T\beta - L_i > \frac{\epsilon}{2}) + I(|X_i^T\beta - L_i| \leq \frac{\epsilon}{2}) \left[\frac{1}{\epsilon}(X_i^T\beta - L_i + \frac{\epsilon}{2}) \right] \right\} \\
&\quad + \tau I(T_i > R_i) \left\{ I(X_i^T\beta - R_i \leq -\frac{\epsilon}{2}) + I(|X_i^T\beta - R_i| < \frac{\epsilon}{2}) \left[-\frac{1}{\epsilon}(X_i^T\beta - R_i - \frac{\epsilon}{2}) \right] \right\} \\
&\quad + (1-\tau) I(T_i \leq R_i) \left\{ I(X^T\beta - R_i > \frac{\epsilon}{2}) + I(|X_i^T\beta - R_i| \leq \frac{\epsilon}{2}) \left[\frac{1}{\epsilon}(X_i^T\beta - R_i + \frac{\epsilon}{2}) \right] \right\} \Bigg)
\end{aligned}
$$

We assume the following conditions for the consistency theorem.

Condition 11. *Let $\beta \in \mathcal{B}$ where $\mathcal{B}$ is a compact subset of $\Re^k$ which contains β_τ as an interior point.*

Condition 12. *$M_T \equiv \sup_{T,X} f_{t|X}(T|X) < \infty$, $M_L \equiv \sup_{L,X} f_{L|X}(L|X) < \infty$, and $M_R \equiv \sup_{R,X} f_{R|X}(R|X) < \infty$.*

Theorem 4.3.2. *Under Conditions 9–12, $\hat{\beta}_{n,\epsilon}$ converges to β_τ in probability as $n \to \infty$ and $\epsilon \to 0$.*

The proof follows by first showing that the collection of functions in $Z_n(\beta)$ is a VC-subgraph class and hence $Z_n(\beta)$ converges almost surely uniformly to $Z(\beta)$. In addition, $|Z_{n,\epsilon}(\beta) - Z_n(\beta)| \to 0$ as $n \to \infty$ and $\epsilon \to 0$; thus we can conclude that $Z_{n,\epsilon}$ converges almost surely uniformly to $Z(\beta)$. Next, we prove that $Z(\cdot)$ is continuous. Condition 9 and 10 provided sufficient conditions for identifiability and hence, β_τ is the unique minimizer of $Z(\cdot)$. Since we assumed $\mathcal{B}$ is compact, we can then conclude that $\hat{\beta}_{n,\epsilon}$ converges to β_τ in probability by a standard argument for M-estimators (see Theorem 2.1 of Newey and McFadden (1994)). A detailed proof is provided in Section 4.7.

4.3.3 Asymptotic Distribution

In this section, we show that $n^{1/3}(\hat{\beta}_{n,\epsilon} - \beta_\tau)$ converges to a nondegenerate distribution. The convergence rate is atypical because our objective function (4.27) is non-smooth and not everywhere differentiable; this is sometimes called the "sharp-edge effect" (Kim and Pollard 1990). We will make the following assumptions which guarantee the asymptotic distribution will be nondegenerate.

Condition 13. $\epsilon = o(n^{-2/3})$.

Condition 14. *The true distribution P of L, R, T and X is absolutely continuous with respect to Lebesgue measure.*

Condition 15. *X is bounded.*

Condition 16. *Let $V(\beta_\tau)_{i,j} = P_x\left[X_i X_j\{f_{L|X}(X^T\beta_\tau|X) + f_{R|X}(X^T\beta_\tau|X)\}f_{T|X}(X^T\beta_\tau|X)\right]$ and $V(\beta_\tau)$ is positive definite where X_i and X_j are elements of X.*

We may now proceed with the main result.

Theorem 4.3.3. *Under Conditions 9–16, the process $\left\{n^{2/3}\left[Z_n\left(\beta_\tau + sn^{-1/3}\right) - Z_n\left(\beta_\tau\right)\right] : s \in \Re^k\right\}$ converges in distribution to a Gaussian process $\left\{\Gamma(s) : s \in \Re^k\right\}$ with continuous sample paths, mean $s^T V(\beta_\tau)s/2$, and covariance H, where V is the Hessian matrix of $Z(\beta)$ at β_τ, and H is defined in Equation (4.37) of Section 4.7. Furthermore, $n^{1/3}(\hat{\beta}_{n,\epsilon} - \beta_\tau) \rightarrow_d \arg\min \Gamma(s)$.*

Theorem 2 follows by verifying the conditions of the main theorem from Kim and Pollard (1990). Provided that V is positive definite, we can conclude that $n^{1/3}(\hat{\beta}_{n,\epsilon} - \beta_\tau)$ converges to a nondegenerate distribution. A detailed proof is provided Section 4.7.

Subsampling can produce consistent estimated sampling distributions for our estimator as an immediate consequence of Theorem 2.2.1 from Politis et al. (1999). In our study, we choose block size $b = N^\gamma$ where $\gamma = \{1/3, 1/2, 2/3, 3/4, 0.8, 5/6, 6/7, 0.9, 0.95\}$ thus

$b \to \infty$ and $b/N \to 0$ as $N \to \infty$. $n^{1/3}(\hat{\beta}_{n,\epsilon} - \beta_\tau)$ converges to a nondegenerate continuous distribution. All conditions in Theorem 2.2.1 from Politis et al. (1999) are met thus we can construct confidence intervals as detailed in Section 4.2.3.

4.4 Simulation Studies

Two simulation studies were carried out to examine the finite sample performance of our estimator. In the first scenario, *Simulation 1*, the conditional quantile functions had identical linear coefficient and differed only intercept. In the second scenario, *Simulation 2*, both the intercept and covariate effects varied over the quantiles. *Simulation 1* represents a situation where the errors are independent and identically distributed and *Simulation 2* represents a situation where the errors were heteroscedastic.

In *Simulation 1*, the covariate is $X \equiv (1, X_1, X_2)^T$ where $X_1 \sim$ Uniform $[0, 2]$ and $X_2 \sim$ Bernoulli(0.5). The unobserved failure times were generated from the linear model, $T = 2 + 3\,X_1 + X_2 + 0.3\,U$. The observation times, L and R, were generated from the linear model, $1.9 + 3.2\,X_1 + 0.8\,V$ when $X_2 = 0$ and $3.1 + 2.8\,X_1 + 0.8\,V$ when $X_2 = 1$. Both U and V were generated from $N(0, 1)$. Twenty-eight percent of events occurred prior to the left observation time and 44% of events occurred between the left and right observation times. In *Simulation 2*, the covariate setup is the same as in *Simulation 1*. Unobserved failure times were generated from the linear model, $T = 2 + 3\,X_1 + X_2 + (0.3 + 0.2\,X_1)\,U$. The observation times, L and R, were generated from the linear model, $1.9 + 3.2\,X_1 + 0.8\,V$. The proportion of events occurring prior to left observation time and in between the left and right observation times was about 34% and 40%, respectively.

For each scenario, we report the mean bias, mean squared error, and median absolute deviation based on 1000 simulations. Sample sizes were chosen to be $n = 200, 400,$ and 800 for each simulation setup. We are interested in estimation of 25th, median, and 75th percentiles. For each simulated dataset, the procedure described at the end of Section 4.2.2

was used to estimate β. Symmetric confidence intervals were calculated based on a stochastic approximation with 500 subsamples. To decreases the computational burden, the block size was determined via a pilot simulation in the same fashion as described in Banerjee and McKeague (2007). The optimal subsampling block size was determined from the following selected block sizes: $\{n^{1/3}, n^{1/2}, n^{2/3}, n^{3/4}, n^{0.8}, n^{5/6}, n^{6/7}, n^{0.9}, n^{0.95}\}$.

Table 4.5 and Table 4.6 summarize the results for *Simulation 1* and *Simulation 2* with sample size equal to 200, 400, and 800 at the 25th, median, and 75th percentiles. In the tables, "Truth" is the true parameter value; "Bias" is the mean bias of the estimates from all replicates; "MSE" is the mean squared error; "MAD" is the median absolute deviation of the estimates; "CP" is the average coverage from subsampling symmetric confidence intervals; "Length" is the average confidence interval length.

The tables show that the regression coefficient estimators have negligible bias. In *Simulation 1*, the bias has a decreasing trend as the sample size increases for all quantiles and parameters. The mean squared errors and median absolute deviations decrease as the sample size increases for all quantiles and parameters. The subsampling confidence interval coverage fluctuates around the nominal level 95% with slight over coverage for some instances. In *Simulation 2*, the bias for all quantiles is small for all sample sizes. There is a general decreasing trend for bias when the sample size increases. The mean squared errors and median absolute deviations decrease as the sample size increases for all quantiles and parameters. The average 95% confidence interval coverage rate is a bit high for the smallest sample size but gets closer to the nominal 0.95 level as the sample size increases.

4.5 Application

We now apply the proposed method to analyze the Atherosclerosis Risk in Communities (ARIC) study data. The detailed study design and objective are described elsewhere (The ARIC Investigators 1989). ARIC enrolled 15,792 participants aged 45 to 64

Table 4.5: Simulation results for *Simulation 1*, based on 1000 simulation replicates.

N	τ	Parameter	Truth	Bias	MSE	MAD	CP	Length
200	0.25	β_0	1.7977	0.0124	0.0222	0.1024	0.9500	0.5561
		β_1	3.0000	-0.0041	0.0137	0.0776	0.9670	0.4441
		β_2	1.0000	0.0136	0.0178	0.0898	0.9620	0.5056
	0.50	β_0	2.0000	0.0024	0.0187	0.0915	0.9610	0.5188
		β_1	3.0000	-0.0018	0.0115	0.0707	0.9560	0.4025
		β_2	1.0000	0.0067	0.0156	0.0854	0.9650	0.4523
	0.75	β_0	2.2023	0.0016	0.0252	0.1065	0.9400	0.5621
		β_1	3.0000	-0.0083	0.0149	0.0774	0.9470	0.4379
		β_2	1.0000	-0.0015	0.0185	0.0856	0.9590	0.4924
400	0.25	β_0	1.7977	0.0054	0.0130	0.0761	0.9570	0.4383
		β_1	3.0000	-0.0016	0.0076	0.0586	0.9660	0.3364
		β_2	1.0000	0.0060	0.0112	0.0686	0.9620	0.3939
	0.50	β_0	2.0000	0.0048	0.0096	0.0652	0.9650	0.3891
		β_1	3.0000	-0.0027	0.0059	0.0492	0.9750	0.3055
		β_2	1.0000	-0.0002	0.0093	0.0645	0.9610	0.3495
	0.75	β_0	2.2023	-0.0013	0.0134	0.0784	0.9650	0.4290
		β_1	3.0000	0.0005	0.0080	0.0611	0.9620	0.3329
		β_2	1.0000	0.0010	0.0104	0.0656	0.9520	0.3855
800	0.25	β_0	1.7977	0.0028	0.0074	0.0550	0.9670	0.3272
		β_1	3.0000	-0.0018	0.0046	0.0453	0.9660	0.2560
		β_2	1.0000	0.0005	0.0060	0.0526	0.9630	0.3011
	0.50	β_0	2.0000	-0.0038	0.0063	0.0494	0.9670	0.3009
		β_1	3.0000	0.0003	0.0035	0.0386	0.9740	0.2286
		β_2	1.0000	0.0035	0.0049	0.0438	0.9550	0.2754
	0.75	β_0	2.2023	-0.0000	0.0071	0.0581	0.9620	0.3319
		β_1	3.0000	-0.0013	0.0043	0.0409	0.9650	0.2570
		β_2	1.0000	0.0022	0.0061	0.0494	0.9680	0.2996

Truth is the true parameter value; Bias is mean of bias from 1000 replicates; MSE is mean squared error; MAD is median absolute deviation of the estimates; CP is the empirical coverage probabilities with a nominal level of 0.95 from subsampling symmetric confidence intervals with 500 subsamples; Length is mean confidence interval length.

Table 4.6: Simulation results for *Simulation 2*, based on 1000 simulation replicates.

N	τ	Parameter	Truth	Bias	MSE	MAD	CP	Length
200	0.25	β_0	1.7977	0.0162	0.0359	0.1230	0.9560	0.7387
		β_1	2.8651	-0.0110	0.0317	0.1136	0.9720	0.6823
		β_2	1.0000	-0.0005	0.0358	0.1289	0.9590	0.7421
	0.50	β_0	2.0000	0.0065	0.0286	0.1153	0.9680	0.6508
		β_1	3.0000	-0.0179	0.0251	0.1075	0.9570	0.5744
		β_2	1.0000	0.0141	0.0295	0.1158	0.9530	0.6305
	0.75	β_0	2.2023	0.0033	0.0336	0.1251	0.9560	0.7263
		β_1	3.1349	-0.0163	0.0265	0.1051	0.9620	0.6353
		β_2	1.0000	0.0044	0.0341	0.1178	0.9630	0.7157
400	0.25	β_0	1.7977	-0.0041	0.0196	0.0919	0.9720	0.5572
		β_1	2.8651	-0.0028	0.0177	0.0848	0.9650	0.5178
		β_2	1.0000	0.0059	0.0213	0.0974	0.9620	0.5632
	0.50	β_0	2.0000	-0.0035	0.0164	0.0840	0.9580	0.4974
		β_1	3.0000	-0.0056	0.0133	0.0782	0.9600	0.4331
		β_2	1.0000	0.0025	0.0174	0.0859	0.9580	0.4914
	0.75	β_0	2.2023	0.0011	0.0195	0.0960	0.9670	0.5534
		β_1	3.1349	-0.0032	0.0156	0.0816	0.9640	0.4766
		β_2	1.0000	-0.0043	0.0206	0.0956	0.9670	0.5444
800	0.25	β_0	1.7977	-0.0023	0.0111	0.0707	0.9690	0.4158
		β_1	2.8651	-0.0054	0.0097	0.0630	0.9660	0.3838
		β_2	1.0000	0.0062	0.0132	0.0794	0.9650	0.4289
	0.50	β_0	2.0000	0.0008	0.0098	0.0649	0.9580	0.3748
		β_1	3.0000	-0.0038	0.0078	0.0563	0.9500	0.3335
		β_2	1.0000	-0.0007	0.0093	0.0634	0.9680	0.3797
	0.75	β_0	2.2023	0.0055	0.0127	0.0749	0.9600	0.4204
		β_1	3.1349	-0.0007	0.0096	0.0653	0.9570	0.3667
		β_2	1.0000	-0.0018	0.0137	0.0791	0.9560	0.4231

Truth is the true parameter value; Bias is mean of bias from 1000 replicates; MSE is mean squared error; MAD is median absolute deviation of the estimates; CP is the empirical coverage probabilities with a nominal level of 0.95 from subsampling symmetric confidence intervals with 500 subsamples; Length is mean confidence interval length.

years between 1987 and 1989 (visit 1) from 4 communities and followed them for a maximum of 5 visits. The first 4 visits were approximately 3 years apart and the last visit (visit 5) was done between 2011 and 2013 for a maximum follow-up of 24 years. We focus on the African American population with at least 2 visits (N=3861). We choose to focus on African Americans because the prevalence of hypertension in African Americans is among the highest in the world (AHA Writng Group 2010). We excluded those participants who had hypertension at visit 1 (N=2381) and participants with missing body mass index (BMI) (N=2). The final analysis included 1478 African American adults aged 45-64 (median=51) of which 60.4% (N=893) were female.

Hypertension was defined as diastolic blood pressure $\geq$ 90 mm Hg or systolic blood pressure (BP) $\geq$ 140 mm Hg or reporting the use of medication known to treat hypertension. Among 1478 adults in the analysis sample, 866 (58.6%) developed hypertension by the end of the study. The (unobserved) failure time of interest is the logarithm of number of months to hypertension development after visit 1. The analysis examines the effect of age, gender, BMI, and systolic BP on the quantiles of logarithm of time to hypertension. These covariates are chosen because they are strong predictors of incident hypertension (AHA Writng Group 2010).

Before applying our proposed method to the data, nonparametric maximum likelihood estimation (NPMLE) of the (unobserved) failure time distribution function was carried out for the data (Wellner and Zhan 1997). NPMLE was used to determine whether certain quantiles can be reasonably estimated. The NPMLE of distribution functions stratified by each covariate is shown in Figure 4.13. The continuous covariates were dichotomized by the median of each covariate.

Based on the NPMLE in Figure 4.13, it shows that we may have difficulty distinguish quantiles above 0.6. The long flat segment of NPMLE most likely due to the fact that visit

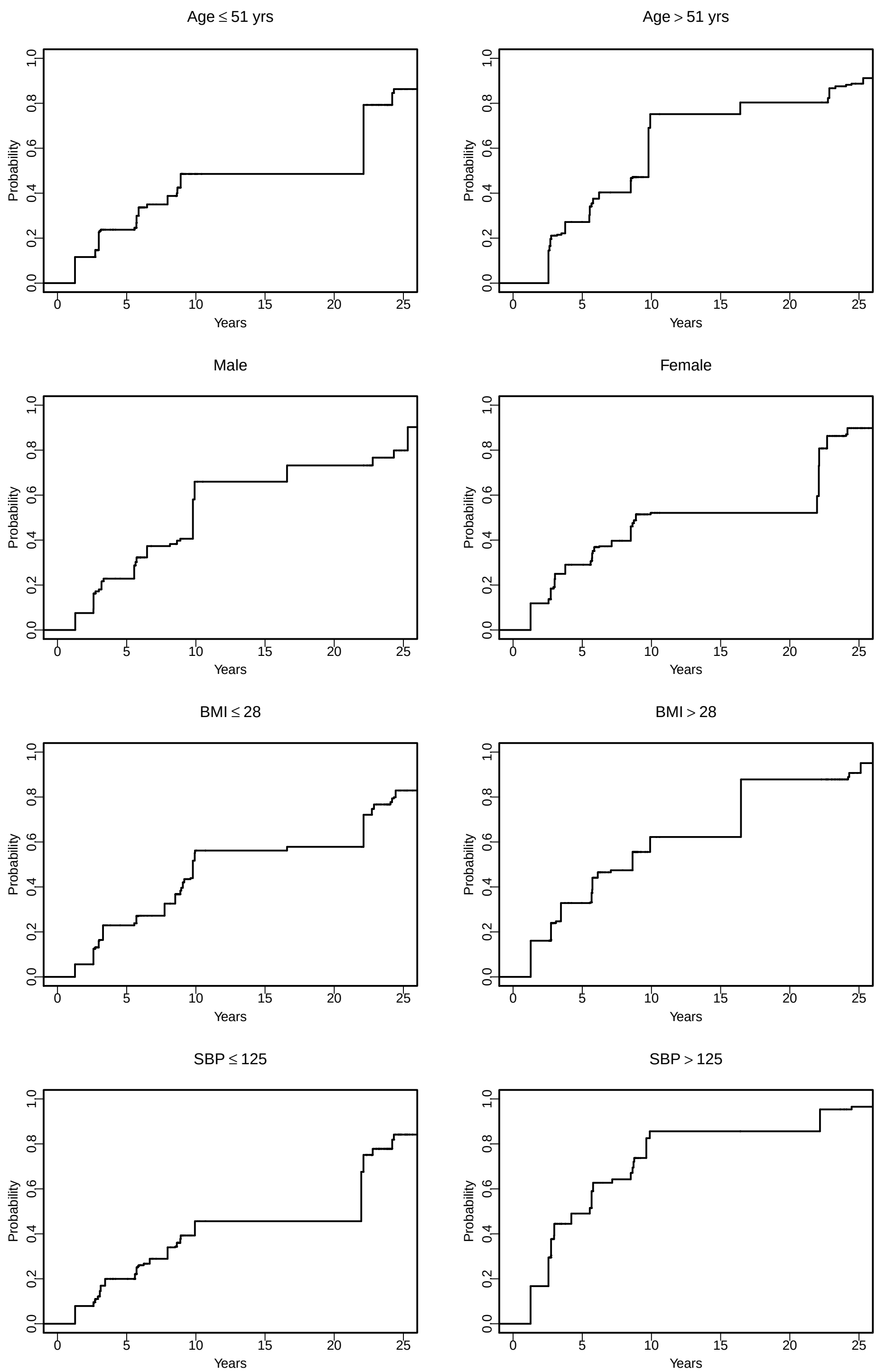

Figure 4.13: Nonparametric maximum likelihood estimator (NPMLE) of the (unobserved) failure time distribution function stratified by covariates.

Table 4.7: Results of analyzing “Atherosclerosis Risk in Communities (ARIC) study” data, effect on time to hypertension onset in years

Quantile	Parameter	Estimate	Lower C.I.	Upper C.I.
0.2	Intercept	2.875	2.115	3.636
	Age	0.013	-0.410	0.437
	Male	0.334	-0.562	1.230
	BMI	-0.258	-0.619	0.104
	Systolic BP	-1.521	-2.099	-0.943
0.3	Intercept	4.152	3.811	4.493
	Age	0.234	-0.094	0.562
	Male	0.658	0.023	1.292
	BMI	-0.651	-1.150	-0.152
	Systolic BP	-1.130	-1.665	-0.596
0.4	Intercept	5.156	4.754	5.559
	Age	0.194	-0.100	0.489
	Male	0.170	-0.671	1.011
	BMI	-0.212	-0.509	0.086
	Systolic BP	-0.462	-0.974	0.050
0.5	Intercept	5.298	5.061	5.535
	Age	0.074	-0.096	0.244
	Male	0.323	-0.220	0.867
	BMI	0.029	-0.142	0.201
	Systolic BP	-0.308	-0.523	-0.094

BMI: Body Mass Index; BP: Blood Pressure.

4 and visit 5 are almost 15 years apart. Since we only have enough information for estimation for quantiles between 0.2 to around 0.6, we focused on 0.2 to 0.5 quantiles with increments of 0.1. We centered the age at 50 then divided it by 5, centered the BMI at 28 then divided by 5, and centered systolic BP at 115 then divided by 10. Our proposed model was fitted for the lower quantiles and symmetric confidence intervals were constructed by subsampling where the block size was chosen based on the algorithm presented in Section 4.2.3. The results are summarized in Table 4.7 and Figure 4.14.

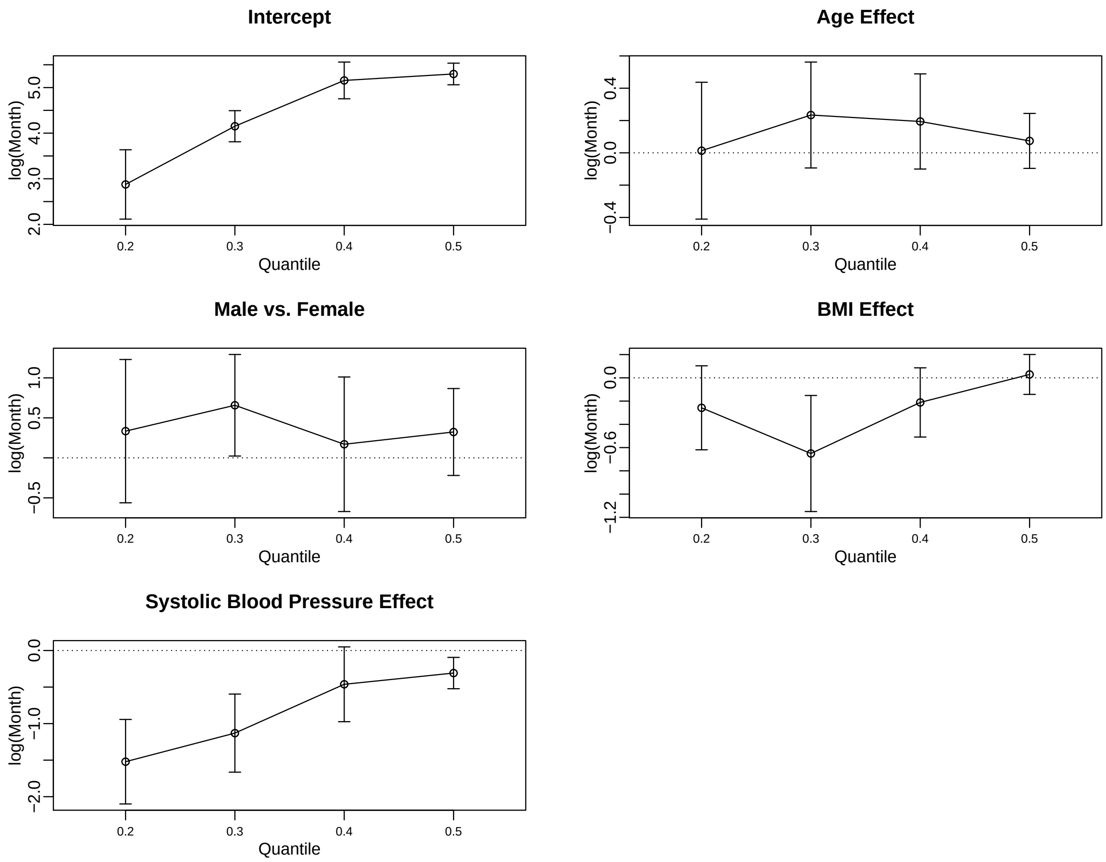

Figure 4.14: ARIC data: effect on time to hypertension onset. The vertical bars are symmetric confidence interval constructed using a subsampling method.

The results indicate that age has no statistically significant effect on log-time to hypertension onset at all quantiles examined. Gender has positive estimates at all quantiles examined but it is only statistically significant at 0.3 quantile. BMI has negative estimates at the lower quantiles examined but it is only significant at 0.3 quantile. Systolic BP is negatively associated with log-time to hypertension onset, i.e. higher systolic BP at visit 1 associate with shorter time to hypertension. Interestingly, the effect of systolic blood pressure appears to be the strongest at the lowest quantile and the effect extenuates at higher quantiles.

Each 5 unit increase of BMI at visit 1 shortens the time to hypertension by about 50% at 0.3 quantile. Each 10 mm Hg increase of systolic BP at visit 1 shorten the time to hypertension by about 80% at 0.2 quantile but the effect diminished to about 30% at 0.5 quantile. The systolic BP effect on shortening time to hypertension is large in the quantiles examined which may suggest that lowering the systolic BP can be most beneficial to delay hypertension onset.

Interestingly, the age is not significantly associated with time to hypertension onset which is different from the conclusion of most literature. One possible explanation is that older age is a significant predictor of greater awareness, treatment, and control of high blood pressure (Wyatt et al. 2008); therefore, older African American may be controlling their blood pressure by means other than medication and hence older age is not associated with shorter time to hypertension. The age effect on hypertension among African American warrents more investigation.

4.6 Discussion

To solve the non-convex objective function in (4.26), we used the difference between two convex hinge functions (4.27) to approximate the objective function. We also offer a practical solution to the issue of selecting good initial values. The computational procedure proposed is efficient and easy to implement.

Identifiability is a common issue for interval-censored data. We provided theoretical justification in the article and in practice, we recommend using nonparametric maximum likelihood estimator (NPMLE) of the (unobserved) failure time distribution function to explore whether certain quantiles are identifiable. This appeared to work well in our real data example.

The method proposed in this paper can be extended to handle multivariate interval-censored data and it can also be adapted to analyze right censored data. Nonparametric quantile regression models are also possible which will add greater flexibility in the functional form. Furthermore, models with varying coefficient or mixed censoring mechanism will be useful in practice which warrant future investigation.

4.7 Proofs of Lemma and Theorems

Proof of Lemma 4.3.1.

$$
\begin{aligned}
Z(\beta) - Z(\beta_\tau) = P_{L,R,X}\, [\, & \tau\{I(T > L)I(X^T\beta - L \le 0) + I(T > R)I(X^T\beta - R \le 0)\} \\
& + (1-\tau)\{I(T \le L)I(X^T\beta - L > 0) + I(T \le R)I(X^T\beta - R > 0)\}] \\
& - \tau\{I(T > L)I(X^T\beta_\tau - L \le 0) - I(T > R)I(X^T\beta_\tau - R \le 0)\} \\
& -(1-\tau)\{I(T \le L)I(X^T\beta_\tau - L > 0) - I(T \le R)I(X^T\beta_\tau - R > 0)\}] \\
= P_{L,R,X}\, & \{[\tau I(T > L) - (1-\tau)I(T \le L)] \\
& \big[I(X^T\beta \le L < X^T\beta_\tau) - I(X^T\beta_\tau \le L < X^T\beta)\big] \\
& [\tau I(T > R) - (1-\tau)I(T \le R)] \big[I(X^T\beta \le R < X^T\beta_\tau) - I(X^T\beta_\tau \le R < X^T\beta)\big]\big\} \\
= P_{L,X}\, & \big\{[\tau - I(T \le L)] \big[I(X^T\beta \le L < X^T\beta_\tau) - I(X^T\beta_\tau \le L < X^T\beta)\big]\big\} \\
+ P_{R,X}\, & \big\{[\tau - I(T \le R)] \big[I(X^T\beta \le R < X^T\beta_\tau) - I(X^T\beta_\tau \le R < X^T\beta)\big]\big\} \\
= P_{L,X}\, & \big\{\big[\tau - F_{T|X}(L|X)\big] \big[I(X^T\beta \le L < X^T\beta_\tau) - I(X^T\beta_\tau \le L < X^T\beta)\big]\big\} \\
+ P_{R,X}\, & \big\{\big[\tau - F_{T|X}(R|X)\big] \big[I(X^T\beta \le R < X^T\beta_\tau) - I(X^T\beta_\tau \le R < X^T\beta)\big]\big\} \\
= \int_X \int_{L|X} & \big[\tau - F_{T|X}(L|X)\big] \big[I(X^T\beta \le L < X^T\beta_\tau) - I(X^T\beta_\tau \le L < X^T\beta)\big]\, dP_{L|X}\, dP_X \\
+ \int_X \int_{R|X} & \big[\tau - F_{T|X}(R|X)\big] \big[I(X^T\beta \le R < X^T\beta_\tau) - I(X^T\beta_\tau \le R < X^T\beta)\big]\, dP_{R|X}\, dP_X \\
= \int_X \int_{(X^T\beta \le L < X^T\beta_\tau)\bigcup(X^T\beta_\tau \le L < X^T\beta)} & \left| \int_{X^T\beta_\tau}^{l} f_{T|X}(t)\, f_{L|X}(l)\, dt \right| dl\, dP_X \\
& + \int_X \int_{(X^T\beta \le R < X^T\beta_\tau)\bigcup(X^T\beta_\tau \le R < X^T\beta)} \left| \int_{X^T\beta_\tau}^{r} f_{T|X}(t)\, f_{R|X}(r)\, dt \right| dr\, dP_X
\end{aligned}
$$

Condition 10 insures that the integrand of the inner integrals is strictly positive. Since the integrands, $f_{T|X}(t)\, f_{L|X}(l)$, and $f_{T|X}(t)\, f_{R|X}(r)$, are positive, we can conclude that the inner most integrals are positive.We can then conclude that, $Z(\beta) - Z(\beta_\tau) > 0$ for all $\beta \neq \beta_\tau$ and hence, β_τ is identifiable. □

Proof of Theorem 4.3.2. We shall prove this theorem by showing

$$\sup_{\beta \in \mathcal{B}} |Z_{n,\epsilon}(\beta) - Z(\beta)| \to 0 \quad \text{almost surely} \tag{4.33}$$

and then showing $Z(\beta)$ is continuous. By Lemma 4.3.1, β_τ is the unique minimizer of $Z(\cdot)$ and since $\mathcal{B}$ is assumed, we can use Theorem 2.1 of Newey and McFadden (1994) to conclude that $\hat{\beta}_{n,\epsilon} \to \beta_\tau$ in probability.

We can show Equation (4.33) is true by proving $\sup_{\beta \in \mathcal{B}} |Z_n(\beta) - Z(\beta)| \to 0$ almost surely and $\sup_{\beta \in \mathcal{B}} |Z_{n,\epsilon}(\beta) - Z_n(\beta)| \to 0$ almost surely since

$$\sup_{\beta \in \mathcal{B}} |Z_{n,\epsilon}(\beta) - Z(\beta)| \leq \sup_{\beta \in \mathcal{B}} |Z_{n,\epsilon}(\beta) - Z_n(\beta)| + \sup_{\beta \in \mathcal{B}} |Z_n(\beta) - Z(\beta)| \tag{4.34}$$

The class of indicator functions $I(T \leq W)$, $I(T > W)$, for $W \in \{R, L\}$, $\mathcal{I}_1 \equiv \{I(X^T\beta - W \leq 0) : \beta \in \mathcal{B}, W \in \{R, L\}\}$, and $\mathcal{I}_2 \equiv \{I(X^T\beta - W > 0) : \beta \in \mathcal{B}, W \in \{R, L\}\}$ are Vapnik-Červonenkis (VC)-subgraph classes. τ and $1 - \tau$ are fixed functions and thus by Lemma 2.6.18 (i) and (vi) of van der Vaart and Wellner (1996), the classes $\tau\, I(T > W)\, \mathcal{I}_1$ and $(1 - \tau)\, I(T \leq W)\, \mathcal{I}_2$ are also VC-subgraph classes for $W \in \{R, L\}$ Finally, (v) of the same lemma gives that $\mathcal{Z} \equiv \{Z_n(\beta) : \beta \in \mathcal{B}\}$ is a VC-subgraph class. Since $\mathcal{Z}$ is a VC-subgraph class, it is also a Glivenko-Cantelli class; hence, $\sup_{\beta \in \mathcal{B}} |Z_n(\beta) - Z(\beta)| \to 0$ almost surely.

Since $I(|X^T\beta - W| \leq \epsilon/2)$ is a VC class of functions for $W \in \{R, L\}$, $\mathbf{P}_n[I(|X^T\beta - W| \leq \epsilon/2)]$ converges to $P[I(|X^T\beta - W| \leq \epsilon/2)]$ uniformly over $\mathcal{B}$ where $\mathbf{P}_n$ is the empirical measure and P is the true underlying measure. Thus, we have

$$
\begin{aligned}
\sup_{\beta\in\mathcal{B}} |Z_{n,\epsilon}(\beta) - Z_n(\beta)| \leq\ & \sup_{\beta\in\mathcal{B}} \mathbf{P}_n[I(|X^T\beta - L| \leq \epsilon/2) + I(|X^T\beta - R| \leq \epsilon/2)] \\
& \underset{n\uparrow\infty}{\longrightarrow} \sup_{\beta\in\mathcal{B}} P[I(|X^T\beta - L| \leq \epsilon/2) + I(|X^T\beta - R| \leq \epsilon/2)] \\
& \leq \sup_{\beta\in\mathcal{B}} P_X[P_{L|X}(|X^T\beta - L| \leq \epsilon/2|X)] + \sup_{\beta\in\mathcal{B}} P_X[P_{R|X}(|X^T\beta - R| \leq \epsilon/2|X)] \\
& \leq P_X(\epsilon M_L|X) + P_X(\epsilon M_R|X) = \epsilon(M_L + M_R).
\end{aligned} \tag{4.35}
$$

By Condition 12, Equation (4.35) is bounded and converges to 0 as $\epsilon \to 0$, thus we can conclude that $\sup_{\beta\in\mathcal{B}} |Z_{n,\epsilon}(\beta) - Z_n(\beta)| \to 0$ almost surely as $n \to \infty$ and $\epsilon \to 0$. Since each term on the right hand side of Equation (4.34) converges to 0 almost surely, we can conclude that $\sup_{\beta\in\mathcal{B}} |Z_{n,\epsilon}(\beta) - Z(\beta)| \to 0$ almost surely as $n \to \infty$ and $\epsilon \to 0$.

To show that $Z(\cdot)$ is continuous, we re-express $Z(\cdot)$ as

$$
\begin{aligned}
Z(\beta) = E\big[&\tau I(T > L) I(X^T\beta - L \leq 0) + (1-\tau) I(T \leq L) I(X^T\beta - L > 0) \\
&+ \tau I(T > R) I(X^T\beta - R \leq 0) + (1-\tau) I(T \leq R) I(X^T\beta - R > 0)\big] \\
= & \int_X \left\{ \tau \int_{X^T\beta}^{\infty} \left[1 - F_{T|X}(l)\right] dF_{L|X} + (1-\tau) \int_0^{X^T\beta} F_{T|X}(l) dF_{L|X} \right\} dF_X \\
& + \int_X \left\{ \tau \int_{X^T\beta}^{\infty} \left[1 - F_{T|X}(r)\right] dF_{R|X} + (1-\tau) \int_0^{X^T\beta} F_{T|X}(r) dF_{R|X} \right\} dF_X \\
= & \int_X \left\{ \int_{X^T\beta}^{\infty} \tau \, dF_{L|X} + \int_0^{X^T\beta} F_{T|X}(l) \, dF_{L|X} - \tau \int_0^{\infty} F_{T|X}(l) \, dF_{L|X} \right\} dF_X \\
& + \int_X \left\{ \int_{X^T\beta}^{\infty} \tau \, dF_{R|X} + \int_0^{X^T\beta} F_{T|X}(r) \, dF_{R|X} - \tau \int_0^{\infty} F_{T|X}(r) \, dF_{R|X} \right\} dF_X.
\end{aligned} \tag{4.36}
$$

Only the first two inner integrals for L and R are functions of β. Under Condition 12, both of these inner integrals are bounded and continuous with respect to β; therefore, $Z(\cdot)$

is continuous. $\square$

Before proceeding with the proof, we will state the main theorem from Kim and Pollard (1990). The theorem concerns estimators defined by minimization of process $\mathbf{P}_n g(\cdot\,,\theta) = \frac{1}{n}\sum_{i\le n} g(\xi_i,\theta)$, where $\{\xi_i\}$ is a sequence of independent observations taken from a distribution P and $\{g(\cdot,\theta) : \theta\in\Theta\}$ is a class of functions indexed by a subset Θ of $\Re^d$. $\mathbf{P}_n$ denotes the expectation with respect to the empirical process. The envelope $G_R(\cdot)$ is defined as the supremum of $|g(\cdot,\theta)|$ over the class $\mathbf{g}_R = \{g(\cdot,\theta) : |\theta-\theta_o|\le R\}$.

Kim and Pollard (1990) *Let $\{\theta_n\}$ be a sequence of estimators for which*

(i) $\mathbf{P}_n g(\cdot\,,\theta_n) \le \inf_{\theta\in\Theta}\mathbf{P}_n g(\cdot,\theta) + o_p(n^{-2/3})$.

Suppose

(ii) θ_n converges in probability to the unique θ_0 that minimizes $Pg(\cdot,\theta)$;

(iii) θ_0 is an interior point of Θ.

Let the functions be standardized so that $g(\cdot,\theta_0)\equiv 0$. If the classes $\mathbf{g}_R$, for R near 0, are uniformly manageable for the envelopes G_R and satisfy

(iv) $Pg(\cdot,\theta)$ is twice differentiable with second derivative matrix V at θ_0;

(v) $H(s,t) = \lim_{\alpha\to\infty}\alpha Pg(\cdot,\theta_0+s/\alpha)g(\cdot,\theta_0+t/\alpha)$ exists for each s, t in $\Re^d$ and

$$\lim_{\alpha\to\infty}\alpha Pg(\cdot,\theta_0+t/\alpha)^2 I\left\{|g(\cdot,\theta_0+t/\alpha)|>\epsilon\alpha\right\}=0$$

for each $\epsilon>0$ and t in $\Re^d$;

(vi) $PG_R^2 = O(R)$ as $R\to 0$ and for each $\epsilon>0$ there is a constant K such that $PG_R^2 I\{G_R>K\}<\epsilon R$ for R near 0;

(vii) $P|g(\cdot,\theta_1)-g(\cdot,\theta_2)| = O(|\theta_1-\theta_2|)$ near θ_0;

then the process $n^{2/3}\mathbf{P}_n g(\cdot,\theta_0+tn^{-1/3})$ converges in distribution to a Gaussian process $Z(t)$ with continuous sample paths, expected value $t^T Vt/2$ and covariance kernel H.

If V is positive definite and if Z has nondegenerate increments, then $n^{1/3}(\theta_n-\theta_0)$ converges in distribution to the (almost surely unique) random vector that minimizes $Z(t)$.

Now we proceed with our proof of theorem 2.

Proof of Theorem 3.3.3. It will be convenient to define a version of the original objective function centered at the true value β_τ,

$$\begin{aligned} g(\beta) = \tau\{&I(T > L)I(X^T\beta - L \le 0) + I(T > R)I(X^T\beta - R \le 0)\} \\ &+ (1-\tau)\{I(T \le L)I(X^T\beta - L > 0) + I(T \le R)I(X^T\beta - R > 0)\} \\ &- \tau\{I(T > L)I(X^T\beta_\tau - L \le 0) - I(T > R)I(X^T\beta_\tau - R \le 0)\} \\ &- (1-\tau)\{I(T \le L)I(X^T\beta_\tau - L > 0) - I(T \le R)I(X^T\beta_\tau - R > 0)\} \\ = [\tau - I(T \le L)]&\left[I(X^T\beta \le L < X^T\beta_\tau) - I(X^T\beta_\tau \le L < X^T\beta)\right] \\ &+ [\tau - I(T \le R)]\left[I(X^T\beta \le R < X^T\beta_\tau) - I(X^T\beta_\tau \le R < X^T\beta)\right]. \end{aligned}$$

Under the true distribution P, we have $P(g(\beta)) = Z(\beta) - Z(\beta_\tau)$. The minimum value of $P(g(\cdot))$ is then obtained at the arg min of $Z(\cdot)$ and $P(g(\beta_\tau)) = 0$. The estimator we use here is $\hat{\beta}_{n,\epsilon}$ which is the minimizer of $Z_{n,\epsilon}(\cdot)$ defined as Equation (4.27).

The first condition of the main theorem from Kim and Pollard (1990) is satisfied under Condition 13. By the definition of $\hat{\beta}_{n,\epsilon}$, we have $Z_{n,\epsilon}(\hat{\beta}_{n,\epsilon}) \le \inf_\beta Z_{n,\epsilon}(\beta) + o_p(n^{-2/3})$. We also have $Z_n(\hat{\beta}_{n,\epsilon}) - \inf_\beta Z_n(\beta) \ge 0$ by definition.

$$\begin{aligned} Z_n(\hat{\beta}_{n,\epsilon}) - \inf_\beta Z_n(\beta) \;\le\; & Z_{n,\epsilon}(\hat{\beta}_{n,\epsilon}) + \mathbf{P}_n[I(|L - X^T\beta_{n,\epsilon}| < \epsilon/2)] + \mathbf{P}_n[I(|R - X^T\beta_{n,\epsilon}| < \epsilon/2)] \\ & - \inf_{\beta\in\beta}\left\{Z_{n,\epsilon}(\beta) - \mathbf{P}_n[I(|L - X^T\beta| < \epsilon/2)] - \mathbf{P}_n[I(|R - X^T\beta| < \epsilon/2)]\right\} \\ \le\; & Z_{n,\epsilon}(\hat{\beta}_{n,\epsilon}) - \inf_\beta Z_{n,\epsilon}(\beta) + 2\sup_\beta\left\{\mathbf{P}_n[I(|L - X^T\beta| < \epsilon/2)] + \mathbf{P}_n[I(|R - X^T\beta| < \epsilon/2)]\right\} \\ =\; & Z_{n,\epsilon}(\hat{\beta}_{n,\epsilon}) - \inf_\beta Z_{n,\epsilon}(\beta) \\ & + 2\sup_\beta\left\{\mathbf{P}(|L - X^T\beta| < \epsilon/2) + \mathbf{P}(|R - X^T\beta| < \epsilon/2)\right\} + o_p(\epsilon\, n^{-1/2}) \\ \le\; & Z_{n,\epsilon}(\hat{\beta}_{n,\epsilon}) - \inf_\beta Z_{n,\epsilon}(\beta) + 2\epsilon(M_L + M_R) + o_p(n^{-7/6}) = o_p(n^{-2/3}) \end{aligned}$$

Therefore, $Z_n(\hat{\beta}_{n,\epsilon}) \leq \inf_\beta Z_n(\beta) + o_p(n^{-2/3})$ which satisfies the first condition. The second condition, $\hat{\beta}_{n,\epsilon} \to \beta_\tau$ in probability, has been verified in Theorem 4.3.2. The third condition is satisfied by assuming Condition 10.

The remaining four conditions of the theorem deal with the nature of expectation of g under the measure P. $P(g)$ may be expressed as

$$\begin{aligned}
P(g(\beta)) = \; & P_{L,R,X}\left\{[\tau - I(T \leq L)]\left[I(X^T\beta \leq L < X^T\beta_\tau) - I(X^T\beta_\tau \leq L < X^T\beta)\right]\right. \\
& \left. + [\tau - I(T \leq R)]\left[I(X^T\beta \leq R < X^T\beta_\tau) - I(X^T\beta_\tau \leq R < X^T\beta)\right]\right\} \\
= \; & P_{L,X}\left\{\left[\tau - F_{T|X}(L|X)\right]\;\left[I(X^T\beta \leq L < X^T\beta_\tau) - I(X^T\beta_\tau \leq L < X^T\beta)\right]\right\} \\
& + P_{R,X}\left\{\left[\tau - F_{T|X}(R|X)\right]\;\left[I(X^T\beta \leq R < X^T\beta_\tau) - I(X^T\beta_\tau \leq R < X^T\beta)\right]\right\}
\end{aligned}$$

where $F_{T|X}$ is the conditional distribution of T given X. Since $|\tau - F_{T|X}(\cdot|X)| < 1$, the expectation is dominated by:

$$\begin{aligned}
P|g(\beta)| \leq \; & P_{L,X}\left\{I(X^T\beta \leq L < X^T\beta_\tau) - I(X^T\beta_\tau \leq L < X^T\beta)\right\} \\
& + P_{R,X}\left\{I(X^T\beta \leq R < X^T\beta_\tau) - I(X^T\beta_\tau \leq R < X^T\beta)\right\} \\
\leq \; & P_{L,X}\left\{I(X^T\beta \leq L < X^T\beta_\tau) + I(X^T\beta_\tau \leq L < X^T\beta)\right\} \\
& + P_{R,X}\left\{I(X^T\beta \leq R < X^T\beta_\tau) + I(X^T\beta_\tau \leq R < X^T\beta)\right\} \leq 2
\end{aligned}$$

Since P is absolutely continuous with respect to Lebesgue measure, for any sequence $d_n \to 0$, the dominated convergence theorem tells us $P(g(\beta + d_n)) \to P(g(\beta))$. In other words, $P(g(\beta))$ is continuous with respect to β.

We may expand $P(g(\beta))$ with a Taylor expansion. The first derivative is found by interchanging integration (expectation) and differentiation to find

$$\begin{aligned}
&\frac{\partial}{\partial \beta_i} P(g(\beta)) = \\
&\quad \int_X \left[\tau - F_{T|X}(X^T\beta|X)\right] X_i f_{L|X}(X^T\beta|X)[-I(X^T\beta < X^T\beta_\tau) - I(X^T\beta_\tau < X^T\beta)]\, dP_X \\
&\qquad + \int_X \left[\tau - F_{T|X}(X^T\beta|X)\right] X_i f_{R|X}(X^T\beta|X)[-I(X^T\beta < X^T\beta_\tau) - I(X^T\beta_\tau < X^T\beta)]\, dP_X \\
&\quad = \int_X \left[F_{T|X}(X^T\beta|X) - \tau\right] X_i (f_{L|X}(X^T\beta|X) + f_{R|X}(X^T\beta|X))\, dP_X,
\end{aligned}$$

where X_i is an element of $\mathbf{X_i}$. Evaluated at β_τ, the term $F_{T|X}(X^T\beta_\tau|X) - \tau$ is equal to zero by definition of the τth quantile, making the derivative equal zero as would be expected for an extrema. Taking one step further, the second derivative would be

$$\begin{aligned}
V(\beta)_{i,j} &= \frac{\partial^2}{\partial \beta_i \partial \beta_j} P(g(\beta)) \\
&= \int_X \left\{F_{T|X}(X^T\beta|X) - \tau\right\} X_i X_j \left(\frac{\partial}{\partial L} f_{L|X}(X^T\beta|X) + \frac{\partial}{\partial R} f_{R|X}(X^T\beta|X)\right) dP_X \\
&\quad + \int_X f_{T|X}(X^T\beta|X) X_i X_j [f_{L|X}(X^T\beta|X) + f_{R|X}(X^T\beta|X)]\, dP_X.
\end{aligned}$$

At β_τ, the first integral vanishes and only the second remains taking the form

$$V(\beta_\tau)_{i,j} = \int_X f_{T|X}(X^T\beta|X) X_i X_j [f_{L|X}(X^T\beta|X) + f_{R|X}(X^T\beta|X)]\, dP_X$$

As the entries are dominated by $|V(\beta_\tau)_{i,j}| \leq (M_R + M_L) M_T M_{|X|}^2$, where $M_{|X|}$ is the bound over all $|X_i|$ and M_C and M_T are defined in Condition 12, $V(\beta_\tau)_{i,j}$ will be well defined verifying the fourth condition of the theorem. Writing $V(\beta_\tau) = P_X(XX^T h(X))$ with $h(X) = [f_{L|X}(X^T\beta|X) + f_{R|X}(X^T\beta|X)] f_{T|X}(X^T\beta|X) \geq 0$, show that $V(\beta_\tau)$ would be a symmetric positive semi-definite matrix since it is a positive mixture of the positive semi-definite terms XX^T. A sufficient condition for $V(\beta_\tau)$ to be positive definite is that the Lebesgue measure of the set $\{X : h(X) > 0\}$ is greater than zero.

To control asymptotic covariance of $Z(s)$, let

$$
\begin{aligned}
H(s,q) &= \lim_{\alpha\to\infty} \alpha P[g(\beta_\tau + \frac{q}{\alpha})\, g(\beta_\tau + \frac{s}{\alpha})] \\
&= \lim_{\alpha\to\infty} \alpha P_{L,R,X}\Bigg\{\Bigg[\Big\{\big[\tau - F_{T|X}(L|X)\big]\ \Big[I(X^T(\beta_\tau + \frac{q}{\alpha}) \le L < X^T\beta_\tau) \\
&\qquad\qquad -I(X^T\beta_\tau \le L < X^T(\beta_\tau + \frac{q}{\alpha}))\Big]\Big\} \\
&+ \Big\{\big[\tau - F_{T|X}(R|X)\big]\ \Big[I(X^T(\beta_\tau + \frac{q}{\alpha}) \le R < X^T\beta_\tau) - I(X^T\beta_\tau \le R < X^T(\beta_\tau + \frac{q}{\alpha}))\Big]\Big\}\Bigg] \\
&\times \Bigg[\Big\{\big[\tau - F_{T|X}(L|X)\big]\ \Big[I(X^T(\beta_\tau + \frac{s}{\alpha}) \le L < X^T\beta_\tau) - I(X^T\beta_\tau \le L < X^T(\beta_\tau + \frac{s}{\alpha}))\Big]\Big\} \\
&+ \Big\{\big[\tau - F_{T|X}(R|X)\big]\ \Big[I(X^T(\beta_\tau + \frac{s}{\alpha}) \le R < X^T\beta_\tau) - I(X^T\beta_\tau \le R < X^T(\beta_\tau + \frac{s}{\alpha}))\Big]\Big\}\Bigg]\Bigg\} \\
&= \lim_{\alpha\to\infty} \alpha\Bigg[P_{L,X}\Bigg\{\ \big[\tau - F_{T|X}(L|X)\big]^2 \\
&\quad \Bigg[I\left(\frac{X^Tq \vee X^Ts}{\alpha} \le L - X^T\beta_\tau < 0\right) + I\left(0 \le L - X^T\beta_\tau < \frac{X^Tq \wedge X^Ts}{\alpha}\right)\Bigg]\Bigg\} \\
&+ P_{R,X}\Bigg\{\ \big[\tau - F_{T|X}(R|X)\big]^2 \\
&\qquad \Bigg[I\left(\frac{X^Tq \vee X^Ts}{\alpha} \le R - X^T\beta_\tau < 0\right) + I\left(0 \le R - X^T\beta_\tau < \frac{X^Tq \wedge X^Ts}{\alpha}\right)\Bigg]\Bigg\} \\
&+ P_{L,R,X}\Bigg\{\ \big[\tau - F_{T|X}(L|X)\big] \times \big[\tau - F_{T|X}(R|X)\big] \\
&\quad \Big(I(X^T(\beta_\tau + \frac{q}{\alpha}) \le L < X^T\beta_\tau) I(X^T(\beta_\tau + \frac{s}{\alpha}) \le R < X^T\beta_\tau) \\
&\quad + I(X^T\beta_\tau \le L < X^T(\beta_\tau + \frac{q}{\alpha})) I(X^T\beta_\tau \le R < X^T(\beta_\tau + \frac{s}{\alpha})) \\
&\quad + I(X^T(\beta_\tau + \frac{q}{\alpha}) \le R < X^T\beta_\tau) I(X^T(\beta_\tau + \frac{s}{\alpha}) \le L < X^T\beta_\tau) \\
&\quad + I(X^T\beta_\tau \le R < X^T(\beta_\tau + \frac{q}{\alpha})) I(X^T\beta_\tau \le R < X^T(\beta_\tau + \frac{s}{\alpha})) \\
&\quad - I(X^T(\beta_\tau + \frac{s}{\alpha}) \le L < X^T\beta_\tau \le R < X^T(\beta_\tau + \frac{q}{\alpha})) \\
&\quad -I(X^T(\beta_\tau + \frac{q}{\alpha}) \le L < X^T\beta_\tau \le R < X^T(\beta_\tau + \frac{s}{\alpha}))\Big)\Bigg\}\Bigg]
\end{aligned}
\tag{4.37}
$$

where $\vee$ and $\wedge$ denote maximum and minimum, respectively. Using Condition 12 and 15, we have

$$
\begin{aligned}
|P[g(\beta_\tau + q/\alpha)g(\beta_\tau + s/\alpha)]| \;&\le\; P\{|g(\beta_\tau + q/\alpha)||g(\beta_\tau + s/\alpha)|\} \;\le\; 2 \cdot P\{|g(\beta_\tau + s/\alpha)|\} \\
&\le\; 2\int_X (M_L + M_R)\frac{|X^T s|}{\alpha}\, dP_X \;\le\; 2(M_L + M_R)\frac{(\|s\|_1)M_{|X|}}{\alpha} \;=\; O(\alpha^{-1}),
\end{aligned}
$$

hence, along with Condition 14, $H(s, r)$ is well defined by the dominated convergence theorem satisfying the fifth condition.

Let G_Q be the envelope of $\mathbf{g}_Q \equiv \{g(\beta) : \|\beta - \beta_\tau\|_\infty < Q \le \tilde{\epsilon}\}$, i.e.,

$$
\begin{aligned}
G_Q &= |\tau - I(T \le L)|\, I(|L - X^T\beta_\tau| < Q \max|X_j|) + \\
&\qquad |\tau - I(T \le R)|\, I(|R - X^T\beta_\tau| < Q \max|X_j|) \\
&\le I(|L - X^T\beta_\tau| < Q \max|X_j|) + I(|R - X^T\beta_\tau| < Q \max|X_j|)
\end{aligned}
$$

A sufficient condition for the class $\mathbf{g}_Q$ to be uniformly manageable is that its envelope function G_Q is uniformly square integrable given that $\{g(\beta)\}$ is VC-subgraph. Since G_Q is bounded by two, it is uniformly square integrable for Q close to zero. Together with the fact that $\{g(\beta)\}$ is a VC-subgraph, we conclude that $\mathbf{g}_Q$ is uniformly manageable. Then

$$
\begin{aligned}
P(G_Q^2) &\le \int_X \int_{\Re} I(|L - X^T\beta_\tau| < Q \max|X_j|) + I(|R - X^T\beta_\tau| < Q \max|X_j|)\, dP_{L,R|X}\, dP_X \\
&\le \int_X 2(M_L + M_R) Q \max(|X_j|)\, dP_X \le Q[2(M_L + M_R)M_{|X|}] = O(Q).
\end{aligned}
$$

For any $\epsilon > 0$, we can use $K = 3$, then $E(G_Q^2\, I(G_R > K)) = 0 < \epsilon Q$ since G_R is less than K everywhere. Combining these two traits shows that the sixth condition of the theorem is satisfied. The final condition is verified by letting $G_{Q,\beta}$ be the envelope of $\{g(\tilde{\beta}) - g(\beta) :$

$\|\tilde{\beta} - \beta\|_\infty < Q\}$, i.e.,

$$G_{Q,\beta} = |\tau - I(T \le L)|\, I(|L - X^T\beta| < Q \max |X_j|) + |\tau - I(T \le R)|\, I(|R - X^T\beta| < Q \max |X_j|).$$

Using the same integration inequalities as used in the preceding for G_Q we find that $P|g(\tilde{\beta}) - g(\beta)| \le O(\|\tilde{\beta} - \beta\|_\infty) = O(\|\tilde{\beta} - \beta\|_1)$ over all $\beta, \tilde{\beta}$ in an $\tilde{\epsilon}$ neighborhood of β_τ since $\|\cdot\|_\infty$ and $\|\cdot\|_1$ are equivalent metrics.

As the seven conditions are satisfied, the conclusion of the main theorem in Kim and Pollard (1990) follows which in turn proved Theorem 2. $\square$

CHAPTER 5: SEMIPARAMETRIC METHODS FOR ACCELERATED FAILURE TIME MODELS FOR INTERVAL-CENSORED DATA

5.1 Introduction

Interval-censored data occurs when continuous inspection is not feasible and the event of interest is only known to occur during a time interval. This is common in longitudinal studies, clinical trials, and epidemiological studies since the information is only collected at the time of a follow-up or clinical visit. A special case of interval-censored data is called "current status data" or "Class I" interval-censored data. In current status data, each subject is only observed once so the information we have is the observation time and whether an event has occurred prior to the observation time. Several likelihood-based methods have been developed to perform regression analysis on interval-censored data. For current status data, Huang (1996) proposed efficient estimation for the proportional hazard models, Rossini and Tsiatis (1996) studied proportional odds models, and both Lin et al. (1998) and Martinussen and Scheike (2002) developed methods for additive hazard models. For general interval-censored data, methods for proportional hazards models (Finkelstein and Wolfe 1985, Betensky et al. 2002, Huang and Wellner 1997) and proportional odds models (Huang and Rossini 1997, Huang and Wellner 1997, Rabinowitz et al. 2000, Shen 1998) are also available.

We consider an accelerated failure time (AFT) model where the logarithm of the survival time is a linear function of the covariates,

$$\log(T) = X^T\beta + \xi, \tag{5.38}$$

where T denotes the survival time and X is the p-dimensional covariate vector. The error term, ξ, is independent and identically distributed (iid) with distribution G and is independent of X. When the distribution of the error term is left unspecified, the AFT model can be considered as a semiparametric alternative approach to the proportional hazards and proportional odds models for survival analysis. A likelihood-based approach for fitting the AFT model is difficult because the regression parameter β and the distribution G are both present in the likelihood function. Rabinowitz et al. (1995) proposed a class of score statistics that may be used for estimation and confidence procedures. Under the current status data setting, Murphy et al. (1999) and Shen (2000) developed likelihood-based methods. Murphy et al. (1999) considered a penalized nonparametric maximum likelihood estimator and Shen (2000) constructed likelihoods based on the random-sieve likelihood concept. Under a general interval-censored data setting, Betensky et al. (2001) studied a simple numerically efficient estimation procedure. The examination time and event time were assumed to be independent in Betensky et al. (2001). To overcome the numerical difficulty presented in previous methods and to include higher-dimensional covariates, Tian and Cai (2006) proposed to construct the estimator by inverting a Wald-type test for testing a null proportional hazards model. The method proposed by Tian and Cai (2006) can be implemented using a grid search when the covariate is one-dimensional and a Markov chain Monte Carlo based procedure when the covariates are high-dimensional. Tian and Cai (2006) applied their method to both current status data and general interval-censored data.

We propose a semiparametric method to analyze interval-censored data using the accelerated failure time model. We take advantage of the quantile regression framework and construct the estimators by combining information over multiple quantiles. The proposed method is applicable to both current status data and general interval-censored data. We start with fitting the model to general interval-censored data.

5.2 Model and Inference Procedure

5.2.1 Models and Data

Under the AFT model defined in (5.38), our interest lies in the estimation and inference of β. Consider a quantile regression model (Koenker and Bassett 1978) for the logarithm of failure time,

$$Q_{\log(T)}(\tau \mid X) = \beta_0(\tau) + X^T\beta(\tau), \qquad \tau \in (0,1), \tag{5.39}$$

where $Q_{\log(T)}(\tau \mid X)$ is the conditional quantile for logarithm of the survival time and defined as $Q_{\log(T)}(\tau \mid X) = \inf\{t : \mathrm{pr}(\log(T) \leq t \mid X) \geq \tau\}$. The unknown intercept, $\beta_0(\tau)$ and the vector of unknown regression coefficients, $\beta(\tau)$, represent the covariate effects on the τth quantile of $\log(T)$ which may depend on τ. $\beta_0(\tau)$ and each element of $\beta(\tau)$ can be interpreted as an estimated difference in τth quantile by one unit change of the corresponding covariate while other variables in the model are held constant. Under the assumption that the error term, ξ, in (5.38) is iid and is independent of X, the quantile functions of $\log(T)$ should have the same coefficients at different quantiles except the intercept term. This is equivalent to formulating the quantiles of $\log(T)$ as

$$Q_{\log(T)}(\tau|X) = Q_\xi(\tau|X) + X^T\beta, \qquad \tau \in (0,1). \tag{5.40}$$

We can then use an estimate for (5.40) as an estimate for (5.38).

Let $\{(L_i, R_i, \delta_{i1}, \delta_{i2}, X_i) : i = 1, \cdots, n\}$ be n independent and identically distributed realizations of $(L, R, \delta_1, \delta_2, X)$. The random observation times L and R satisfy $L < R$ with probability 1 and are assumed to be independent of ξ but they may depend on X. δ_1 and δ_2 are defined as $\delta_1 = I(T \leq L)$ and $\delta_2 = I(T \leq R)$ where $I(\cdot)$ denotes the indicator function and T is the survival time. It is assumed that T is conditionally independent of L and

R given X. The τth conditional quantile of $\log(T)$ conditional on X can be characterized as the solution to the expected loss minimization problem (Powell 1994),

$$Z(\beta) = E[E\{\rho_\tau(\log(T) - [\beta_0(\tau) + X^T\beta(\tau))]|X\}], \tag{5.41}$$

where $\rho_\tau(u) = u\{\tau - I(u < 0)\}$. Using the unique "equivariance to monotone transformations" (Koenker 2005) property of quantile regression, we can apply the monotone non-decreasing transformations, $h_1(T|L,R) = I(T > L)$ and $h_2(T|L,R) = I(T > R)$, to the conditional quantile, $\beta_0(\tau) + X^T\beta(\tau)$ and use the transformed conditional quantile, $I(\beta_0(\tau) + X^T\beta(\tau) > L)$ and $I(\beta_0(\tau) + X^T\beta(\tau) > R)$, in the analysis. We can thus substitute $(1-\delta_1)$, $I(\beta_0(\tau) + X^T\beta > L)$, $(1-\delta_1-\delta_2)$, and $I(\beta_0(\tau) + X^T\beta > R)$ in Equation (5.41) and get

$$Z^1(\beta) = E\{E[\rho_\tau\{(1-\delta_1) - [\beta_0(\tau) + I(X^T\beta(\tau)] > L)\}|X,L,R]\} \quad \text{and} \tag{5.42}$$

$$Z^2(\beta) = E\{E[\rho_\tau\{(1-\delta_1-\delta_2) - I([\beta_0(\tau) + X^T\beta(\tau)] > R)\}|X,L,R]\}. \tag{5.43}$$

We can estimate $\beta_0(\tau)$ and $\beta(\tau)$ by minimizing (5.42) and (5.43) simultaneously (see Section 4.2.1 for more details). A sufficient condition to guarantee the identifiability for a fixed quantile is provided in Section 4.3.1.

5.2.2 Parameter Estimation and Algorithm

Let β^* denote the true parameter vector in (5.38). To obtain consistent estimates of β^*, we can use the estimation routine developed in Section 4.2.2 for general interval-censored data. Let $\hat{\beta}_{0,n,\epsilon}(\tau)$ and $\hat{\beta}_{n,\epsilon}(\tau)$ denote the estimates of $\beta_0(\tau)$ and β, respectively, from a single τ. Under the iid assumption of ξ, any consistent $\hat{\beta}_{n,\epsilon}(\tau)$ can be used as the estimate of β^*. To increase efficiency, we propose combining the quantile estimates over k quantiles.

Specifically, we propose the following estimate of β^*,

$$\check{\beta}_{n,\epsilon} = \boldsymbol{\omega}\,\hat{\boldsymbol{\beta}}_{n,\epsilon}, \tag{5.44}$$

where

$$\boldsymbol{\omega} = \begin{bmatrix} \omega_{1,1} & \cdots & \omega_{1,k} \\ \vdots & \vdots & \vdots \\ \omega_{p,1} & \cdots & \omega_{p,k} \end{bmatrix} \text{ and } \hat{\boldsymbol{\beta}}_{n,\epsilon} = \begin{bmatrix} \hat{\beta}_{n,\epsilon}(\tau_1)^T \\ \vdots \\ \hat{\beta}_{n,\epsilon}(\tau_k)^T \end{bmatrix}. \tag{5.45}$$

$\hat{\beta}_{n,\epsilon}(\tau_i)$ is the quantile estimate from individual quantiles τ_i and $\boldsymbol{\omega}$ is a weight matrix where each column is a weight vector for a specific quantile. It is desired that $\boldsymbol{\omega}$ is the set of weights which minimize the asymptotic variance of the random variables $\tilde{\beta}_{n,\epsilon}$. More specifically, we define each row by the following

$$\boldsymbol{\omega}_i = \underset{\Pi \in \Re^k : \sum \Pi_j = 1}{\arg\inf} \; \Pi^T \, \Sigma_i \, \Pi,$$

for $i = 1, \ldots, p$ where Σ_i is the (asymptotic) covariance matrix for the estimators of the ith element of β across the quantiles. We intensionally do not apply the same weighting scheme to $\beta_0(\tau)$ since it can be considered as being absorbed into the error term, ξ, in (5.38).

Since $\boldsymbol{\omega}$ is unknown in practice, we can use the estimated weight matrix, $\hat{\boldsymbol{\omega}}_{n,\epsilon}$, to calculate the final estimators,

$$\tilde{\beta}_{n,\epsilon} = \hat{\boldsymbol{\omega}}_{n,\epsilon}\,\hat{\boldsymbol{\beta}}_{n,\epsilon}, \tag{5.46}$$

where

$$\hat{\boldsymbol{\omega}}_{n,\epsilon} = \begin{bmatrix} \hat{\omega}_{1,1} & \cdots & \hat{\omega}_{1,k} \\ \vdots & \vdots & \vdots \\ \hat{\omega}_{p,1} & \cdots & \hat{\omega}_{p,k} \end{bmatrix}. \tag{5.47}$$

Our proposed method for estimating $\hat{\boldsymbol{\omega}}_{n,\epsilon}$ is described in Section 5.2.3. We use the term "weighted quantile average estimator" for $\tilde{\beta}_{n,\epsilon}$ henceforth.

5.2.3 Inference

A $k \times k$ empirical variance-covariance matrix can be computed from the subsampling estimates for each estimator $\hat{\beta}(\tau_1), \cdots, \hat{\beta}(\tau_k)$. $\boldsymbol{\omega}$, is then estimated as a solution to

$$\hat{\boldsymbol{\omega}}_{n,\epsilon,i} = \underset{\Pi \in \Re^k : \sum \Pi_j = 1}{\arg\inf} \; \Pi^T \, \hat{\Sigma}_{n,\epsilon,i} \, \Pi,$$

where $\hat{\Sigma}_{n,\epsilon,i}$ is the estimated covariance matrix of the ith element of β across quantiles observed from the subsample estimates and solved using any standard quadratic programming technique.

After obtaining $\hat{\boldsymbol{\omega}}_{n,\epsilon}$, we can use the weighted subsampling results to construct confidence intervals. Let $K_1, K_2, \cdots, K_M$ denote M subsamples of $\{(L_i, R_i, \delta_{i1}, \delta_{i2}, X_i)_{i=1,\cdots,n}\}$ of size b. Let $\tilde{\beta}_{n,\epsilon,q} = \hat{\boldsymbol{\omega}}_{n,\epsilon} \, \hat{\boldsymbol{\beta}}_{n,\epsilon,q}$ denote the estimated values produced using the K_qth dataset, where

$$\tilde{\boldsymbol{\beta}}_{n,\epsilon,q} = \begin{bmatrix} \tilde{\beta}_{n,\epsilon,q}(\tau_1)^T \\ \vdots \\ \tilde{\beta}_{n,\epsilon,q}(\tau_k)^T \end{bmatrix}.$$

Define

$$L_{n,i}(x) = M^{-1} \sum_{q=1}^{M} I \left\{ b^{1/3} (\tilde{\beta}_{n,\epsilon,q,i} - \tilde{\beta}_{n,\epsilon,i}) \leq x \right\} \quad \text{and} \quad c_{n,i}(\gamma) = \inf\{x : L_{n,i}(x) \geq \gamma\},$$

for $i = 1, \ldots, p$.

An asymptotic $1 - \alpha$ level confidence interval for the ith element of β^* can then be constructed with

$$\left[\tilde{\beta}_{n,\epsilon,i} - n^{-1/3} c_{n,i} \left(1 - \frac{\alpha}{2} \right), \, \tilde{\beta}_{n,\epsilon,i} - n^{-1/3} c_{n,i} \left(\frac{\alpha}{2} \right) \right].$$

Symmetric confidence intervals can be obtained by modifying the above approach slightly. Define

$$\tilde{L}_{n,i}(x) = M^{-1} \sum_{q=1}^{M} I\left\{ b^{1/3} |\tilde{\beta}_{n,\epsilon,q,i} - \tilde{\beta}_{n,\epsilon,i})| \leq x \right\} \quad \text{and} \quad \tilde{c}_{n,i}(\gamma) = \inf\{x : L_{n,i}(x) \geq \gamma\},$$

Again, if $b \to \infty$ as $n \to \infty$ and $b/n \to 0$, a symmetric confidence interval for the ith element of $\tilde{\beta}_{n,\epsilon}$ can be constructed as

$$\left[\tilde{\beta}_{n,\epsilon,i} - n^{-1/3} \tilde{c}_{n,i}(1-\alpha),\, \tilde{\beta}_{n,\epsilon,i} + n^{-1/3} \tilde{c}_{n,i}(1-\alpha) \right]. \tag{5.48}$$

Symmetric confidence intervals are desirable because they often have nicer properties than the nonsymmetric version in finite samples (Banerjee and Wellner 2005). This fact was also observed in our simulation studies; hence, symmetric confidence intervals are recommended and used in this paper.

To avoid large scale computation issues, a stochastic approximation from Politis et al. (1999) is employed where only B randomly chosen datasets from $\{1, 2, \cdots, N_n\}$ are used in the above calculation. Furthermore, the block size is chosen using the method implemented in Delgado et al. (2001).

The block size can be chosen based on the algorithm described in the end of Section 3.2.3 with a slight modification. For each of the block size considered, we can calculate $\hat{\boldsymbol{\omega}}_{n,\epsilon}$ using the subsample estimates. At Step 3 of the algorithm, we can use $\tilde{\beta}_{n,\epsilon}$ to calculate the average coverage then follow the algorithm to choose block size.

5.3 Asymptotic Properties

In this section, we prove the consistency and asymptotic distribution of the proposed estimators under current status data setting. The consistency and asymptotic distribution of the proposed estimators under Case II interval-censored data setting. can be generalized using the objective function, conditions, and Theorems in Section 4.3.

5.3.1 Consistency

Theorem 5.3.1. *If Conditions 1-4 of Section 3.3 are satisfied for each τ_i, $i = 1, \ldots, k$, and $\hat{\boldsymbol{\omega}}_{n,\epsilon} \to_p \boldsymbol{\omega}$ then $\tilde{\beta}_{n,\epsilon}$ converges in probability to β^* as $n \to \infty$ and $\epsilon \to 0$.*

Proof.

$$\tilde{\beta}_{n,\epsilon} = \hat{\boldsymbol{\omega}}_{n,\epsilon}\,\hat{\boldsymbol{\beta}}_{n,\epsilon} \to_d \boldsymbol{\omega}\,\boldsymbol{\beta}^* = \begin{bmatrix} \sum_{j=1}^{k} \omega_{1,j}\beta_1^* \\ \vdots \\ \sum_{j=1}^{k} \omega_{p,j}\beta_p^* \end{bmatrix} = \beta^*,$$

where the limit is an application of the continuous mapping theorem along with Theorem 3.3.2 and the final equality follows since the row sum of $\boldsymbol{\omega}$ is 1. Since the limit is to a single element, convergence in distribution is equivalent to convergence in probability. □

5.3.2 Asymptotic Distribution

Theorem 5.3.2. *If Conditions 1-8 of Section 3.3 hold for each τ_i ($i = 1, \ldots k$), and $\hat{\boldsymbol{\omega}}_{n,\epsilon} \to_p \omega$ then $n^{1/3}(\tilde{\beta}_{n,\epsilon} - \beta^*)$ has the same asymptotic distribution as $n^{1/3}\boldsymbol{\omega}\,(\hat{\boldsymbol{\beta}}_{n,\epsilon} - \boldsymbol{\beta}^*)$. The asymptotic form of $n^{1/3}(\tilde{\beta}_{n,\epsilon} - \beta^*)$ is a generalization of that described in Theorem 3.3.3.*

Proof.

$$(\tilde{\beta}_{n,\epsilon} - \beta^*) = (\hat{\boldsymbol{\omega}}_{n,\epsilon}\,\hat{\boldsymbol{\beta}}_{n,\epsilon} - \beta^*) = \hat{\boldsymbol{\omega}}_{n,\epsilon}\,(\hat{\boldsymbol{\beta}}_{n,\epsilon} - \boldsymbol{\beta}^*)$$

This final form has an asymptotic distribution identical the asymptotic distribution of $\boldsymbol{\omega}\,(\hat{\boldsymbol{\beta}}_{n,\epsilon} - \boldsymbol{\beta}^*)$, through an application of Slutsky's theorem.

The asymptotic distribution of $(\hat{\boldsymbol{\beta}}_{n,\epsilon} - \boldsymbol{\beta}^*)$ can be found by defining

$$
\begin{aligned}
g(\beta) = & \sum_{i=1}^{k} [\tau_i I(\delta = 0) I(c \geq \beta_i^* x) + (1 - \tau_i) I(\delta = 1) I(\beta_i^* x > c)] \\
& - \sum_{i=1}^{k} [\tau_i I(\delta = 0) I(c \geq \beta_i x) + (1 - \tau) I(\delta = 1) I(\beta_i x > c)] \\
= & \sum_{i=1}^{k} [I(\delta = 1) - \tau_i] \left[I(\beta_i x < c \leq Cx) - \sum_{i=1}^{k} I(\beta_i^* x < c < \beta_i x) \right],
\end{aligned}
$$

where β_i^* and β_i refer to the rows of the respective matrices, and these rows would only contain parameters for a single quantile regression estimate. With this g, the 7 conditions of Kim and Pollard (1990) theorem follow using Conditions 1–8 of Section 3.3 and duplicate the proof of Theorem 3.3.3 using sums of the individual quantile regressions. □

5.4 Simulation Studies

Three simulation studies were carried out to examine the finite sample performance of our estimators. All three simulation studies consider an AFT model where the true model is $\log(T) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \xi$ and $(\beta_0, \beta_1, \beta_2) = (2, 1, 1)$, $X_1 \sim$ Bernoulli(0.5), $X_2 \sim$ Uniform$[-1, 1]$.

In *Simulation 1*, ξ were generated from (standard) normal distributions and observation times, L and R, were generated from the linear model, $1.9 + I(X_1 = 1) + 1.1\, X_2 + 1.2\, V$ where V are iid $N(0, 1)$. In *Simulation 2*, ξ were generated from (standard) logistic distributions and the observation times, L and R, were generated from the linear model, $1.9 + 1.1\, X_2 + 1.5\, V + X_1(0.9 + 0.2\, X_2)$where V are iid $N(0, 1)$. In *Simulation 3*, ξ were generated from (standard, minimum) extreme-value distributions and the observation times, L and R, were generated from the linear model, $1.9 + 1.3\, X_2 + V + X_1(1 - 0.1\, X_2 + 0.5\, V)$ where V are generated from extreme-value distribution with mean equals 0.1 and scale equals 0.9.

For each scenario, we report the mean bias, mean squared error (MSE), standard deviation (Std), coverage probability (CP), and confidence interval length based on 1000 simulations. Sample sizes were chosen to be $n = 200, 400,$ and 800 for each simulation setup. In simulation studies, we considered $k = 3, 5,$ and 7. When $k = 3$, we used quantiles $0.25, 0.50,$ and 0.75. When $k = 5$, we used quantiles $1/6, 1/3, 0.50, 2/3,$ and 5/6. When $k = 7$, we used quantiles $1/6, 0.25, 1/3, 0.50, 2/3, 0.75,$ and 5/6.

For each simulated dataset, the procedure described at the end of Section 4.2.2 was used to estimate $\beta(\tau)$. Symmetric confidence intervals were calculated based on a stochastic approximation with 500 subsamples. To decrease the computational burden, the block size was determined via a pilot simulation in the same fashion as described in Banerjee and McKeague (2007). The optimal subsampling block size was determined from the following selected block sizes: $\{n^{1/3}, n^{1/2}, n^{2/3}, n^{3/4}, n^{0.8}, n^{0.85}, n^{0.9}, n^{0.95}\}$. The weight vector, ω_j, were estimated using method described in Section 5.2.3. To compare performance, we also used a naive weight vector which is defined as $[(1/k, \cdots, 1/k)_{p\times 1}]^T$. We used relative efficiency (RE) and confidence interval length ratio to compare the efficiency between the two estimates. Relative efficiency is defined as the ratio of MSE where a value of RE≥ 1 indicates better performance of the weighted quantile average estimator. The numerator used in the relative efficiency calculation is the smallest MSE among all individual quantiles, $1/6, 0.25, 1/3, 0.50, 2/3, 0.75,$ and 5/6. Confidence interval length ratio is defined as the ratio of confidence interval length where a value of confidence interval length ratio≥ 1 indicates better performance of the weighted quantile average estimator. The numerator used in the confidence interval length ratio calculation is the narrowest confidence interval length among all individual quantiles, $1/6, 0.25, 1/3, 0.50, 2/3, 0.75,$ and 5/6. In *Simulation 1* and *Simulation 2*, the MSE and confidence interval length of $\hat{\beta}(\tau = 0.5)$ was used as the numerator. In *Simulation 3*, the MSE and confidence interval length of $\hat{\beta}(\tau = 2/3)$ was used as the numerator.

Table 5.8–5.10, Table 5.11–5.13, and Table 5.14–5.11 summarize the results for *Simulation 1*, *Simulation 2*, and *Simulation 3* with sample size equal to 200, 400, and 800 at each individual quantile along with the two weighted results. In the tables, "Truth" is the true parameter value; "Bias" is the mean bias of the estimates from all replicates; "MSE" is the mean squared error; "Std" is the standard deviation of estimates; "CP" is the average coverage from subsampling symmetric confidence intervals; and "Length" is the average confidence interval length. "Equal$Z$" is the naive weighted estimates using Z quantiles and "Optimal$Z$" is the weighted estimates using $\hat{\omega}_j$ with Z quantiles where $Z \in \{3, 5, 7\}$.

Table 5.8 to 5.16 show that the regression coefficient estimators from individual quantiles have negligible bias and it is also true for the weighted estimators. The bias in general has a decreasing trend as the sample size increases for parameters from individual quantiles and for weighted estimator. The mean squared errors and standard deviations decrease as the sample size increases. The subsampling confidence interval coverage fluctuates around the nominal level 95% with slight over coverage for some instances. Table 5.17 shows the relative efficiency and Table 5.18 shows the confidence interval length ratio of $\tilde{\beta}$. There is an efficiency gain even for naive weight and this efficiency gain is more pronounced when the distribution is symmetric, as seen in the Normal and Logistic cases. The relative efficiency using $\hat{\omega}_j$ is almost always higher than using the naive weight but the difference is more pronounced for the asymmetric extreme-value distribution. The relative efficiency in general increased as the sample sizes increased. The incremental gain in relative efficiency was observed when number of quantiles, k, increased. However, the gain between combining 5 or 7 quantiles are not obvious which indicating that using a small number of quantiles may be sufficient to achieve a substantial improvement. Similar conclusions can be drawn for confidence interval length ratio.

5.5 Application

We apply our estimation and inference procedure on two real datasets in this section. The first dataset contains current status data from a calcification study (Yu et al. 2001). The calcification study investigated the effects of clinical variables on the time to calcification of intraocular lenses, which is an infrequently reported complication of cataract treatment. A patient's calcification status was determined by an ophthalmologist at a random time within 36 months after implantation of the intraocular lenses. The severity of calcification was graded 0, 1, $\cdots$, 5. For this analysis, we defined severity > 1 as calcified and severity ≤ 1 as no calcification. The covariates considered are gender (female vs. male) and age. Xue et al. (2004) and Cheng and Wang (2011) showed that the relationship between age and time to calcification is not simply linear; instead, patients around 60 years old enjoyed the longest time to calcification; therefore, the age variable entered the model as a linear spline with a knot at 60 then divided by 10. The dataset contains 379 records and we used the 378 records which have complete data for our analysis.

The nonparametric maximum likelihood estimator (NPMLE) (Wellner and Zhan 1997) of the (unobserved) logarithm of time to calcification distribution function stratified by each covariate is shown in Figure 5.15. The continuous covariate, age, was dichotomized by the mean of age. Based on the NPMLE in Figure 5.15, it is clear that the data may provide enough information for estimation only for quantiles below 0.2; thus, we focused on the 0.1, 0.15, and 0.20 quantiles.

The estimation procedure proposed in Section 3.2.2 was used for the 3 lower quantiles. We used block sizes $b = N^{\gamma}$ where $\gamma = \{1/3, 1/2, 2/3, 3/4, 0.8, 0.85, 0.9, 0.95\}$. Five-hundred subsamples for each of the block size were generated. Weights $\hat{\boldsymbol{\omega}}_{\boldsymbol{j}}$ were estimated and symmetric confidence intervals were constructed using the method described in Section 5.2.3. The results are summarized in Table 5.19.

Female sex and advanced age after 60 yrs old are associated with accelerated failure.

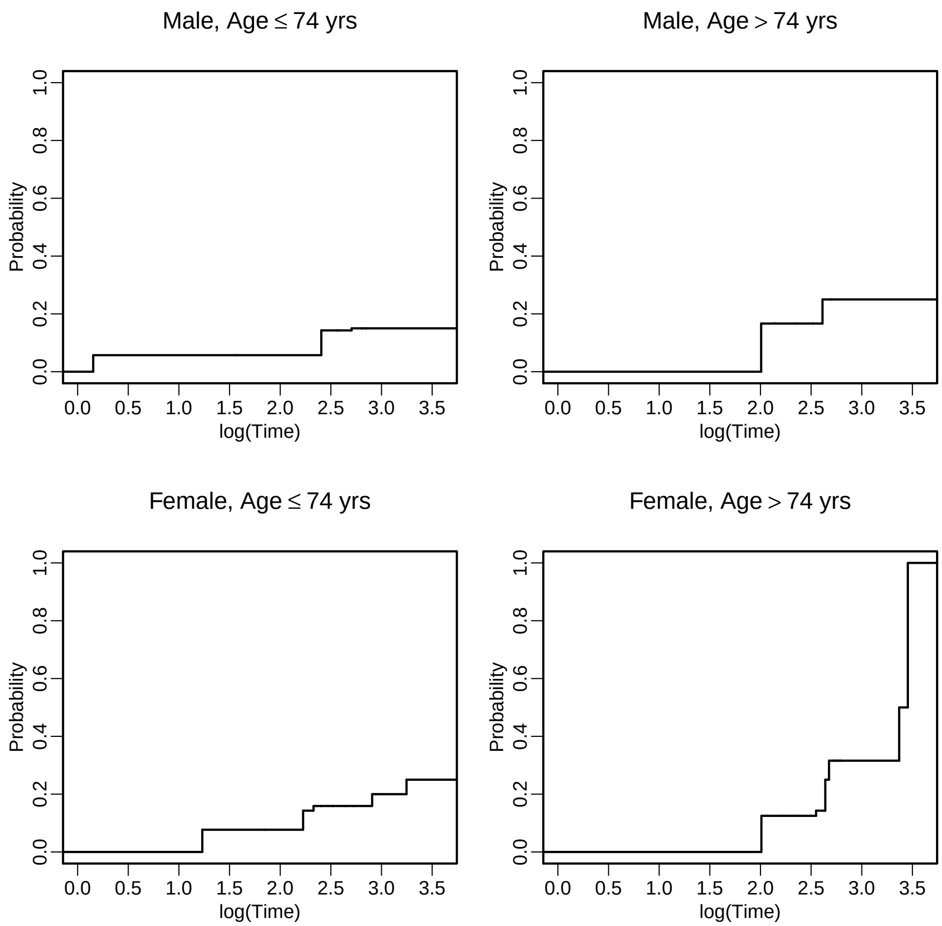

Figure 5.15: Nonparametric maximum likelihood estimator (NPMLE) of the (unobserved) logarithm of time to calcification distribution function stratified by covariates.

Advanced age prior to 60 years old; however, is associated with decelerated failure. The confidence interval for the weighted β is narrower than all the individual quantiles examined; unfortunately, it is not statistically significant at the 0.95 level.

The second dataset is a general interval-censored failure time data. The breast cosmesis data consists of 94 patients who were given either radiation therapy alone (RT, N=46) or radiation therapy plus adjuvant chemotherapy (RCT, N=48). Every 4 to 6 months, patients returned to the clinic for a check up and the cosmetic appearance of the patient was evaluated during the clinic visit. The outcome of interest is time to breast retraction, an undesired cometic effect. There are 38 patients who did not experience breast retraction during the study. A detailed description of this study can be found in Finkelstein and Wolfe (1985) and Finkelstein (1986).

The NPMLE (Wellner and Zhan 1997) of the (unobserved) failure time distribution function is shown in Figure 5.16. Based on the NPMLE in Figure 5.16, it is clear that the data may provide enough information for estimation only for quantiles below 0.5; thus, we focused on the 0.2, 0.4, and 0.5 quantiles. The 0.3 quantile was not used because we encountered convergence issues which is an indication that we may not have enough data to estimate that quantile.

The estimation procedure proposed in Section 4.2.2 was used for the 3 lower quantiles. We used block sizes $b = N^{\gamma}$ where $\gamma = \{1/3, 1/2, 2/3, 3/4, 0.8, 0.85, 0.9, 0.95\}$. Five-hundred subsamples for each of the block sizes were generated. Weights $\hat{\boldsymbol{\omega}}_{\boldsymbol{j}}$ were estimated and symmetric confidence intervals were constructed using the method described in Section 5.2.3. The results are summarized in Table 5.20. The weights calculated using the observed subsampling variance-covariance matrix are 0.282, 0.477, and 0.241 for 0.20, 0.40, and 0.50 quantiles, respectively.

We obtained a weighted estimate of 0.595 for RT vs. RCT with a confidence interval (0.164, 1.026) which indicated that patients receiving radiation therapy alone experienced

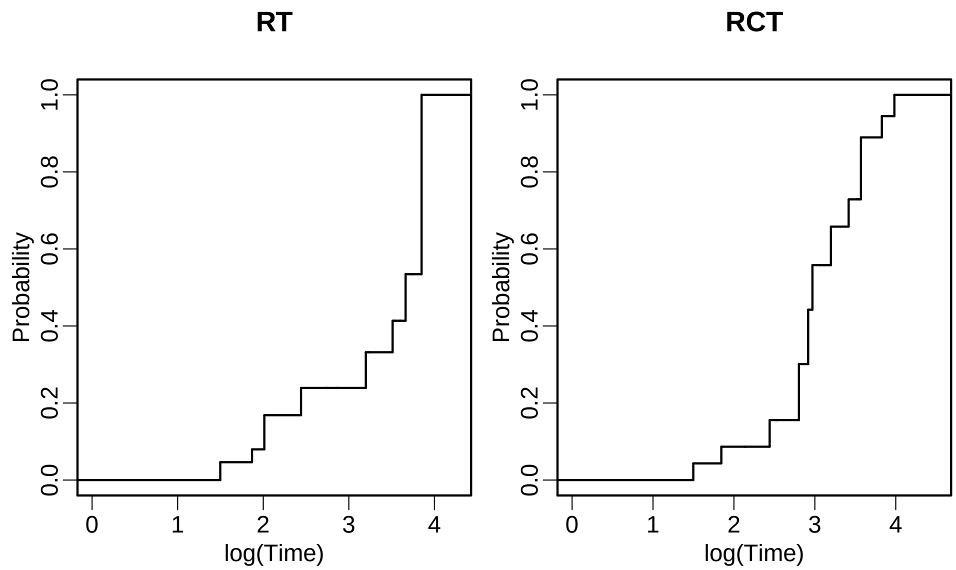

Figure 5.16: Nonparametric maximum likelihood estimator (NPMLE) of the (unobserved) logarithm of failure time distribution function for patients received radiation therapy alone (RT) and patients received radiation therapy plus adjuvant chemotherapy (RCT).

a significantly delay in breast retraction. The confidence interval length is 0.862 for the weighted estimate which is much narrower than the confidence interval length from the individual quantiles (1.596, 1.952, and 1.853 for the 0.2, 0.4, and 0.5 quantiles, respectively). The much narrower confidence interval length resulted in a significant point estimate for the weighted estimate while the estimates from the individual quantiles are not statistically significant.

5.6 Discussion

We have taken advantage of the quantile regression framework to provide a semiparametric method for accelerated failure time models using interval-censored data. We proposed to combine information across multiple quantiles to improve upon the efficiency of the estimators. Based on the simulation studies, the weighted estimators have ignorable bias. A subsampling procedure was used to produce confidence intervals for weighted estimators and it appeared to perform well. Combining a small number of quantiles, such as 3, can provide a marked gain in relative efficiency. The gain of using $\hat{\boldsymbol{\omega}}_{n,\epsilon}$ is slightly more than using a naive weighting scheme, i.e. taking the average across quantiles, but the naive weight method still has decent performance.

Table 5.8: Normal distribution, N=200

τ	**Truth**	**Bias**	**MSE**	**Std**	**CP**	**Length**
1/6	$\beta_1 = 1$	0.003	0.166	0.408	0.960	1.467
	$\beta_2 = 1$	-0.032	0.144	0.378	0.954	1.320
0.25	$\beta_1 = 1$	-0.010	0.137	0.370	0.953	1.346
	$\beta_2 = 1$	-0.017	0.102	0.318	0.953	1.172
1/3	$\beta_1 = 1$	-0.003	0.116	0.340	0.953	1.261
	$\beta_2 = 1$	-0.014	0.084	0.289	0.953	1.096
0.5	$\beta_1 = 1$	-0.000	0.105	0.324	0.961	1.197
	$\beta_2 = 1$	-0.011	0.081	0.285	0.952	1.044
2/3	$\beta_1 = 1$	-0.004	0.109	0.330	0.962	1.263
	$\beta_2 = 1$	-0.022	0.096	0.308	0.966	1.108
0.75	$\beta_1 = 1$	0.012	0.139	0.372	0.956	1.372
	$\beta_2 = 1$	-0.013	0.109	0.331	0.959	1.216
5/6	$\beta_1 = 1$	0.028	0.174	0.417	0.953	1.509
	$\beta_2 = 1$	-0.027	0.153	0.390	0.949	1.357
Equal3	$\beta_1 = 1$	0.001	0.053	0.230	0.969	0.928
	$\beta_2 = 1$	-0.014	0.042	0.204	0.959	0.839
Equal5	$\beta_1 = 1$	0.005	0.043	0.207	0.967	0.850
	$\beta_2 = 1$	-0.021	0.038	0.194	0.960	0.784
Equal7	$\beta_1 = 1$	0.004	0.043	0.209	0.958	0.842
	$\beta_2 = 1$	-0.020	0.037	0.192	0.961	0.772
Optimal3	$\beta_1 = 1$	-0.001	0.052	0.228	0.962	0.898
	$\beta_2 = 1$	-0.022	0.039	0.196	0.961	0.798
Optimal5	$\beta_1 = 1$	0.003	0.043	0.207	0.955	0.814
	$\beta_2 = 1$	-0.029	0.036	0.187	0.955	0.728
Optimal7	$\beta_1 = 1$	0.001	0.043	0.207	0.949	0.799
	$\beta_2 = 1$	-0.030	0.035	0.185	0.953	0.715

Table 5.9: Normal distribution, N=400

τ	**Truth**	**Bias**	**MSE**	**Std**	**CP**	**Length**
1/6	$\beta_1 = 1$	0.001	0.098	0.313	0.970	1.175
	$\beta_2 = 1$	-0.007	0.076	0.276	0.962	1.013
0.25	$\beta_1 = 1$	-0.002	0.077	0.278	0.969	1.050
	$\beta_2 = 1$	-0.001	0.057	0.238	0.962	0.915
1/3	$\beta_1 = 1$	-0.007	0.067	0.258	0.969	0.993
	$\beta_2 = 1$	-0.011	0.048	0.219	0.962	0.843
0.5	$\beta_1 = 1$	-0.002	0.067	0.258	0.960	0.952
	$\beta_2 = 1$	-0.011	0.048	0.219	0.966	0.819
2/3	$\beta_1 = 1$	-0.013	0.074	0.271	0.966	1.002
	$\beta_2 = 1$	-0.001	0.051	0.225	0.957	0.859
0.75	$\beta_1 = 1$	-0.007	0.085	0.292	0.961	1.083
	$\beta_2 = 1$	-0.005	0.063	0.251	0.972	0.931
5/6	$\beta_1 = 1$	0.005	0.111	0.333	0.970	1.232
	$\beta_2 = 1$	-0.007	0.080	0.282	0.966	1.063
Equal3	$\beta_1 = 1$	-0.004	0.030	0.174	0.963	0.705
	$\beta_2 = 1$	-0.006	0.022	0.148	0.973	0.625
Equal5	$\beta_1 = 1$	-0.003	0.024	0.155	0.965	0.633
	$\beta_2 = 1$	-0.007	0.018	0.133	0.974	0.571
Equal7	$\beta_1 = 1$	-0.004	0.023	0.153	0.960	0.623
	$\beta_2 = 1$	-0.006	0.017	0.132	0.970	0.561
Optimal3	$\beta_1 = 1$	-0.004	0.029	0.169	0.963	0.678
	$\beta_2 = 1$	-0.007	0.020	0.142	0.970	0.597
Optimal5	$\beta_1 = 1$	-0.005	0.023	0.151	0.959	0.608
	$\beta_2 = 1$	-0.010	0.017	0.130	0.960	0.539
Optimal7	$\beta_1 = 1$	-0.005	0.022	0.149	0.954	0.591
	$\beta_2 = 1$	-0.009	0.016	0.128	0.959	0.526

Table 5.10: Normal distribution, N=800

τ	**Truth**	**Bias**	**MSE**	**Std**	**CP**	**Length**
1/6	$\beta_1 = 1$	-0.007	0.059	0.243	0.965	0.931
	$\beta_2 = 1$	-0.012	0.044	0.209	0.971	0.806
0.25	$\beta_1 = 1$	-0.005	0.046	0.215	0.968	0.832
	$\beta_2 = 1$	-0.006	0.034	0.185	0.961	0.709
1/3	$\beta_1 = 1$	-0.003	0.044	0.210	0.968	0.781
	$\beta_2 = 1$	-0.006	0.030	0.173	0.961	0.659
0.5	$\beta_1 = 1$	-0.008	0.041	0.202	0.966	0.740
	$\beta_2 = 1$	0.001	0.028	0.168	0.961	0.634
2/3	$\beta_1 = 1$	0.003	0.044	0.210	0.951	0.781
	$\beta_2 = 1$	0.006	0.031	0.175	0.963	0.664
0.75	$\beta_1 = 1$	0.004	0.056	0.236	0.952	0.850
	$\beta_2 = 1$	0.001	0.033	0.180	0.975	0.713
5/6	$\beta_1 = 1$	0.013	0.068	0.260	0.963	0.952
	$\beta_2 = 1$	-0.004	0.047	0.216	0.969	0.832
Equal3	$\beta_1 = 1$	-0.003	0.019	0.136	0.967	0.536
	$\beta_2 = 1$	-0.002	0.013	0.112	0.970	0.468
Equal5	$\beta_1 = 1$	-0.001	0.013	0.113	0.973	0.477
	$\beta_2 = 1$	-0.003	0.009	0.097	0.974	0.421
Equal7	$\beta_1 = 1$	-0.001	0.012	0.111	0.969	0.464
	$\beta_2 = 1$	-0.003	0.009	0.094	0.971	0.412
Optimal3	$\beta_1 = 1$	-0.003	0.017	0.131	0.970	0.520
	$\beta_2 = 1$	-0.002	0.012	0.109	0.966	0.451
Optimal5	$\beta_1 = 1$	-0.001	0.012	0.112	0.970	0.458
	$\beta_2 = 1$	-0.004	0.009	0.096	0.963	0.401
Optimal7	$\beta_1 = 1$	-0.001	0.012	0.110	0.969	0.443
	$\beta_2 = 1$	-0.004	0.009	0.095	0.963	0.390

Table 5.11: Logistic distribution, N=200

τ	**Truth**	**Bias**	**MSE**	**Std**	**CP**	**Length**
1/6	$\beta_1 = 1$	-0.160	0.436	0.641	0.958	2.408
	$\beta_2 = 1$	-0.173	0.404	0.612	0.943	2.199
0.25	$\beta_1 = 1$	-0.059	0.303	0.547	0.948	1.952
	$\beta_2 = 1$	-0.109	0.261	0.500	0.929	1.709
1/3	$\beta_1 = 1$	-0.009	0.272	0.522	0.944	1.822
	$\beta_2 = 1$	-0.069	0.208	0.451	0.942	1.582
0.5	$\beta_1 = 1$	0.014	0.240	0.490	0.950	1.725
	$\beta_2 = 1$	-0.048	0.172	0.412	0.945	1.492
2/3	$\beta_1 = 1$	0.016	0.297	0.545	0.953	1.866
	$\beta_2 = 1$	-0.041	0.213	0.460	0.960	1.622
0.75	$\beta_1 = 1$	-0.005	0.318	0.564	0.956	2.035
	$\beta_2 = 1$	-0.027	0.256	0.505	0.958	1.771
5/6	$\beta_1 = 1$	0.002	0.420	0.649	0.962	2.302
	$\beta_2 = 1$	-0.012	0.342	0.585	0.961	2.064
Equal3	$\beta_1 = 1$	-0.017	0.124	0.351	0.954	1.349
	$\beta_2 = 1$	-0.061	0.101	0.312	0.953	1.182
Equal5	$\beta_1 = 1$	-0.027	0.117	0.341	0.949	1.259
	$\beta_2 = 1$	-0.069	0.092	0.296	0.936	1.127
Equal7	$\beta_1 = 1$	-0.029	0.111	0.332	0.945	1.243
	$\beta_2 = 1$	-0.069	0.092	0.296	0.929	1.108
Optimal3	$\beta_1 = 1$	-0.019	0.129	0.358	0.941	1.308
	$\beta_2 = 1$	-0.066	0.092	0.295	0.938	1.126
Optimal5	$\beta_1 = 1$	-0.017	0.114	0.337	0.939	1.203
	$\beta_2 = 1$	-0.069	0.080	0.275	0.935	1.041
Optimal7	$\beta_1 = 1$	-0.025	0.110	0.332	0.930	1.171
	$\beta_2 = 1$	-0.076	0.080	0.272	0.935	1.016

Table 5.12: Logistic distribution, N=400

τ	**Truth**	**Bias**	**MSE**	**Std**	**CP**	**Length**
1/6	$\beta_1 = 1$	-0.070	0.294	0.538	0.957	1.919
	$\beta_2 = 1$	-0.089	0.223	0.464	0.957	1.714
0.25	$\beta_1 = 1$	-0.018	0.181	0.425	0.959	1.592
	$\beta_2 = 1$	-0.030	0.150	0.387	0.951	1.364
1/3	$\beta_1 = 1$	0.014	0.150	0.388	0.961	1.441
	$\beta_2 = 1$	-0.025	0.116	0.340	0.952	1.224
0.5	$\beta_1 = 1$	0.003	0.147	0.383	0.959	1.409
	$\beta_2 = 1$	-0.026	0.096	0.308	0.962	1.174
2/3	$\beta_1 = 1$	0.029	0.167	0.408	0.956	1.513
	$\beta_2 = 1$	-0.023	0.114	0.336	0.954	1.269
0.75	$\beta_1 = 1$	0.027	0.196	0.442	0.961	1.632
	$\beta_2 = 1$	-0.037	0.149	0.385	0.950	1.406
5/6	$\beta_1 = 1$	0.018	0.278	0.527	0.961	1.889
	$\beta_2 = 1$	-0.036	0.217	0.465	0.961	1.678
Equal3	$\beta_1 = 1$	0.004	0.066	0.256	0.973	1.048
	$\beta_2 = 1$	-0.031	0.053	0.228	0.964	0.911
Equal5	$\beta_1 = 1$	-0.001	0.054	0.233	0.967	0.957
	$\beta_2 = 1$	-0.040	0.047	0.214	0.966	0.853
Equal7	$\beta_1 = 1$	0.001	0.052	0.227	0.964	0.938
	$\beta_2 = 1$	-0.038	0.046	0.212	0.956	0.835
Optimal3	$\beta_1 = 1$	-0.001	0.063	0.251	0.967	1.014
	$\beta_2 = 1$	-0.035	0.048	0.217	0.962	0.874
Optimal5	$\beta_1 = 1$	-0.002	0.053	0.229	0.960	0.910
	$\beta_2 = 1$	-0.042	0.041	0.198	0.958	0.790
Optimal7	$\beta_1 = 1$	-0.004	0.050	0.225	0.959	0.881
	$\beta_2 = 1$	-0.044	0.040	0.195	0.956	0.766

Table 5.13: Logistic distribution, N=800

τ	**Truth**	**Bias**	**MSE**	**Std**	**CP**	**Length**
1/6	$\beta_1 = 1$	-0.028	0.175	0.418	0.966	1.562
	$\beta_2 = 1$	-0.027	0.131	0.362	0.962	1.348
0.25	$\beta_1 = 1$	0.015	0.122	0.349	0.958	1.278
	$\beta_2 = 1$	-0.017	0.089	0.297	0.953	1.088
1/3	$\beta_1 = 1$	0.004	0.095	0.308	0.969	1.160
	$\beta_2 = 1$	-0.004	0.079	0.281	0.960	0.989
0.5	$\beta_1 = 1$	0.001	0.092	0.303	0.954	1.102
	$\beta_2 = 1$	0.003	0.056	0.238	0.971	0.938
2/3	$\beta_1 = 1$	0.003	0.106	0.325	0.966	1.198
	$\beta_2 = 1$	-0.008	0.075	0.273	0.964	1.023
0.75	$\beta_1 = 1$	0.007	0.123	0.351	0.976	1.333
	$\beta_2 = 1$	-0.003	0.098	0.313	0.963	1.160
5/6	$\beta_1 = 1$	0.006	0.203	0.450	0.965	1.581
	$\beta_2 = 1$	-0.009	0.134	0.366	0.972	1.378
Equal3	$\beta_1 = 1$	0.007	0.043	0.208	0.964	0.816
	$\beta_2 = 1$	-0.006	0.031	0.176	0.971	0.718
Equal5	$\beta_1 = 1$	-0.003	0.035	0.186	0.964	0.742
	$\beta_2 = 1$	-0.009	0.026	0.162	0.958	0.657
Equal7	$\beta_1 = 1$	0.001	0.031	0.177	0.964	0.723
	$\beta_2 = 1$	-0.009	0.025	0.159	0.950	0.642
Optimal3	$\beta_1 = 1$	0.003	0.042	0.204	0.956	0.790
	$\beta_2 = 1$	-0.006	0.028	0.167	0.971	0.686
Optimal5	$\beta_1 = 1$	-0.006	0.032	0.178	0.961	0.704
	$\beta_2 = 1$	-0.012	0.022	0.149	0.961	0.613
Optimal7	$\beta_1 = 1$	-0.005	0.030	0.174	0.954	0.680
	$\beta_2 = 1$	-0.015	0.021	0.145	0.949	0.593

Table 5.14: Extreme-value distribution, N=200

τ	Truth	Bias	MSE	Std	CP	Length
1/6	$\beta_1 = 1$	-0.061	0.407	0.635	0.943	2.187
	$\beta_2 = 1$	-0.120	0.312	0.544	0.948	1.950
0.25	$\beta_1 = 1$	-0.025	0.268	0.518	0.958	1.870
	$\beta_2 = 1$	-0.067	0.211	0.455	0.964	1.675
1/3	$\beta_1 = 1$	0.023	0.189	0.434	0.963	1.592
	$\beta_2 = 1$	-0.040	0.162	0.400	0.945	1.388
0.5	$\beta_1 = 1$	-0.017	0.128	0.358	0.962	1.365
	$\beta_2 = 1$	-0.049	0.107	0.323	0.952	1.187
2/3	$\beta_1 = 1$	-0.014	0.120	0.347	0.955	1.276
	$\beta_2 = 1$	-0.024	0.083	0.287	0.963	1.134
0.75	$\beta_1 = 1$	-0.018	0.124	0.352	0.957	1.319
	$\beta_2 = 1$	-0.021	0.097	0.311	0.964	1.187
5/6	$\beta_1 = 1$	-0.053	0.138	0.368	0.954	1.396
	$\beta_2 = 1$	-0.031	0.111	0.332	0.970	1.299
Equal3	$\beta_1 = 1$	-0.020	0.071	0.267	0.968	1.070
	$\beta_2 = 1$	-0.046	0.057	0.234	0.964	0.972
Equal5	$\beta_1 = 1$	-0.025	0.064	0.252	0.968	1.070
	$\beta_2 = 1$	-0.053	0.050	0.217	0.964	0.972
Equal7	$\beta_1 = 1$	-0.024	0.064	0.252	0.954	0.973
	$\beta_2 = 1$	-0.050	0.049	0.216	0.961	0.888
Optimal3	$\beta_1 = 1$	-0.019	0.064	0.252	0.963	1.004
	$\beta_2 = 1$	-0.045	0.049	0.216	0.970	0.902
Optimal5	$\beta_1 = 1$	-0.027	0.056	0.235	0.943	0.902
	$\beta_2 = 1$	-0.059	0.042	0.198	0.958	0.815
Optimal7	$\beta_1 = 1$	-0.027	0.056	0.234	0.951	0.890
	$\beta_2 = 1$	-0.051	0.041	0.196	0.956	0.803

Table 5.15: Extreme-value distribution, N=400

τ	Truth	Bias	MSE	Std	CP	Length
1/6	$\beta_1 = 1$	-0.011	0.243	0.493	0.963	1.802
	$\beta_2 = 1$	-0.054	0.202	0.446	0.956	1.583
0.25	$\beta_1 = 1$	0.018	0.176	0.419	0.956	1.526
	$\beta_2 = 1$	-0.031	0.130	0.360	0.956	1.312
1/3	$\beta_1 = 1$	0.022	0.131	0.362	0.950	1.278
	$\beta_2 = 1$	-0.019	0.091	0.301	0.949	1.097
0.5	$\beta_1 = 1$	0.010	0.084	0.290	0.955	1.087
	$\beta_2 = 1$	-0.019	0.063	0.250	0.967	0.934
2/3	$\beta_1 = 1$	-0.003	0.069	0.263	0.962	0.999
	$\beta_2 = 1$	-0.014	0.047	0.217	0.959	0.860
0.75	$\beta_1 = 1$	-0.013	0.068	0.261	0.968	1.017
	$\beta_2 = 1$	-0.027	0.055	0.233	0.970	0.896
5/6	$\beta_1 = 1$	-0.032	0.084	0.288	0.958	1.078
	$\beta_2 = 1$	-0.038	0.064	0.250	0.975	0.976
Equal3	$\beta_1 = 1$	0.005	0.041	0.202	0.962	0.820
	$\beta_2 = 1$	-0.026	0.034	0.184	0.957	0.719
Equal5	$\beta_1 = 1$	-0.003	0.034	0.184	0.970	0.742
	$\beta_2 = 1$	-0.029	0.029	0.168	0.958	0.666
Equal7	$\beta_1 = 1$	-0.002	0.033	0.181	0.959	0.729
	$\beta_2 = 1$	-0.029	0.029	0.167	0.957	0.654
Optimal3	$\beta_1 = 1$	-0.001	0.034	0.184	0.961	0.764
	$\beta_2 = 1$	-0.028	0.029	0.167	0.956	0.667
Optimal5	$\beta_1 = 1$	-0.012	0.028	0.167	0.960	0.678
	$\beta_2 = 1$	-0.031	0.024	0.150	0.953	0.601
Optimal7	$\beta_1 = 1$	-0.013	0.028	0.166	0.954	0.661
	$\beta_2 = 1$	-0.033	0.023	0.150	0.948	0.588

Table 5.16: Extreme-value distribution, N=800

τ	Truth	Bias	MSE	Std	CP	Length
1/6	$\beta_1 = 1$	0.019	0.154	0.393	0.963	1.462
	$\beta_2 = 1$	-0.027	0.128	0.357	0.960	1.258
0.25	$\beta_1 = 1$	0.005	0.101	0.318	0.969	1.225
	$\beta_2 = 1$	-0.017	0.079	0.280	0.966	1.048
1/3	$\beta_1 = 1$	0.015	0.075	0.274	0.969	1.018
	$\beta_2 = 1$	-0.011	0.053	0.230	0.961	0.867
0.5	$\beta_1 = 1$	0.009	0.049	0.221	0.962	0.847
	$\beta_2 = 1$	0.004	0.036	0.190	0.961	0.716
2/3	$\beta_1 = 1$	0.009	0.040	0.201	0.961	0.780
	$\beta_2 = 1$	-0.013	0.033	0.175	0.959	0.671
0.75	$\beta_1 = 1$	0.002	0.042	0.205	0.968	0.788
	$\beta_2 = 1$	-0.013	0.030	0.178	0.964	0.677
5/6	$\beta_1 = 1$	-0.008	0.049	0.220	0.965	0.839
	$\beta_2 = 1$	-0.013	0.037	0.193	0.968	0.740
Equal3	$\beta_1 = 1$	0.005	0.023	0.151	0.972	0.630
	$\beta_2 = 1$	-0.009	0.019	0.136	0.964	0.547
Equal5	$\beta_1 = 1$	0.009	0.018	0.134	0.968	0.560
	$\beta_2 = 1$	-0.012	0.015	0.123	0.963	0.497
Equal7	$\beta_1 = 1$	0.007	0.017	0.130	0.974	0.548
	$\beta_2 = 1$	-0.013	0.018	0.121	0.955	0.482
Optimal3	$\beta_1 = 1$	0.001	0.019	0.139	0.973	0.580
	$\beta_2 = 1$	-0.011	0.015	0.124	0.967	0.502
Optimal5	$\beta_1 = 1$	-0.002	0.015	0.122	0.963	0.506
	$\beta_2 = 1$	-0.016	0.012	0.109	0.961	0.448
Optimal7	$\beta_1 = 1$	-0.003	0.014	0.119	0.967	0.491
	$\beta_2 = 1$	-0.018	0.012	0.107	0.958	0.434

Table 5.17: Relative efficiency

		β_1			β_2		
ξ	**Weight**	$N = 200$	$N = 400$	$N = 800$	$N = 200$	$N = 400$	$N = 800$
Simulation 1	Equal3	1.9836	2.2115	2.1862	1.9424	2.1955	2.2586
Normal	Equal5	2.4471	2.7788	3.1769	2.1373	2.6973	3.0158
	Equal7	2.4114	2.8685	3.3223	2.1952	2.7670	3.1872
	Optimal3	2.0101	2.3422	2.3621	2.0984	2.3891	2.3725
	Optimal5	2.4559	2.9367	3.2689	2.2721	2.8223	3.0791
	Optimal7	2.4540	2.9953	3.3664	2.3157	2.9174	3.1468
Simulation 2	Equal3	1.9424	2.2376	2.1281	1.6966	1.8040	1.8137
Logistic	Equal5	2.0566	2.7077	2.6446	1.8562	2.0241	2.1358
	Equal7	2.1625	2.8478	2.9240	1.8657	2.0702	2.2285
	Optimal3	1.8638	2.3346	2.2075	1.8747	1.9917	2.0158
	Optimal5	2.1078	2.7868	2.9062	2.1375	2.3406	2.5203
	Optimal7	2.1730	2.9064	3.0159	2.1517	2.3969	2.6439
Simulation 3	Equal3	1.6844	1.6876	1.7808	1.4611	1.3784	1.6548
Extreme-value	Equal5	1.8840	2.0442	2.2294	1.6634	1.6359	1.9990
	Equal7	1.8832	2.1170	2.4054	1.6928	1.6540	2.0610
	Optimal3	1.8896	2.0387	2.0952	1.7098	1.6438	1.9833
	Optimal5	2.1504	2.4681	2.7084	1.9996	2.0094	2.5258
	Optimal7	2.1665	2.5012	2.8335	2.0254	2.0180	2.5742

In *Simulation 1* and *Simulation 2*, MSE of $\hat{\beta}(\tau = 0.5)$ was used as the numerator. In *Simulation 3*, MSE of $\hat{\beta}(\tau = 2/3)$ was used as the numerator.

Table 5.18: Confidence Interval Length Ratio

		β_1			β_2		
ξ	**Weight**	$N = 200$	$N = 400$	$N = 800$	$N = 200$	$N = 400$	$N = 800$
Simulation 1	Equal3	1.2892	1.3511	1.3804	1.2442	1.3111	1.3542
Normal	Equal5	1.4075	1.5042	1.5506	1.3321	1.4330	1.5055
	Equal7	1.4203	1.5293	1.5924	1.3519	1.4585	1.5387
	Optimal3	1.3327	1.4039	1.4218	1.3085	1.3720	1.4034
	Optimal5	1.4699	1.5658	1.6150	1.4330	1.5208	1.5782
	Optimal7	1.4982	1.6097	1.6693	1.4603	1.5567	1.6248
Simulation 2	Equal3	1.2784	1.3450	1.3503	1.2626	1.2474	1.3075
Logistic	Equal5	1.3699	1.4723	1.4846	1.3237	1.3768	1.4285
	Equal7	1.3883	1.5029	1.5259	1.3470	1.4060	1.4617
	Optimal3	1.3187	1.3897	1.3939	1.3253	1.3439	1.3679
	Optimal5	1.4336	1.5485	1.5656	1.4331	1.4866	1.5311
	Optimal7	1.4727	1.5998	1.6199	1.4690	1.5334	1.5820
Simulation 3	Equal3	1.1921	1.2179	1.2377	1.1667	1.1965	1.2273
Extreme-value	Equal5	1.1921	1.3461	1.3914	1.1667	1.2914	1.3504
	Equal7	1.3116	1.3714	1.4216	1.2777	1.3152	1.3827
	Optimal3	1.2710	1.3084	1.3439	1.2575	1.2902	1.3383
	Optimal5	1.4138	1.4737	1.5416	1.3909	1.4319	1.5010
	Optimal7	1.4332	1.5128	1.5891	1.4129	1.4631	1.5484

Table 5.19: Results of Calcification Data Analysis

τ	Covariate	Point Estimate	Confidence Interval	C.I. Length
0.10	Intercept	3.003	(1.355, 4.651)	3.296
	Female vs. Male	-0.224	(-0.905, 0.457)	1.363
	Age (per 10 yrs) prior to 60 yrs	0.832	(-3.748, 5.413)	9.161
	Age (per 10 yrs) after to 60 yrs	-0.469	(-1.604, 0.667)	2.271
0.15	Intercept	3.133	(0.211, 6.055)	5.843
	Female vs. Male	0.388	(-3.115, 3.891)	7.006
	Age (per 10 yrs) prior to 60 yrs	1.416	(-4.202, 7.034)	11.235
	Age (per 10 yrs) after to 60 yrs	-0.576	(-1.568, 0.417)	1.985
0.20	Intercept	4.061	(0.831, 7.291)	6.462
	Female vs. Male	-0.851	(-3.778, 2.077)	5.855
	Age (per 10 yrs) prior to 60 yrs	2.177	(-2.146, 6.500)	8.465
	Age (per 10 yrs) after to 60 yrs	-0.321	(-1.498, 0.856)	2.354
Weighted	Female vs. Male	-0.224	(-0.902, 0.454)	1.355
	Age (per 10 yrs) prior to 60 yrs	1.159	(-2.943, 5.260)	8.202
	Age (per 10 yrs) after to 60 yrs	-0.463	(-1.265, 0.340)	1.605

Table 5.20: Results of Breast Cosmesis Data Analysis

		RT vs. RCT
τ	Intercept	Point Estimate (Confidence Interval)
0.2	1.096	0.392 (-0.406, 1.190)
0.4	1.476	0.569 (-0.407, 1.545)
0.5	3.571	0.884 (-0.042, 1.811)
Weighted	—	0.595 (0.164, 1.026)

RT: radiation therapy alone and RCT: radiation therapy plus adjuvant chemotherapy.

CHAPTER 6: SUMMARY AND FUTURE RESEARCH

In failure time analysis, the data are rarely normally distributed and tend to be right-skewed. Commonly used methods for failure time analysis describe covariate effects through relative risk which portrays the odds of event occurrence but does not convey the information in the original time scale. Quantile regression models can describe the association at different quantiles, providing more detailed relationships when data is skewed and when heterogeneity is present. The estimate from a quantile regression can be interpreted as the direct effect on the response variable which is appealing for failure time analysis. Many methods have been proposed for right-censored failure time data, but there is limited method available to apply quantile regression on interval-censored failure time data. This dissertation innovates conditional quantile regression models to analyze interval-censored failure time data.

The method for Case I interval-censored data, also known as current status data, has been developed. The asymptotic statistical properties of the estimator have been proven and the small sample performances have been demonstrated via simulation studies. The proposed method has also been applied on the females' data from "The Voluntary HIV-1 Counseling and Testing Efficacy Study Group".

We extended the method developed for current status data to Case II interval-censored data. We found that a weighted support vector machine (weighted SVM) method produces a reasonable initial value for the estimation routine. Using weighted SVM to produce initial values is an improvement upon the method we used for current status data which was a coarse grid search. We also rigourously proved the asymptotic properties of our estimator. The numerical performances have been shown using simulation studies. Our method has been applied to the African American data from the Atherosclerosis Risk in Communities (ARIC) study.

Under the accelerated failure time model, the quantiles should be simply a shift at the intercept. Taking advantage of this fact, we used our method to estimate the covariate effects under the accelerated failure time model for interval-censored data. The estimates from different quantiles are combined to increase efficiency. This provided a semiparametric approach to analyze accelerated failure time model with interval-censored data.

It will be advantageous to construct a formal test to test the validity of using the AFT model for a given dataset. The test can be used to check the iid assumption of the error term prior to performing a data analysis. If the test indicates a heteroscedastic error then an AFT model should not be used and alternative methods should be employed.

It might be of interest of investigate more on the efficiency gain of combining information across quantiles. One open question is that whether combining quantiles can improve the cube-root convergence rate to close to the typical $\sqrt{n}$ convergence rate. The other question is whether it is possible to construct efficient estimators in this manner. These are interesting statistical questions and warrant further investigations.

REFERENCES

Abrevaya, J. (2002), "The effects of demographics and maternal behavior on the distribution of birth outcomes," in *Economic Applications of Quantile Regression*, Springer, pp. 247–257.

Abrevaya, J. and Huang, J. (2005), "On the Bootstrap of the Maximum Score Estimator," *Econometrica*, 73, 1175–1204.

AHA Writng Group (2010), "Executive Summary: Heart Disease and Stroke StatisticsŮ2010 Update: A Report From the American Heart Association," *Circulation*, 121, 948–954.

Ash, R. and Doléans-Dade, C. (2000), *Probability and Measure Theory*, Harcourt/Academic Press.

Banerjee, M. and McKeague, I. W. (2007), "Confidence Sets for Split Points in Decision Trees," *The Annals of Statistics*, 35, 543–574.

Banerjee, M. and Wellner, J. A. (2005), "Confidence Intervals for Current Status Data," *Scandinavian Journal of Statistics*, 32, 405–424.

Barrodale, I. and Roberts, F. D. (1973), "An improved algorithm for discrete l_1 linear approximation," *SIAM Journal on Numerical Analysis*, 10, 839–848.

Betensky, R. A., Lindsey, J. C., Ryan, L. M., and Wand, M. (2002), "A local likelihood proportional hazards model for interval censored data," *Statistics in Medicine*, 21, 263–275.

Betensky, R. A., Rabinowitz, D., and Tsiatis, A. A. (2001), "Computationally simple accelerated failure time regression for interval censored data," *Biometrika*, 88, 703–711.

Bofingeb, E. (1975), "ESTIMATION OF A DENSITY FUNCTION USING ORDER STATISTICS1," *Australian Journal of Statistics*, 17, 1–7.

Buchinsky, M. and Hahn, J. (1998), "An Alternative Estimator for the Censored Quantile Regression Model," *Econometrica*, 66, 653–671.

Cade, B. S. and Noon, B. R. (2003), "A Gentle Introduction to Quantile Regression for Ecologists," *Frontiers in Ecology and the Environment*, 1, 412–420.

Chatterjee, S., Bose, A., et al. (2005), "Generalized bootstrap for estimating equations," *The Annals of Statistics*, 33, 414–436.

Chen, C. (2007), "A finite smoothing algorithm for quantile regression," *Journal of Computational and Graphical Statistics*, 16.

Cheng, G. and Wang, X. (2011), "Semiparametric additive transformation model under current status data," *Electronic Journal of Statistics*, 5, 1735–1764.

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

Coates, T. J., Grinstead, O. A., Gregorich, S. E., Sweat, M. D., C., K. M., Sangiwa, G., Balmer, D., and Furlonge, C. (2000), "Efficacy of voluntary HIV-1 counselling and testing in individuals and couples in Kenya, Tanzania, and Trinidad: a randomised trial," *The Lancet*, 356, 103 – 112.

Delgado, M. A., Rodriguez-Poo, J. M., and Wolf, M. (2001), "Subsampling inference in cube root asymptotics with an application to ManskiŠs maximum score estimator," *Economics Letters*, 73, 241 – 250.

Dunham, J. B., Cade, B. S., and Terrell, J. W. (2002), "Influences of spatial and temporal variation on fish-habitat relationships defined by regression quantiles," *Transactions of the American Fisheries Society*, 131, 86–98.

Feng, X., He, X., and Hu, J. (2011), "Wild bootstrap for quantile regression," *Biometrika*, 98, 995–999.

Finkelstein, D. M. (1986), "A proportional hazards model for interval-censored failure time data," *Biometrics*, 845–854.

Finkelstein, D. M. and Wolfe, R. A. (1985), "A semiparametric model for regression analysis of interval-censored failure time data," *Biometrics*, 933–945.

Gentleman, R. and Vandal, A. (2011), *Icens: NPMLE for Censored and Truncated Data*, r package version 1.24.0.

Gutenbrunner, C. and Jurecková, J. (1992), "Regression rank scores and regression quantiles," *The Annals of Statistics*, 305–330.

Gutenbrunner, C., Jurečková, J., Koenker, R., and Portnoy, S. (1993), "Tests of linear hypotheses based on regression rank scores," *Journaltitle of Nonparametric Statistics*, 2, 307–331.

Hájek, J., Šidák, Z., and Sen, P. K. (1967), *Theory of rank tests*, Academic press New York.

Hall, P. and Sheather, S. J. (1988), "On the distribution of a studentized quantile," *Journal of the Royal Statistical Society. Series B (Methodological)*, 381–391.

He, S. and Yang, G. L. (2003), "Estimation of regression parameters with left truncated data," *Journal of statistical planning and inference*, 117, 99–122.

He, X. and Hu, F. (2002), "Markov chain marginal bootstrap," *Journal of the American Statistical Association*, 97, 783–795.

Hendricks, W. and Koenker, R. (1992), "Hierarchical spline models for conditional quantiles and the demand for electricity," *Journal of the American Statistical Association*, 87, 58–68.

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

Horowitz, J. (2010), *Semiparametric and Nonparametric Methods in Econometrics*, Springer Series in Statistics, Springer New York.

Huang, J. (1996), "Efficient estimation for the proportional hazards model with interval censoring," *The Annals of Statistics*, 24, 540–568.

Huang, J. and Rossini, A. (1997), "Sieve estimation for the proportional-odds failure-time regression model with interval censoring," *Journal of the American Statistical Association*, 92, 960–967.

Huang, J. and Wellner, J. A. (1997), "Interval censored survival data: a review of recent progress," in *Proceedings of the First Seattle Symposium in Biostatistics*, Springer, pp. 123–169.

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

Huber, P. J. (1973), "Robust regression: asymptotics, conjectures and Monte Carlo," *The Annals of Statistics*, 799–821.

Ji, S., Peng, L., Cheng, Y., and Lai, H. (2012), "Quantile regression for doubly censored data," *Biometrics*, 68, 101–112.

Kamenga, M., Sweat, M., De Zoysa, I., Dallabetta, G., and Coates, T. J. (2000), "The Voluntary HIV-1 Counseling and Testing Efficacy Study: Design and Methods," *AIDS and Behavior*, 4, 5–14.

Kim, J. and Pollard, D. (1990), "Cube Root Asymptotics," *The Annals of Statistics*, 18, 191–219.

Kim, Y.-J., Cho, H., Kim, J., and Jhun, M. (2010), "Median regression model with interval censored data," *Biometrical Journal*, 52, 201–208.

Koenker, R. (2005), *Quantile Regression*, Econometric Society Monographs, Cambridge University Press.

— (2008), "Censored quantile regression redux," *Journal of Statistical Software*, 27, 1–24.

— (2013), *quantreg: Quantile Regression*, r package version 4.97.

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

Koenker, R. and d'Orey, V. (1994), "Remark AS R92: A remark on algorithm AS 229: Computing dual regression quantiles and regression rank scores," *Applied Statistics*, 410–414.

Koenker, R. and Hallock, K. F. (2001), "Quantile Regression," *Journal of Economic Perspectives*, 15, 143–156.

Koenker, R. and Machado, J. A. (1999), "Goodness of fit and related inference processes for quantile regression," *Journal of the american statistical association*, 94, 1296–1310.

Koenker, R. W. and d'Orey, V. (1987), "Algorithm AS 229: Computing regression quantiles," *Applied Statistics*, 383–393.

Leng, C. and Tong, X. (2013), "A quantile regression estimator for censored data," *Bernoulli*, 19, 344–361.

Lin, D., Oakes, D., and Ying, Z. (1998), "Additive hazards regression with current status data," *Biometrika*, 85, 289–298.

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

Lustig, I. J., Marsten, R. E., and Shanno, D. F. (1992), "On implementing Mehrotra's predictor-corrector interior-point method for linear programming," *SIAM Journal on Optimization*, 2, 435–449.

Madsen, K. and Nielsen, H. B. (1993), “A Finite Smoothing Algorithm for Linear l_1 Estimation,” *SIAM Journal on Optimization*, 3, 223–235.

Manski, C. F. (1975), “Maximum score estimation of the stochastic utility model of choice,” *Journal of Econometrics*, 3, 205 – 228.

Martinussen, T. and Scheike, T. H. (2002), “Efficient estimation in additive hazards regression with current status data,” *Biometrika*, 89, 649–658.

McKeague, I. W., Subramanian, S., and Sun, Y. (2001), “Median regression and the missing information principle,” *Journal of nonparametric statistics*, 13, 709–727.

Mehrotra, S. (1992), “On the implementation of a primal-dual interior point method,” *SIAM Journal on optimization*, 2, 575–601.

Mohammadi, L. and Van De Geer, S. (2005), “Asymptotics in empirical risk minimization,” *The Journal of Machine Learning Research*, 6, 2027–2047.

Murphy, S. A., Van der Vaart, A., and Wellner, J. (1999), “Current status regression,” *Mathematical methods of statistics*, 8, 407–425.

Newey, W. K. and McFadden, D. (1994), “Chapter 36 Large sample estimation and hypothesis testing,” in *Handbook of Econometrics*, eds. Engle, R. F. and McFadden, D. L., Elsevier, vol. 4 of *Handbook of Econometrics*, pp. 2111 – 2245.

Parzen, E. (1979), “Nonparametric statistical data modeling,” *Journal of the American statistical association*, 74, 105–121.

Parzen, M., Wei, L., and Ying, Z. (1994), “A resampling method based on pivotal estimating functions,” *Biometrika*, 81, 341–350.

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

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

Politis, D., Romano, J., and Wolf, M. (1999), *Subsampling*, Springer series in statistics, Springer.

Portnoy, S. (2003), “Censored Regression Quantiles,” *Journal of the American Statistical Association*, 98, 1001–1012.

Portnoy, S., Koenker, R., et al. (1997), "The Gaussian hare and the Laplacian tortoise: computability of squared-error versus absolute-error estimators," *Statistical Science*, 12, 279–300.

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

— (1986), "Censored regression quantiles," *Journal of Econometrics*, 32, 143 – 155.

— (1991), "Estimation of monotonic regression models under quantile restrictions," *Nonparametric and semiparametric methods in Econometrics,(Cambridge University Press, New York, NY)*, 357–384.

— (1994), "Estimation of semiparametric models," *Handbook of econometrics*, 4, 2443–2521.

Rabinowitz, D., Betensky, R. A., and Tsiatis, A. A. (2000), "Using conditional logistic regression to fit proportional odds models to interval censored data," *Biometrics*, 56, 511–518.

Rabinowitz, D., Tsiatis, A., and Aragon, J. (1995), "Regression with interval-censored data," *Biometrika*, 82, 501–513.

Rao, C. R. and Zhao, L. (1992), "Approximation to the distribution of M-estimates in linear models by randomly weighted bootstrap," *Sankhyā: The Indian Journal of Statistics, Series A*, 323–331.

Rossini, A. and Tsiatis, A. (1996), "A semiparametric proportional odds regression model for the analysis of current status data," *Journal of the American Statistical Association*, 91, 713–721.

Sheather, S. and Maritz, J. (1983), "An estimate of the asymptotic standard error of the sample median," *Australian Journal of Statistics*, 25, 109–122.

Shen, P.-S. (2014), "A weighted quantile regression for left-truncated and right-censored data," *Journal of Statistical Computation and Simulation*, 84, 596–604.

Shen, X. (1998), "Propotional odds regression and sieve maximum likelihood estimation," *Biometrika*, 85, 165–177.

— (2000), "Linear regression with current status data," *Journal of the American Statistical Association*, 95, 842–852.

Siddiqui, M. M. (1960), "Distribution of quantiles in samples from a bivariate population," *J. Res. Nat. Bur. Standards B*, 64, 145–150.

Sun, J. (2007), *The Statistical Analysis of Interval-censored Failure Time Data*, Statistics for Biology and Health, Springer.

Sun, X., Peng, L., Huang, Y., and Lai, H. J. (2015), "Generalizing Quantile Regression for Counting Processes with Applications to Recurrent Events," *Journal of the American Statistical Association.*

The ARIC Investigators (1989), "The Atherosclerosis Risk in Communities (ARIC) study: design and objectives," *American Journal of Epidemiology*, 129, 687–702.

Tian, L. and Cai, T. (2006), "On the accelerated failure time model for current status and interval censored data," *Biometrika*, 93, 329–342.

Tobin, J. (1958), "Estimation of relationships for limited dependent variables," *Econometrica: journal of the Econometric Society*, 24–36.

Tukey, J. W. (1965), "Which part of the sample contains the information?" *Proceedings of the National Academy of Sciences of the United States of America*, 53, 127.

Turnbull, B. W. (1976), "The empirical distribution function with arbitrarily grouped, censored and truncated data," *Journal of the Royal Statistical Society. Series B (Methodological)*, 290–295.

van der Vaart, A. and Wellner, J. (1996), *Weak Convergence and Empirical Processes: With Applications to Statistics*, Springer Series in Statistics, Springer.

Wang, H. J. and Wang, L. (2009), "Locally Weighted Censored Quantile Regression," *Journal of the American Statistical Association*, 104, 1117–1128.

Wellner, J. A. and Zhan, Y. (1997), "A Hybrid Algorithm for Computation of the Nonparametric Maximum Likelihood Estimator from Censored Data," *Journal of the American Statistical Association*, 92, 945–959.

Welsh, A. (1988), "Asymptotically efficient estimation of the sparsity function at a point," *Statistics & probability letters*, 6, 427–432.

Wey, A., Wang, L., and Rudser, K. (2014), "Censored quantile regression with recursive partitioning-based weights," *Biostatistics*, 1, 170–181.

Wyatt, S. B., Akylbekova, E. L., Wofford, M. R., Coady, S. A., Walker, E. R., Andrew,

M. E., Keahey, W. J., Taylor, H. A., and Jones, D. W. (2008), "Prevalence, awareness, treatment, and control of hypertension in the Jackson Heart Study," *Hypertension*, 51, 650–656.

Xue, H., Lam, K. F., and Li, G. (2004), "Sieve Maximum Likelihood Estimator for Semiparametric Regression Models With Current Status Data," *Journal of the American Statistical Association*, 99, 346–356.

Yang, S. (1999), "Censored Median Regression Using Weighted Empirical Survival and Hazard Functions," *Journal of the American Statistical Association*, 94, 137–145.

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

Yu, A. K. F., Kwan, K. Y. W., Chan, D. H. Y., and Fong, D. Y. T. (2001), "Clinical features of 46 eyes with calcified hydrogel intraocular lenses," *J Cataract Refract Surg*, 27, 1596–1606.

Yuille, A. L. and Rangarajan, A. (2003), "The concave-convex procedure," *Neural Comput.*, 15, 915–936.

Zhou, W. (2011), "A weighted quantile regression for randomly truncated data," *Computational Statistics & Data Analysis*, 55, 554–566.