Statistica Sinica **24** (2014), 221-249
doi:http://dx.doi.org/10.5705/ss.2012.089

CENSORED QUANTILE REGRESSION VIA BOX-COX TRANSFORMATION UNDER CONDITIONAL INDEPENDENCE

Chenlei Leng and Xingwei Tong

University of Warwick and Beijing Normal University

Abstract: We propose a new quantile regression model when data are subject to censoring. Our model does not require any global linearity assumption, or independence of the covariates and the censoring time. We develop a class of power-transformed quantile regression models such that the transformed survival time can be better characterized by linear regression quantiles. Consistency and asymptotic normality of the resulting estimators are shown. A re-sampling based approach is proposed for statistical inference. Empirically, the new estimator is shown to outperform its competitors under conditional independence, and perform similarly under unconditional independence. The proposed method is illustrated with a data analysis.

Key words and phrases: Accelerated failure time model, Box-Cox transformation, censored quantile regression, empirical process, estimating equation, martingale.

1. Introduction

Consider the accelerated failure time (AFT) model (Kalbfleisch and Prentice (2002))

$$\log(T) = \beta_0^T Z + \varepsilon,$$

where T is the survival time, Z is the covariate vector and the errors are i.i.d. with mean zero. This model provides a direct interpretation of the (log) survival time using the covariates and is a useful alternative to the Cox model (Kalbfleisch and Prentice (2002)).

The usual AFT model excludes error heterogeneity, and cannot be used to quantify the covariate effects in lower or higher quantiles of the survival time if such heterogeneity exists (Koenker and Geling (2001)). Recent years have seen quantile regression as a complement to the classical conditional mean model (Koenker and Bassett (1978); Koenker (2005)). When data are subject to censoring, quantile regression has emerged as a flexible method that is able to assess the distributional information of the survival time based on covariates. Powell (1986) studied censored quantile regression for fixed censoring.

Ying, Jung, and Wei (1995) proposed a novel median regression model under random censoring. Yang (1999) used an empirically covariate-weighted cumulative hazard function to study median regression. Koenker and Geling (2001) explored the usefulness of this model by analyzing a medfly longevity dataset. Portnoy (2003) proposed an innovative redistribution-of-mass method that generalized the classical Kaplan-Meier estimator. Peng and Huang (2008) proposed a censored quantile regression approach using martingale-based estimating equations, though their model relies on a strong global linearity assumption. It is thus of interest to develop censored quantile regression methods that are free of this assumption. Wang and Wang (2009) proposed a locally weighted censored quantile regression extending Portnoy's method with less restrictive assumptions. More recently, Huang (2010) proposed a censored quantile regression method based on estimating integral equations.

The logarithmic transformation in the AFT model is often made for convenience. Transformation quantile regression models are flexible and can accommodate a wide variety of models. To identify a proper scale of the survival time that is linearly related to the covariates, the Box-Cox transformation (Box and Cox (1964)) is an attractive option. This model has been studied by Cai, Tian, and Wei (2005) for the conditional mean in survival analysis, by Mu and He (2007) for the quantile regression with no censoring, and by Yin, Zeng, and Li (2008) in censored quantile regression.

We propose a censored quantile regression estimator motivated by the martingale process associated with the counting process. The martingale process provides a general framework for studying asymptotic properties and inferential procedures. Our model does not require a global linearity assumption. We only require the conditional independence of the survival and the censoring time given the covariates.

Using a locally weighted Kaplan-Meier estimator, our approach can be easily implemented with existing quantile regression code (Koenker (2005)). We use our estimator in conjunction with the Box-Cox transformation to identify a proper transformation of the survival time such that the conditional quantiles after this transformation are linearly related to covariates. We study the estimator when multiple continuous covariates are present, extending substantially the univariate results in Wang and Wang (2009). In addition, we propose to use a power transformation such that the conditional quantile of the transformed response is linearly related to covariates.

The rest of the paper is organized as follows. In Section 2, we use a set of mar tingale-based estimation equations to motivate our new quantile regression estimator. We discuss an efficient non-iterative algorithm and study how the Box-Cox transformation can be incorporated. We present the consistency and asymptotic results in Section 3, separately for the case where the transformation parameter is unknown and for the case this parameter needs to be estimated. Extensive numerical simulations reported in Section 4 show that the proposed method works well and compare favorably with other methods. We illustrate the usefulness of the method via analysis of an HMO dataset in Section 5. Concluding remarks are given in Section 6. All proofs are relegated to the Appendix.

2. Martingale-based Estimation

Let T be the survival time and C the censoring time. The data are (Y_i, δ_i, Z_i), $i = 1, \ldots, n$, where $Y_i = \min(T_i, C_i)$ is the observed failure time, $\delta_i = I(T_i \leq C_i)$ is the censoring indicator; and $Z_i = (Z_{i0}, Z_{i1}, \ldots, Z_{ip})' \in \mathbb{R}^{p+1}$ is a $(p+1)$-dimensional covariate with Z_{i0} as the intercept. We discuss the censored quantile regression assuming the logarithm transformation. We then discuss how the formulation can be extended to incorporate an unknown transformation parameter.

2.1. Censored quantile regression

The quantile regression model (QR) linearly relates the τth quantile of the survival time to the covariates Z as

$$Q_{\log T_i}(\tau|Z_i) = Z_i'\beta_0(\tau),$$

where $Q_T(\tau|Z) = \inf\{t : P(T \leq t|Z) \geq \tau\}$ and $\beta_0(\tau)$ is the regression coefficient at the τth quantile. The QR model states that, for a fixed τ,

$$\log(T_i) = Z_i'\beta_0(\tau) + e_i(\tau), \tag{2.1}$$

where $e_i(\tau)$ is a random error whose τth quantile given Z_i is zero. This model is clearly an extension of the AFT model where the errors are iid and independent of the covariates.

Remark 1. In order to estimate $\beta_0(\tau_0)$ at the τ_0th quantile, Peng and Huang (2008) require that the quantile regression model hold for all $\tau \leq \tau_0$ simultaneously. Our model only assumes that it holds at a single quantile. Thus, to estimate the conditional median for example, we only need (2.1) to hold at $\tau = 0.5$.

Let $F_0(\cdot|Z)$ be the continuous conditional distribution function of T given Z. Denote the counting process by $N(t) = \delta I(Y \leq t)$ and the martingale process associated with $N(t)$ as $M(t) = N(t) - \Lambda_0(t \wedge Y|Z)$, where $\Lambda_0(\cdot|Z) = -\log(1 -$

$F_0(\cdot|Z))$ is the conditional cumulative hazards function of T. Since $\mathrm{E}\{M(t)|Z\} = 0$ for $t \geq 0$, we have that

$$\mathrm{E}\frac{1}{n}\sum_{i=1}^{n} Z_i\big\{N_i(\exp(Z_i'\beta_0(\tau))) - \Lambda_0(Y_i \wedge \exp(Z_i'\beta_0(\tau))|Z_i)\big\} = 0, \tag{2.2}$$

where we have used the empirical version of $E\{ZM(\exp(Z'\beta_0(\tau)))|Z\} = 0$. Because the function Λ_0 is unknown, we use the quantile property $F_0(\exp(Z_i'\beta_0(\tau)) |Z_i) = \tau$ to obtain

$$\Lambda_0(Y_i \wedge \exp(Z_i'\beta_0(\tau))|Z_i) = H_\tau(F_0(Y_i|Z_i)), \tag{2.3}$$

where $H_\tau(t) = H(\tau) \wedge H(t)$, and $H(t) = -\log(1-t)$ for $t \in [0,1)$ is a strictly increasing function. Combining (2.2) and (2.3), we have the estimating function

$$U_n(\beta) = \frac{1}{n}\sum_{i=1}^{n} Z_i\big\{N_i(\exp(Z_i'\beta)) - H_\tau(F_0(Y_i|Z_i))\big\} \approx 0 \tag{2.4}$$

to estimate β, where we have written $\beta = \beta(\tau)$ for convenience. The approximation $\approx$ is used because an exact solution may not exist (Ying, Jung, and Wei (1995)).

In principle, as long as we can form unbiased estimating functions, consistent estimates for $\beta(\tau)$ can be obtained. Here we use Peng and Huang's approach due to its martingale structure that naturally accommodates the conditional independence censoring mechanism (Kalbfleisch and Prentice (2002)). The underlying counting process notion provides a unified framework for model checking and diagnosis.

Equation (2.4) cannot be used as an estimating equation since the conditional distribution $F_0(\cdot|z)$ is unknown. Given the observed data $\{Y_i, \delta_i, Z_i\}, i = 1, \ldots, n$, we propose to estimate the distribution function $F_0(\cdot|z)$ non-parametrically using the local Kaplan-Meier estimator $\hat{F}(\cdot|z)$ as

$$\hat{F}(t|z) = 1 - \prod_{j=1}^{n}\left[1 - \frac{W_{nj}(z)}{\sum_{k=1}^{n} I(Y_k \geq Y_j)W_{nk}(z)}\right]^{I(Y_j \leq t, \delta_j = 1)}, \tag{2.5}$$

where $W_{nj}(z)$ is a sequence of weights adding up to 1. When $W_{nj}(z) = 1/n$ for all j, $\hat{F}(t|z)$ reduces to the classical Kaplan-Meier estimator. This estimator reduces to the kernel estimator of the conditional cumulative distribution function when no censoring occurs. Similar to Wang and Wang (2009), we use the kernel weights

$$W_{nj}(z) = K\Big(\frac{z - Z_j}{h_n}\Big)\Big[\sum_{k=1}^{n} K\Big(\frac{z - Z_k}{h_n}\Big)\Big]^{-1},$$

where $K(\cdot)$ is a kernel function and $h_n > 0$ is the bandwidth converging to zero when n goes to infinity. When there is only a single continuous covariate in Z other than the intercept ($p = 1$), we use the bi-quadratic kernel $k(x) = (15/16)(1-x^2)^2 I(|x| \leq 1)$ for the covariate. When $p \geq 2$ and there are multiple continuous covariates in Z, we use a product kernel with a higher order kernel for each covariate, as discussed in Müller (1988, Thm. 5.4). For example, when $p = 2$, we use the kernel $K(x) = (15/32)(3 - 10x^2 + 7x^4)I(|x| \leq 1)$ for each covariate and a product kernel for the two non-constant covariates. If $p = 3$, we use $K(x) = (35/256)(15 - 105x^2 + 189x^4 - 99x^6)I(|x| \leq 1)$. Since these higher order kernels can give an estimate of $F(\cdot|z)$ outside $[0,1]$, we truncate the value to $[0,1]$ as needed. The working estimating equation is then

$$U_n(\beta) = \frac{1}{n}\sum_{i=1}^{n} Z_i\big\{N_i(\exp(Z_i'\beta)) - H_\tau(\hat{F}(Y_i|Z_i))\big\} \approx 0. \tag{2.6}$$

Remark 2. Peng and Huang (2008) observed that

$$H_\tau(F_0(Y_i|Z_i)) = \int_0^\tau I[Y_i \geq \exp(Z_i'\beta_0(u))]dH(u), \tag{2.7}$$

under a global linearity assumption, and proposed to approximate the integral by a grid-based procedure similar to the Euler's forward rule. This approximation requires estimation of all $\beta(\tau)$'s on a fine grid in order to estimate a single quantile at τ. In contrast, we make the linearity assumption only at a particular quantile and our model only needs to estimate $\beta(\tau)$ at the τth quantile.

The working estimating equation in (2.6) is monotonic, and its solution is the minimizer of the quantile regression objective function

$$\frac{1}{n}\sum_{i=1}^{n}\{\delta_i\varphi_\tau(\log Y_i - Z_i'\beta) + \varphi_\tau(Y_i^* - (\tau^{-1}H_\tau(\hat{F}(Y_i|Z_i)) - \delta_i)Z_i'\beta)\}, \tag{2.8}$$

where $\varphi_\tau(s) = s(\tau - I(s \leq 0))$ is the check function (Koenker and Bassett (1978)) and $Y_i^* > \max_j\{(\tau^{-1}H_\tau(\hat{F}(Y_j|Z_j)) - \delta_j)Z_j'\beta\}$. We take $Y_i^* = \max_j\{\log(Y_j)\} + 100,\ i = 1,\ldots,n$, for simulation and data analysis.

The formulation in (2.8) suggests use of the fast quantile regression code developed by Portnoy and Koenker (1997) for computation. In particular, after we have estimated the conditional distribution function F_0, we can make use of R function *qr* in the R package *quantreg* to fit a weighted quantile regression model using the augmented data set $\{(Y_i, Z_i)\}_{i=1}^n$ and $\{Y_i^*, (\tau^{-1}H_\tau(\hat{F}(Y_i|Z_i)) - \delta_i)Z_i\}_{i=1}^n$, with weights $\{\delta_i\}_{i=1}^n$ for $\{(Y_i, Z_i)\}_{i=1}^n$. Thus, the extra effort needed to implement our approach is minimal.

2.2. Transformation quantile regression

We extend the log-transformed linear quantile regression to the power-transformed linear quantile regression in which the unknown transformation parameter needs to be estimated. This is a flexible family of monotone transformations and gives many useful structures (Box and Cox (1964); Cai, Tian, and Wei (2005); Mu and He (2007); Yin, Zeng, and Li (2008)).

The family of power transformations (Box and Cox (1964)) is given by

$$\rho_\gamma(x) = \begin{cases} \frac{x^\gamma - 1}{\gamma} & \text{if } \gamma \neq 0, \\ \log(x) & \text{if } \gamma = 0. \end{cases}$$

A practical range of γ often used is $[-2, 2]$ and can be extended if necessary. For the power-transformed linear quantile regression model, we assume that the τth quantile of $\rho_{\gamma_0(\tau)}(T_i)$ is linearly related to Z_i

$$Q_{\rho_{\gamma_0(\tau)(T_i)}}(\tau|Z_i) = Z_i'\beta_0(\tau),$$

where $\gamma_0(\tau)$ is the corresponding unknown transformation parameter and $\beta_0(\tau)$ is the quantile regression parameter. In the sequel, we suppress the τ in $\beta(\tau)$ and $\gamma(\tau)$ to simplify the notation. We write the inverse power transformation function as

$$\rho_\gamma^{(-1)}(x) = \begin{cases} (\gamma x + 1)^{1/\gamma} & \text{if } \gamma \neq 0, \\ \exp(x) & \text{if } \gamma = 0. \end{cases}$$

If the transformation parameter γ_0 is known, the τth quantile of T_i given Z_i is $\rho_{\gamma_0}^{(-1)}(Z_i'\beta_0)$. Following (2.4), we propose to estimate β_0 by solving

$$U_n^*(\beta; \gamma) = \frac{1}{n}\sum_{i=1}^n Z_i\big\{N_i(\rho_\gamma^{(-1)}(Z_i'\beta)) - H_\tau(\hat{F}(Y_i|Z_i))\big\} \approx 0. \tag{2.9}$$

Similar to (2.8), we need to minimize

$$\frac{1}{n}\sum_{i=1}^n \{\delta_i\varphi_\tau(\rho_\gamma(Y_i) - Z_i'\beta) + \varphi_\tau(Y_i^{**} - (\tau^{-1}H_\tau(\hat{F}(Y_i|Z_i)) - \delta_i)Z_i'\beta)\}$$

for large Y_i^{**}. We denote the resulting estimator as $\hat{\beta}_n(\gamma)$ to emphasize the dependence on the power transformation parameter γ.

To estimate γ, a common approach in the quantile regression literature is to use some discrepancy function based on a goodness-of-fit measure (Mu and He (2007); Yin, Zeng, and Li (2008)). To this end, we define a cumsum process of residuals indexed by z,

$$R_n(z, \gamma) = \frac{1}{n}\sum_{i=1}^n I(Z_i \leq z)\big\{N_i(\rho_\gamma^{(-1)}(Z_i'\hat{\beta}_n(\gamma))) - H_\tau(\hat{F}(Y_i|Z_i))\big\},$$

where $Z_i \le z$ means componentwise inequality. The cumsum process is used widely, for example in Lin, Wei, and Ying (1993, 2002), He and Zhu (2003), Mu and He (2007), and Yin, Zeng, and Li (2008). Following the proof of Theorem 3, the process $n^{1/2}R_n(z,\gamma_0)$ converges to a zero-mean Gaussian process in z, while for any $\gamma \ne \gamma_0$, the process $n^{1/2}R_n(z,\gamma)$ diverges to infinity for all $z \in R^p$ (Mu and He (2007)). Because of this, the estimator $\hat\gamma_n(\tau)$ (or $\hat\gamma_n$) of γ_0 is obtained by minimizing

$$R_n^*(\gamma) = \frac{1}{n}\sum_{i=1}^n \|R_n(Z_i,\gamma)\|^2, \quad \text{for } \gamma \in \Upsilon,$$

where Υ denotes the parameter space for γ. Subsequently, β_0 is estimated as $\hat\beta_n(\hat\gamma_n)$. The objective function only involves a single parameter although it is not differentiable with respect to γ. The standard grid search algorithm can be used to obtain the estimator $\hat\gamma_n$. Typically, we would not use $\hat\gamma_n$ as the final estimate of the transformation parameter, but would use the nearest convenient value in a sequence such as -1, 0, 1 for better interpretation, after checking that such a value lies within a selected confidence interval (Box and Cox (1964)).

Remark 3. When the transformation parameter γ is fixed, the estimating equation approach adopted by Ying, Jung, and Wei (1995) and Yin, Zeng, and Li (2008) takes the form

$$n^{-1}\sum_{i=1}^n Z_i\Big\{\frac{I(\rho_\gamma(Y_i) - \beta(\gamma)'Z_i \ge 0)}{\hat G(\rho_\gamma^{(-1)}(\beta(\gamma)'Z_i))} - \tau\Big\} \approx 0, \tag{2.10}$$

where $\hat G$ is the Kaplan-Meier estimator for the censoring times based on $\{(Y_i, 1-\delta_i)\}_{i=1}^n$. Since $\beta(\gamma)$ in $\hat G$ is unknown, Yin, Zeng, and Li (2008) used an iterative algorithm. Our estimating equation approach requires estimating the conditional cumulative density function of the failure time just once.

3. Theory

Let $G_0(t|z)$ be the conditional distribution function of C given $Z = z$, and $f_Z(z)$ be the marginal density function of Z. We focus on the situation where there is one continuous covariate, deferring the discussion of the multiple covariate case to Appendix D. To derive the asymptotic properties of the proposed estimator, we make some regularity assumptions.

C1. T and C are conditionally independent given the covariate Z.

C2. The true value β_0 of β is in the interior of a bounded convex region $\mathcal{B}$. The support $\mathcal{Z}$ of Z is bounded and compact.

C3. $\inf_{z\in\mathcal{Z}} P(Y \ge \mathcal{T}|z) \ge 1 - \eta_0 > 0$, where $\mathcal{T} = \sup_{z\in\mathcal{Z},\beta\in\mathcal{B}} \exp(z'\beta)$.

C4. The marginal density function $f_Z(z)$, the conditional density functions $f_0(t|z)$ and $g_0(t|z)$ of the failure time T and C are uniformly bounded away from infinity and have bounded (uniformly in t) first and second order partial derivatives with respect to z.

C5. The bandwidth h_n satisfies $h_n = O(n^{-v})$ with $1/4 < v < 1/2$.

C6. The kernel function $K(\cdot)$ has a compact support $[-1, 1]$, and satisfies the Lipschitz condition of order 1, $\int K(u)du = 1, \int uK(u)du = 0, \int K^2(u)du < \infty$ and $\int u^2 K(u)du < \infty$.

C7. For β in the neighborhood of β_0, $E[ZZ' \exp(Z'\beta) f_0(\exp(Z'\beta)|Z)(1 - G_0 (\exp(Z'\beta)|Z))]$ is positive definite.

Condition C1 is standard in survival analysis. Condition C2, C3, and C4 are standard in analyzing failure time data. Condition C5 specifies the bandwidth choice needed for establishing the rate of convergence of the local Kaplan-Meier estimator and subsequently the consistency of $\hat{\beta}_n$. For asymptotic normality, the rate in Condition C5 needs to be strengthened to $1/4 < v < 1/3$. Condition C6 is usually made for kernel smoothing and it holds for the bi-quadratic kernel we use in this paper. Condition C7 ensures that the quantile regression estimator is unique and it is used to establish the asymptotic normality of the estimator.

Theorem 1 (Consistency). *Under* C1−C7*, the solution* $\hat{\beta}_n$ *to* (2.6) *satisfies* $\hat{\beta}_n \to \beta_0$ *in probability as* $n \to \infty$.

Theorem 2 (Asymptotic normality). *Under Assumptions* C1−C7 *and* $1/4 < v < 1/3$*, we have*

$$n^{1/2}(\hat{\beta}_n - \beta_0) \xrightarrow{d} N(0, \Gamma_1^{-1} V_1 \Gamma_1^{-1}),$$

where $\Gamma_1 = E\{ZZ' \exp(Z'\beta_0) f_0(\exp(Z'\beta_0)|Z)(1 - G_0(\exp(Z'\beta_0)|Z))\}$ *and* V_1 *is defined in Lemma A.3 in Appendix B.*

The matrices Γ_1 and V_1 in the limiting covariance matrix depend on the unknown conditional density functions $f_0(\cdot|z)$ and $g_0(\cdot|z)$ and may be difficult to estimate nonparametrically from finite samples. We propose to use the bootstrap re-sampling procedure for inference following Chen, Linton, and van Keilegom (2003). Operationally, we draw with replacement $\{(Y_i^*, \delta_i^*, Z_i^*)\}_{i=1}^n$ from the original data set $\{(Y_i, \delta_i, Z_i)\}_{i=1}^n$ and re-estimate β as $\tilde{\beta}_n$. This process is repeated many times. Denote the estimates as $\tilde{\beta}_n^{(j)}$, $j = 1, \ldots, B$ for a large B. The sampling distribution of $\tilde{\beta}_n^{(j)}$ can be used to approximate the sampling distribution of $\hat{\beta}_n$. The bootstrap variance of the estimates computed from the bootstrap sample $\{(Y_i^*, \delta_i^*, Z_i^*)\}_{i=1}^n$ is a consistent estimate of the asymptotic variance of $\hat{\beta}_n$ (See Hall (1992, p.159)). A sketch of the justification is provided in the Supplementary File.

Next we study the large-sample properties of the power-transformed estimator. To establish these results, we need more regularity conditions.

C3′. Assumption C3 holds with $\mathcal{T} = \sup_{z\in\mathcal{Z},\gamma\in\Upsilon,\beta\in\mathcal{B}} \rho_\gamma^{(-1)}(z'\beta)$.

C7′. $E[ZZ'\dot{\rho}_\gamma^{(-1)}(Z'\beta)f_0(\rho_\gamma^{(-1)}(Z'\beta)|Z)(1-G_0(\rho_\gamma^{(-1)}(Z'\beta)|Z))]$ is positive definite for β in the neighborhood of β_0 and γ in the neighborhood of γ_0, where $\dot{\rho}_\gamma^{(-1)}(x) = \partial\rho_\gamma^{(-1)}(x)/\partial x$.

C8. γ_0 is an interior point of a compact set Υ.

These assumptions are discussed in Yin, Zeng, and Li (2008).

Theorem 3. *Under Assumptions* C1, C2, C3′, C4-C6, C7′, *and* C8*, we have* $\hat\beta_n(\hat\gamma_n) \to \beta_0$ *and* $\hat\gamma_n \to \gamma_0$ *in probability. Furthermore if* $1/4 < v < 1/3$*, we have*

$$n^{1/2}\left\{ \begin{pmatrix} \hat\beta_n(\hat\gamma_n) \\ \hat\gamma_n \end{pmatrix} - \begin{pmatrix} \beta_0 \\ \gamma_0 \end{pmatrix} \right\} \xrightarrow{d} N(0, \Gamma_2^{-1}V_2(\Gamma_2')^{-1}),$$

with Γ_2 *and* V_2 *defined in Appendix C.*

The proofs outlined in the Appendix depend heavily on empirical process theory (van der Vaart and Wellner (1996)). In Appendix D, we outline the results for $p > 1$. For statistical inference of $\hat\gamma_n$ and $\hat\beta_n(\hat\gamma_n)$, we again resort to the bootstrap re-sampling method as discussed before.

We briefly discuss the problem of choosing the bandwidth h_n. In practice, we can use the K-fold cross validation by dividing the dataset into K parts $D_1, \ldots, D_K$ which are roughly equally-sized. For D_j, we fit the model by using the data from the other $K-1$ parts, and calculate a loss from predicting the τth conditional quantile of T for the uncensored data in D_j,

$$\mathrm{L}_j(h) = \frac{1}{\#\{i : i \in D_j\}} \sum_{k\in D_j} \|R_n(Z_k, \hat\gamma^{(-j)})\|^2,$$

where $\hat\gamma^{(-j)}$ is the estimated parameter using data in $D_1, \ldots, D_{j-1}, D_{j+1}, \ldots, D_K$. This procedure is repeated for $j = 1, \ldots, K$ and the average loss $\mathrm{L}(h) = \sum_{j=1}^K \mathrm{L}_j(h)/K$ is computed for bandwidth h. We take $h_n = \mathrm{argmin}\ h\ \mathrm{L}(h)$.

4. Numerical Study

We present three simulations examples to illustrate the finite sample performance of the approach. In Example 1, we compared our method with that of Yin, Zeng, and Li (2008) when T and C are unconditionally independent. In Example 2, we compared them when T are C are conditionally independent. In Example 3, we considered multiple covariates. For these approaches, we used R function *optimize* to locate the optimal γ in the interval $[-2, 2]$. We fixed γ first and then estimated $\beta_n(\gamma)$. The optimal γ minimized $R_n^*(\gamma)$.

Example 1. The main purposes of the simulation were to compare our approach with that in Yin, Zeng, and Li (2008), YZL, when the unconditional independence of T and C is satisfied, and to evaluate the validity of the bootstrap method for statistical inference.

We followed Yin, Zeng, and Li (2008) in generating data from the Box-Cox transformation linear quantile regression model

$$\rho_\gamma(T) = \frac{T^\gamma - 1}{\gamma} = \beta_0 + \beta_1 Z + \varepsilon, \tag{4.1}$$

where $\beta_0 = 0.2$, $\beta_1 = 1$, and γ was taken to be 0, 0.5, 1 respectively. The covariate Z was Unif$[0, 2]$ and the error was $N(0, 0.25)$. Besides this simple example, we also considered the skewed error model and the heteroscedastic models with $\beta_0 = -0.5$, $\beta_1 = 1$, and $\gamma = 0.5$, as in Yin, Zeng, and Li (2008). Here we took ε from a shifted chi-squared distribution with one degree of freedom and median zero. For the heteroscedastic model, we took a heteroscedastic error $Z\varepsilon$, where $Z \sim$ Unif$[0, 2]$ and $\varepsilon \sim N(0, 0.25)$. The censoring times were generated independently from the uniform distribution to yield censoring rate of 20% or 40%. We took $n = 300$ and assessed the median quantile regression at $\tau = 0.5$. For each configuration, we generated 1,000 datasets. To assess the performance of the bootstrap re-sampling method for statistical inference, we generated 2,000 bootstrap samples for each simulated dataset.

In Table 1, we compare our proposed approach with YZL in terms of average estimates over 1,000 simulations, together with the sample standard deviations of the estimates (SD). Note that the results for YZL are taken from Yin, Zeng, and Li (2008). For simplicity, we take the bandwidth as $n^{-1/3+0.01}$. This particular choice of the bandwidth is motivated by the asymptotic result and is by no means optimal. Additional simulations showed that the results were not very sensitive to the bandwidth and the kernel. From Table 1, we see that the proposed estimator generally gives estimates with small bias. As the censoring ($c\%$) increases, the bias and the sample standard deviation increase. This agrees with the asymptotic results. The proposed method performs similarly to YZL when no censoring occurs; when censoring is high ($c\% = 40\%$), the proposed method gives estimates with smaller SD. These observations suggest that estimation of the conditional distribution function $F(T|Z)$ has little effect on the estimation of the transformation parameter and the linear quantile regression parameter. We also compared the SDs with the average of the estimated standard errors based on the bootstrap method (SE). As seen in Table 1, they are very close. For models where unconditional independence is appropriate, our methods performs competitively compared to Yin et al.'s approach.

Table 1. Estimation of the unknown transformation parameter and the linear quantile regression coefficients. SD: the sampling standard deviation; SE: the average of the standard errors using bootstrap.

		γ				β_0				β_1			
		Proposed		YZL		Proposed		YZL		Proposed		YZL	
γ	$c\%$	$\hat{\gamma}$	$\mathrm{SD}_{\mathrm{SE}}$	$\hat{\gamma}$	SD	$\hat{\beta}_0$	$\mathrm{SD}_{\mathrm{SE}}$	$\hat{\beta}_0$	SD	$\hat{\beta}_1$	$\mathrm{SD}_{\mathrm{SE}}$	$\hat{\beta}_1$	SD
Simple													
0	0	-0.012	$0.273_{0.289}$	0.010	0.261	0.210	$0.112_{0.105}$	0.182	0.114	1.032	$0.351_{0.363}$	1.057	0.328
	20	-0.027	$0.289_{0.291}$	0.027	0.315	0.213	$0.110_{0.110}$	0.181	0.130	1.023	$0.351_{0.367}$	1.092	0.409
	40	-0.097	$0.409_{0.425}$	0.183	0.417	0.213	$0.116_{0.119}$	0.148	0.158	1.001	$0.453_{0.477}$	1.295	0.601
0.5	0	.477	$0.439_{0.443}$	0.503	0.440	0.203	$0.111_{0.119}$	0.181	0.119	1.051	$0.423_{0.433}$	1.078	0.427
	20	0.469	$0.440_{0.469}$	0.528	0.513	0.200	$0.108_{0.117}$	0.177	0.128	1.045	$0.423_{0.431}$	1.123	0.497
	40	.423	$0.523_{0.548}$	0.560	0.621	0.210	$0.115_{0.121}$	0.151	0.171	1.020	$0.443_{0.462}$	1.202	0.640
1	0	1.053	$0.621_{0.632}$	0.979	0.555	0.188	$0.138_{0.142}$	0.193	0.116	1.076	$0.543_{0.555}$	1.057	0.432
	20	1.037	$0.681_{0.691}$	0.975	0.692	0.181	$0.159_{0.167}$	0.185	0.132	1.117	$0.554_{0.562}$	1.103	0.583
	40	0.974	$0.705_{0.712}$	0.982	0.873	0.177	$0.154_{0.167}$	0.158	0.167	1.141	$0.537_{0.545}$	1.189	0.751
Skewed													
0.5	0	0.444	$0.554_{0.578}$	0.402	0.554	-0.488	$0.159_{0.173}$	-0.478	0.155	1.004	$0.269_{0.273}$	0.984	0.243
	20	0.403	$0.597_{0.612}$	0.422	0.614	-0.467	$0.159_{0.177}$	-0.475	0.149	0.993	$0.260_{0.270}$	0.993	0.257
	40	0.400	$0.643_{0.656}$	0.508	0.704	-0.456	$0.184_{0.191}$	-0.469	0.158	0.983	$0.290_{0.302}$	1.009	0.299
Heteroscedastic													
0	0	0.041	$0.200_{0.204}$	0.040	0.205	-0.498	$0.016_{0.017}$	-0.499	0.017	1.030	$0.062_{0.065}$	1.007	0.059
	20	0.033	$0.213_{0.220}$	0.059	0.242	-0.498	$0.017_{0.019}$	-0.497	0.019	1.033	$0.063_{0.068}$	1.007	0.066
	40	0.005	$0.301_{0.313}$	0.010	0.328	-0.500	$0.025_{0.027}$	-0.499	0.019	1.029	$0.067_{0.070}$	1.003	0.096
0.5	0	0.508	$0.199_{0.209}$	0.520	0.272	-0.500	$0.017_{0.021}$	-0.498	0.023	1.019	$0.051_{0.053}$	1.000	0.057
	20	0.517	$0.220_{0.239}$	0.526	0.305	-0.499	$0.020_{0.022}$	-0.499	0.026	1.034	$0.055_{0.061}$	1.002	0.066
	40	0.522	$0.271_{0.291}$	0.660	0.361	-0.499	$0.023_{0.024}$	-0.488	0.028	1.034	$0.060_{0.064}$	0.998	0.080
1	0	1.026	$0.220_{0.229}$	1.019	0.220	-0.498	$0.022_{0.024}$	-0.499	0.023	1.025	$0.050_{0.054}$	1.002	0.049
	20	1.008	$0.237_{0.241}$	1.008	0.279	-0.499	$0.023_{0.024}$	-0.500	0.027	1.027	$0.054_{0.055}$	0.999	0.057
	40	1.010	$0.267_{0.270}$	1.012	0.351	-0.500	$0.029_{0.031}$	-0.499	0.033	1.035	$0.062_{0.064}$	0.996	0.073

We also compared the proposed method with YZL when the transformation parameter γ is known. This is termed conditional inference in Mu and He (2007) and Yin, Zeng, and Li (2008). The results are in Table 2. We can see that the biases are very small and that the biases and the sample standard deviations increase with the censoring rate. Furthermore, the proposed method performs competitively compared to YZL, indicating again that estimating the conditional distribution of the failure time does not affect the estimation of the parameters. Comparing Tables 1 and 2 indicates that conditional inference for β is more efficient than that for which γ needs to be estimated. We also note that in some cases, our method has larger biases here compared to Yin et al's. The bias problem becomes less severe if the bandwidth is decreased (results not shown). Furthermore, The bootstrap method for estimating the standard errors is satisfactory, as all the SEs are close to the SDs in Table 2, and all the coverage

Table 2. Estimation under the transformation model when γ is known.

		β_0				β_1			
		Proposed		YZL		Proposed		YZL	
γ	$c\%$	$\hat{\beta}_0$	$\mathrm{SD}_{\mathrm{SE}}$	$\hat{\beta}_0$	SD	$\hat{\beta}_1$	$\mathrm{SD}_{\mathrm{SE}}$	$\hat{\beta}_1$	SD
Simple									
0	0	0.206	$0.074_{0.079}$	0.200	0.073	1.002	$0.058_{0.064}$	1.000	0.066
	20	0.205	$0.077_{0.080}$	0.200	0.080	1.006	$0.070_{0.077}$	0.997	0.071
	40	0.211	$0.081_{0.088}$	0.229	0.087	1.012	$0.080_{0.088}$	0.948	0.078
0.5	0	0.208	$0.071_{0.079}$	0.200	0.073	1.003	$0.065_{0.069}$	1.000	0.062
	20	0.203	$0.079_{0.088}$	0.199	0.081	1.008	$0.071_{0.075}$	1.000	0.073
	40	0.208	$0.086_{0.092}$	0.203	0.094	1.011	$0.079_{0.085}$	0.997	0.089
1	0	0.205	$0.069_{0.070}$	0.202	0.069	1.001	$0.060_{0.066}$	1.001	0.060
	20	0.209	$0.080_{0.084}$	0.204	0.080	1.003	$0.071_{0.078}$	1.000	0.074
	40	0.210	$0.085_{0.091}$	0.203	0.094	1.013	$0.075_{0.083}$	0.997	0.089
Skewed									
0.5	0	-0.471	$0.121_{0.131}$	-0.487	0.127	1.001	$0.103_{0.102}$	0.994	0.113
	20	-0.475	$0.133_{0.138}$	-0.488	0.138	1.012	$0.111_{0.109}$	0.993	0.124
	40	-0.480	$0.154_{0.167}$	-0.486	0.140	1.020	$0.139_{0.133}$	0.967	0.124
Heteroscedastic									
0	0	-0.499	$0.014_{0.014}$	-0.500	0.014	1.020	$0.050_{0.052}$	0.999	0.049
	20	-0.500	$0.014_{0.015}$	-0.499	0.017	1.026	$0.052_{0.051}$	0.996	0.058
	40	-0.498	$0.015_{0.015}$	-0.498	0.022	1.024	$0.059_{0.057}$	0.974	0.067
0.5	0	-0.500	$0.012_{0.013}$	-0.501	0.016	1.016	$0.049_{0.047}$	1.000	0.050
	20	-0.500	$0.017_{0.018}$	-0.499	0.019	1.024	$0.057_{0.055}$	0.998	0.059
	40	-0.499	$0.015_{0.016}$	-0.492	0.024	1.003	$0.057_{0.057}$	0.967	0.068
1	0	-0.499	$0.014_{0.014}$	-0.499	0.014	1.022	$0.051_{0.052}$	1.000	0.051
	20	-0.500	$0.015_{0.016}$	-0.501	0.018	1.028	$0.056_{0.047}$	1.000	0.062
	40	-0.501	$0.015_{0.016}$	-0.499	0.025	1.036	$0.064_{0.067}$	0.994	0.080

probabilities are close to the nominal 95% level in Table 3.

We note that the proposed method requires an estimation of the local Kaplan-Meier survival function and is computationally more intensive than that in Yin, Zeng, and Li (2008). However, the computational overhead is minimal, as estimating $F(t|z)$ using local smoothing can be efficiently implemented (Wang and Wang (2009)). Moreover, the fact that we only need to estimate $F(t|z)$ once implies that the extra computational demand is worth the trade-off with a relaxed and more natural conditional independence assumption.

Example 2. To compare the performance of our approach with that in Yin, Zeng, and Li (2008). under conditional independence, we simulated data from

$$\log T = \beta_0 + \beta_1 Z + 0.2(Z-1)^2\varepsilon,$$

where $\beta_0 = 2$, $\beta_1 = 1$, $Z \sim \mathrm{Unif}[0,1]$, and $\varepsilon = \chi_1^2 - \Phi^{-1}(\tau)$ is a centered chi-square random variable with one degree of freedom with Φ the cumulative distribution

Table 3. The coverage probability (in percentage) of the 95% confidence interval.

	$c\% = 0\%$			$c\% = 20\%$			$c\% = 40\%$		
γ	$\hat{\gamma}$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\gamma}$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\gamma}$	$\hat{\beta}_0$	$\hat{\beta}_1$
Simple, fixed γ									
0	-	93.2	95.2	-	92.8	93.8	-	94.6	95.8
0.5	-	94.2	94.2	-	94.4	95.2	-	93.4	94.4
1	-	93.4	94.4	-	93.8	95.4	-	94.6	94.4
Simple, estimated γ									
0	96.2	92.2	95.0	97.4	92.6	95.4	97.0	93.8	98.0
0.5	97.6	92.8	96.2	97.8	93.2	95.6	97.2	92.4	97.4
1	97.4	92.8	94.7	98.2	91.6	95.8	97.0	92.0	96.8
Skewed, fixed γ									
0.5	-	97.4	93.0	-	97.6	92.2	-	97.2	92.8
Skewed, estimated γ									
0.5	97.0	97.2	97.6	98.0	96.8	98.0	97.8	98.0	96.6
Heteroscedastic, fixed γ									
0	-	96.0	95.2	-	96.6	94.1	-	96.9	93.9
0.5	-	97.5	92.4	-	97.7	92.3	-	97.9	93.2
1	-	92.5	96.8	-	94.1	96.4	-	97.7	94.0
Heteroscedastic, estimated γ									
0	94.2	96.0	95.5	96.9	93.7	97.1	97.0	96.8	95.9
0.5	96.5	94.9	97.4	97.9	95.8	97.5	97.8	98.0	97.9
1	97.0	93.3	93.0	94.1	93.5	94.1	96.9	97.1	96.8

function of χ_1^2. We looked at the parameter estimates at $\tau = 0.5$. The global linearity assumption needed by Peng and Huang (2008) is not satisfied. The censoring time was generated as $C = b^{-1}\log(2 + Z + e)$ where e is a standard normal variate and the constant b was taken to have about 0%, 20%, 40% and 60% observations censored. This censoring time entails that the failure time and the censoring time are conditionally independent given Z, and the marginal independence assumption needed by Yin, Zeng, and Li (2008) is not satisfied.

For every simulation setup, we generated 1,000 datasets. We took $n = 300$ and used the bandwidth $h_n = n^{-1/3+0.01}$. Table 4 summarizes the averages of the parameter estimates, as well as their sample standard deviations. With no censoring, YZL outperformed our method in terms of standard deviation, since we needed to estimate the conditional distribution function. However, our method outperformed YZL whenever censoring was present. YZL gived biased estimates

Table 4. The mean of the estimated parameters of our method (Proposed), Yin et al.'s method (YZL), Peng and Huang's method (PH) and Wang and Wang's method (WW).

			$c\% = 0\%$			$c\% = 20\%$			$c\% = 40\%$			$c\% = 60\%$		
n	Method		$\hat{\gamma}$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\gamma}$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\gamma}$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\gamma}$	$\hat{\beta}_0$	$\hat{\beta}_1$
100	Proposed	Mean	-0.020	2.015	1.002	-0.072	1.953	1.071	-0.124	1.910	1.090	-0.232	1.835	1.198
		SD	0.229	0.381	0.584	0.296	0.492	0.844	0.379	0.610	1.057	0.507	0.859	2.196
	YZL	Mean	-0.014	2.008	1.078	0.009	2.072	1.255	0.143	2.182	1.588	0.147	2.431	2.755
		SD	0.195	0.325	0.509	0.274	0.442	0.736	0.368	0.610	1.218	0.458	0.900	4.009
	Fixed γ													
	Proposed	Mean	-	2.002	0.995	-	2.003	0.995	-	2.005	0.995	-	2.008	0.993
		SD	-	0.020	0.020	-	0.018	0.021	-	0.023	0.030	-	0.032	0.038
	YZL	Mean	-	2.001	0.999	-	1.994	1.007	-	1.987	1.017	-	1.977	1.031
		SD	-	0.017	0.020	-	0.017	0.020	-	0.021	0.025	-	0.026	0.035
	PH	Mean	-	2.002	0.998	-	2.003	0.997	-	2.003	0.997	-	2.007	0.992
		SD	-	0.018	0.020	-	0.019	0.021	-	0.025	0.028	-	0.032	0.037
	WW	Mean	-	2.002	0.998	-	2.002	0.998	-	2.003	0.998	-	2.005	0.995
		SD	-	0.018	0.020	-	0.018	0.021	-	0.024	0.027	-	0.030	0.035
300	Proposed	Mean	0.035	2.076	1.088	0.025	2.060	1.068	0.006	2.030	1.075	-0.049	1.952	0.992
		SD	0.086	0.150	0.234	0.109	0.185	0.281	0.139	0.232	0.342	0.205	0.328	0.461
	YZL	Mean	0.002	2.008	1.021	0.055	2.101	1.190	0.102	2.193	1.396	0.147	2.293	1.671
		SD	0.063	0.109	0.179	0.077	0.138	0.242	0.136	0.224	0.394	0.203	0.337	0.667
	Fixed γ													
	Proposed	Mean	-	2.000	0.998	-	2.001	0.998	-	2.002	0.998	-	2.003	0.997
		SD	-	0.010	0.011	-	0.011	0.012	-	0.013	0.014	-	0.014	0.016
	YZL	Mean	-	2.001	0.999	-	1.994	1.007	-	1.987	1.016	-	1.978	1.028
		SD	-	0.009	0.010	-	0.010	0.012	-	0.011	0.014	-	0.015	0.019
	PH	Mean	-	2.002	0.998	-	2.002	0.998	-	2.002	0.998	-	2.005	0.995
		SD	-	0.009	0.010	-	0.011	0.012	-	0.012	0.013	-	0.016	0.018
	WW	Mean	-	2.001	0.999	-	2.002	0.999	-	2.003	0.997	-	2.005	0.997
		SD	-	0.009	0.010	-	0.010	0.011	-	0.012	0.013	-	0.015	0.016

whenever censoring occurs. This agrees with the argument that YZL works only when the independence assumption holds. For different sample sizes and censoring rates, our method was unbiased, and the sampling standard deviation decreased with the increasing sample size and the decreasing censoring rate. When γ was known in advance, YZL performed competitively with our method and two other approaches including Wang and Wang's and Peng and Huang's, although usually with a larger bias in the high censoring scenarios ($c\% = 40\%$ or 60%).

Taking the results in Example 1 and 2 together, the simulations suggest that our method is preferred over YZL when T and C are conditionally independent given, and that our method performs similarly with YZL when they are unconditionally independent.

Example 3. The proposed method depends on estimation of a nonparametric

Table 5. The effect of dimensionality when γ is estimated. Here p is the number of the continuous non-intercept covariates in the model.

		$c\% = 0\%$			$c\% = 20\%$			$c\% = 40\%$			$c\% = 60\%$		
p		$\hat{\gamma}$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\gamma}$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\gamma}$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\gamma}$	$\hat{\beta}_0$	$\hat{\beta}_1$
2	Mean	-0.000	2.042	1.047	-0.008	2.037	1.054	-0.077	1.948	0.976	-0.169	1.841	0.933
	SD	0.145	0.262	0.388	0.173	0.316	0.486	0.250	0.442	0.609	0.341	0.572	0.997
3	Mean	-0.059	1.997	1.021	-0.115	1.941	1.006	-0.197	1.876	1.015	-0.263	1.870	1.165
	SD	0.260	0.472	0.730	0.328	0.602	0.962	0.428	0.771	1.448	0.539	0.978	1.946
4	Mean	-0.085	2.042	1.278	-0.115	2.058	1.421	-0.139	2.173	1.942	-0.219	2.221	2.471
	SD	0.378	0.827	1.950	0.446	0.980	2.550	0.568	1.423	4.725	0.690	1.837	7.173

Kaplan-Meier curve. A practical difficulty arises when there are multiple continuous covariates in Z. To assess the dimensionality, we set $n = 300$ and added $p-1$ additional independent covariates $Z_2, \ldots, Z_p$, each standard uniform, to the model. Thus, the true coefficients associated with the added covariates were zero at the median. For simplicity, we took the bandwidth $h_n = 0.3$, 0.6, 0.8 for $p = 2, 3$, and 4 respectively. The results are summarized in Table 5 when γ is estimated. Overall, the proposed method performs satisfactorily when either the dimension is small or the censoring rate is not high. We also see that the biases and the sample standard deviations both increased when either dimensionality p or the censoring increased. We also see that the coverage probabilities are not satisfactory,when estimating a multi-dimensional smooth function.

5. Data Analysis

We illustrate the proposed method by analyzing the HMO data in Hosmer and Lemeshow (1999). Information was available for 100 HIV positive subjects who were followed until death from AIDS or AIDS-related complications, to the end of the study or until the subject was lost to follow-up. The primary outcome was survival time after a confirmed diagnosis of HIV. Covariates information was available on Age (in years) and Drug indicating prior drug use (1=yes, 0=no). There were 20 censored subjects. The scatter plot of the data is shown in Figure 1, for which the majority of the survival times are small. We dichotomized the data according to Drug and applied the local Kaplan-Meier estimator separately for each drug category. In order to choose the optimal bandwidth, we used 10 fold cross validation. We limited the range of γ from -2 to 2.

By fitting a simple Cox model treating the censoring time as survival time, we note that censoring time is highly dependent on Drug (p-value=0.007) and may depend on Age (p-value=0.10) based on a twenty observed censoring times. In addition, motivated by Figure 1, we created a 2 by 2 table using the censoring indicator and whether the observed time was above 40 as the classification variables. Appling Fisher's exact test for testing independence, we found the p-value

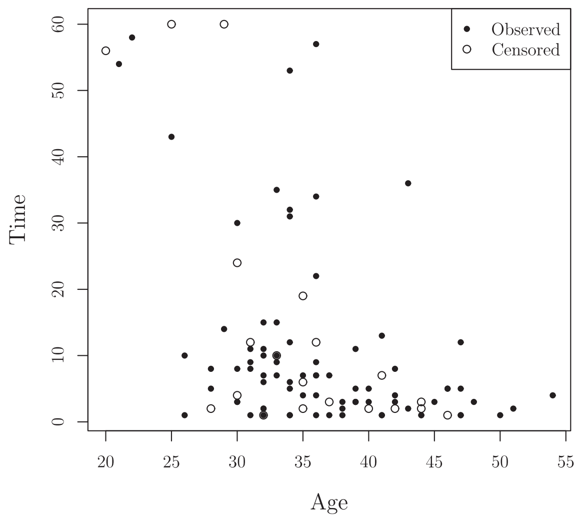

Figure 1. The scatter plot of the HMO study.

0.032, suggesting that C and T may not be independent. To explore the effects of the covariates on different quantiles of the survival time, we considered a series of quantiles ranging from 0.2 to 0.5 with an increment of 0.05. This quantile region is of interest since lower quantiles of the survival time pose an immediate concern to HIV subjects, and have significant biomedical implications in the near term. We used the bootstrap method to re-sample 300 times for any approach we use. We also tried larger quantiles, for example τ as large as 0.9. We found that sometimes the estimate with a bootstrap sample deviated from that of the original sample by a large magnitude, thus giving huge standard errors.

The top two plots in Figure 2 suggest that a log transformation of the survival time is reasonable to analyze this data set. We also see that the confidence intervals for our approach in estimating the transformation parameter are much smaller than those of Yin, Zeng, and Li (2008). Using $\gamma = 0$ as the transformation parameter, we plot in Figure 2 the estimated quantile regression coefficients β in the middle and the bottom panels, together with Peng and Huang's estimates. All three approaches indicate that both Drug and Age are significant across the quantiles we examine. Our results agree with the usual Cox model, for which both covariates are highly significant.

We plot the estimated quantile function $Z'\beta(\tau)$ of the survival time in Figure 3 given Drug (=1) and Age (=35), when the transformation parameter γ is estimated (Mu and He (2007)). The pointwise confidence intervals were computed via bootstrap. We find considerable difference between the two approaches.

When the transformation parameter γ is estimated, we assessed the effects of the covariates by examining their marginal effects. We follow Mu and He (2007)

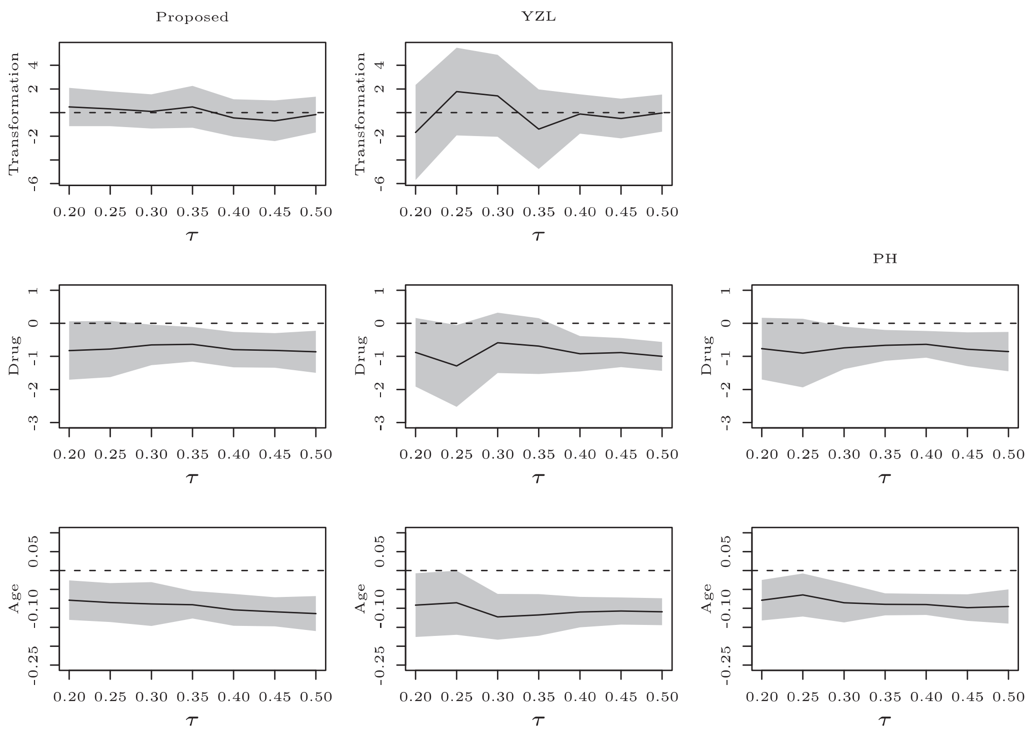

Figure 2. The HMO study: The top panels give the estimated transformation parameters with 95% confidence intervals; the middle and the bottom panels are the estimated linear quantile regression coefficients when $\gamma = 0$. Proposed: the proposed method; YZL: the method in Yin, Zeng, and Li (2008); PH: Peng and Huang's method.

and Yin, Zeng, and Li (2008) to take the marginal effects of the jth covariate at Z_0 as

$$\left.\frac{\partial \rho_{\gamma_\tau}(T|Z)}{\partial Z_{0,j}}\right|_{Z_0} = \begin{cases} \beta_{\tau,j}(\gamma_\tau \beta_\tau^T Z_0 + 1)^{1/\gamma_\tau - 1} & \gamma_\tau \neq 0, \\ \beta_{\tau,j} \exp(\beta_\tau^T Z_0) & \gamma_\tau = 0. \end{cases}$$

Applying the bootstrap re-sampling method to assess these effects, we plot the marginal effects of Drug (=1) and Age (=35) in Figure 4. We see that our proposed method gives tighter confidence intervals, and generally finds more significant quantiles.

6. Conclusion

We have proposed a new approach for linear quantile regression and studied the Box-Cox transformation approach to relate the survival time and the covariates via a linear quantile regression model. Our methodology is based on

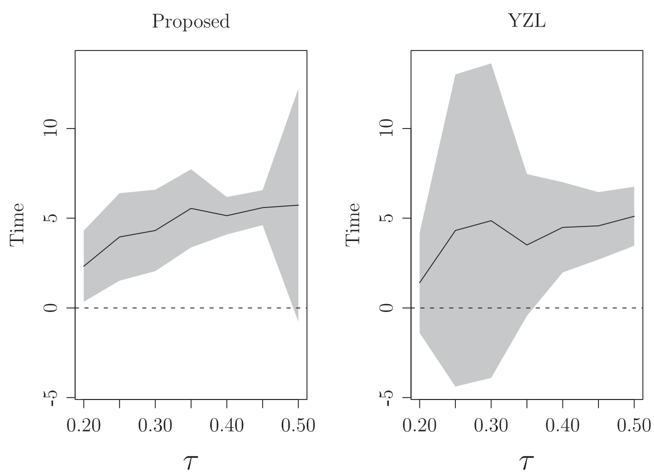

Figure 3. The estimated quantile function of the survival time given Drug (=1) and Age (=35).

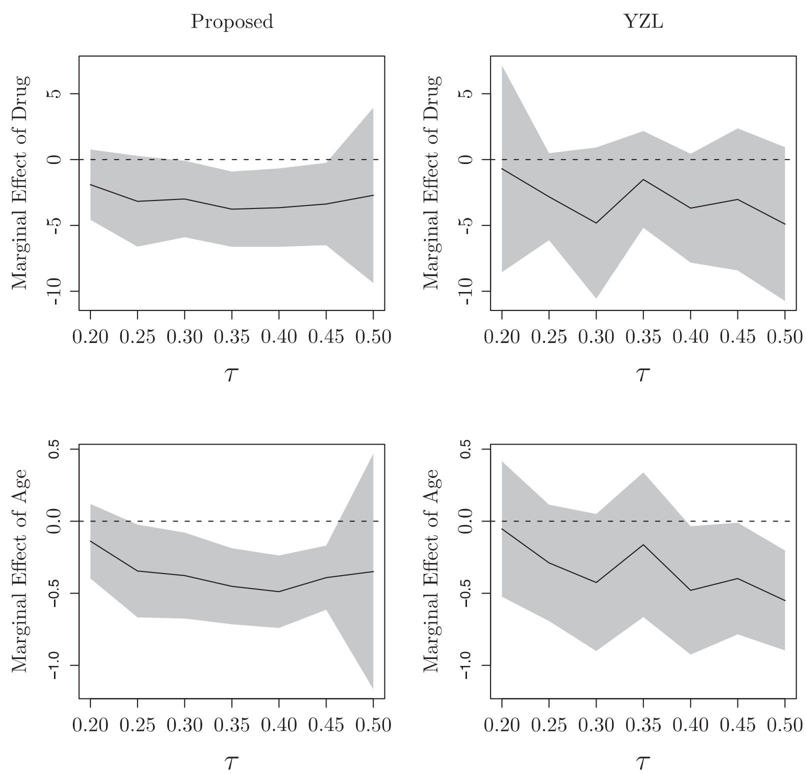

Figure 4. The estimated marginal effects given Drug (=1) and Age (=35).

formulating unbiased estimating equations using martingale residuals. We require no global linearity assumption and the failure and the censoring time only to be conditionally independent given the covariates. In addition, the estimation approach requires no iteration. The price to be paid is the need to estimate a distribution function nonparametrically, which we treat via using local kernel smoothing. Our method performs competitively compared to existing methods.

When the dimensionality of the covariates is high, we note that the estimates can have large biases and sampling standard errors due to the curse of dimensionality. In this case, it is worthwhile to consider model-based dimension reduction on $F(\cdot|Z)$ by, for example, modeling it as a single-index model (Lopez (2009); Wang, Zhou, and Li (2011)), or more simply by using Cox's model to relate the censoring time and the covariates. As an alternative, it is also interesting to develop model selection methods that can choose appropriate variables for the conditional CDF and the quantile regression function. In the data analysis, occassionally a bootstrap sample gave an unstable estimate. While this is not surprising for this small data set, a more promising method is to incorporate suitable random variables in the bootstrap inference (Jin, Ying, and Wei (2001)).

We have focused on continuous covariates mainly in this paper. An immediate extension to multiple discrete covariates is possible, but the estimation of the conditional CDF may suffer from small sample sizes due to an ANOVA-type splitting of the samples. It may be possible to simplify a full ANOVA-type decomposition of the samples by imposing some structural assumption similar to a main effect model for example.

Acknowledgement

Leng is also affiliated with Department of Statistics and Applied Probability, National University of Singapore where this research was supported. Tong is supported by NSF China Zhongdian Project 11131002 and NSFC (No. 11371062), BCMIIS, and the Fundamental Research Funds for the Central Universities.

Appendix

We write $\|\beta\|$ for the Euclidean norm of a finite-dimensional vector β and $\|G(\cdot)\|_\infty$ as the supreme of the absolute value of a function $G(\cdot)$. First, we state a lemma following Theorem 2.1 in Gonzalez-Manteiga and Cadarso-Suarez (1994), see also van Keilegom (1998). Here we focus on the univariate covariate case, $p = 1$, and defer the discussion for $p > 1$ to Appendix D.

Lemma A.1. *If* (C4)$-$(C6) *hold and* $p = 1$*, then*

$$\|\hat{F} - F_0\|_\infty = \sup_t \sup_z |\hat{F}(t|z) - F_0(t|z)| = O_p((\log n)^{1/2} n^{-1/2+v/2} + n^{-2v}).$$

Appendix A: Proof of Theorem 1.

Let $\widetilde{U}_n(\beta) = (1/n)\sum_{i=1}^n \{\Lambda_0(\exp(Z_i'\beta) \wedge Y_i|Z_i) - \Lambda_0(\exp(Z_i'\beta_0) \wedge Y_i|Z_i)\}Z_i$. We have

$$U_n(\beta) - \widetilde{U}_n(\beta) = \frac{1}{n}\sum_{i=1}^n Z_i\{N_i(\exp(Z_i'\beta)) - E[N_i(\exp(Z_i'\beta))|Z_i]\} \\ -\frac{1}{n}\sum_{i=1}^n Z_i\{H_\tau(\hat{F}(Y_i|Z_i)) - H_\tau(F_0(Y_i|Z_i))\}.$$

By Lemma A.1, Assumptions C2 and C3,

$$\frac{1}{n}\sum_{i=1}^n Z_i\{H_\tau(\hat{F}(Y_i|Z_i)) - H_\tau(F_0(Y_i|Z_i))\} = o(1), \tag{A.1}$$

when $0 < v < 1/2$. We show that

$$\sup_{\beta\in\mathcal{B}} |U_{n1}(\beta)| = \sup_{\beta\in\mathcal{B}} \frac{1}{n}|\sum_{i=1}^n Z_i\{N_i(\exp(Z_i'\beta)) - E[N_i(\exp(Z_i'\beta))|Z_i]\}| = o(1). \tag{A.2}$$

For any sufficiently small $\eta > 0$, we divide the parameter space $\mathcal{B}$ into $K = K(\eta) = O(\eta^{-p})$ disjoint balls with each radius proportional to η such that $P(\sup_{\beta\in\mathcal{B}} |U_{n1}(\beta)| > \eta) \le \sum_{j=1}^K P(|U_{n1}(\beta_j)| > \eta/2)$, where β_j is the center of the jth ball. By the Bernstein inequality, $\sum_{n=1}^\infty \sum_{j=1}^K P(|U_{n1}(\beta_j)| > \eta/2) < \infty$. An application of the Borel-Cantelli Theorem gives $\sup_{\beta\in\mathcal{B}} |U_{n1}(\beta)| = o(1)$ and thus

$$\sup_{\beta\in\mathcal{B}} |U_n(\beta) - \widetilde{U}_n(\beta)| = o(1), \quad a.s.. \tag{A.3}$$

From Assumption C7, the first derivative function $A_n(\beta)$ of $\widetilde{U}_n(\beta)$ with respect to β is

$$A_n(\beta) = \frac{1}{n}\sum_{i=1}^n I(Y \ge \exp(Z_i'\beta))Z_iZ_i' \exp(Z_i'\beta)\lambda_0(\exp(Z_i'\beta)|Z_i) \\ \to EZZ' \exp(Z'\beta)f_0(Z'\beta|Z)(1 - G_0(\exp(Z'\beta)|Z))$$

which is positive definite with probability one for β in the neighborhood of β_0. In addition, $\widetilde{U}_n(\beta_0) = 0$. Therefore, $\widetilde{U}_n(\beta)$ is bounded away from zero. This, together with (A.3), yields $\hat{\beta}_n \to \beta_0$ in probability as $n \to \infty$.

Appendix B: Proof of Theorem 2.

We exploit Theorem 2 of Chen, Linton, and van Keilegom (2003) by verifying their conditions (2.1)−(2.4), (2.5′) and (2.6′). For this, write $U_n(\beta, F) =$

$(1/n)\sum_{i=1}^{n} u_i(\beta, F)$ where $u_i(\beta, F) = Z_i\{N_i(\exp(Z_i'\beta)) - H_\tau(F(Y_i|Z_i))\}$ and the function class $\mathcal{F}$ that involves the true F_0 as

$$\mathcal{F} = \{F : F \text{ has a density function } f,\ \sup_{z\in\mathcal{Z}} F(\mathcal{T}|z) \le \eta_0,\ f \text{ satisfies C4}\}.$$

Then $U(\beta, F) = Eu_i(\beta, F) = EZ_i\{\Lambda_0(\exp(Z_i'\beta) \wedge Y_i|Z_i) - H_\tau(F(Y_i|Z_i))\}$. This, together with (2.3), implies that $U(\beta_0, F_0) = 0$. Write $\Gamma_1(\beta_0, F_0)$ as the first derivative function of $U(\beta, F_0)$ with respect to β evaluated at $\beta = \beta_0$, and for all $\beta \in \mathcal{B}$, define the functional derivative of $U(\beta, F)$ at F_0 in the direction $[F - F_0]$ as $\Gamma_2(\beta, F_0)[F - F_0] = \lim_{\eta\to 0}(1/\eta)[U(\beta, F_0 + \eta(F - F_0)) - U(\beta, F_0)]$.

Proof of Theorem 2. Condition (2.1) of Chen, Linton, and van Keilegom (2003) can be easily verified by the subgradient condition of quantile regression (Koenker (2005)). Conditions (2.4), (2.5′), and (2.6) hold directly by Lemma A.1, A.2, and A.3, respectively. From the definition of Γ_1,

$$\Gamma_1 = \left.\frac{\partial U(\beta, F_0)}{\partial \beta}\right|_{\beta=\beta_0} = EZZ' \exp(Z'\beta_0) f_0(\exp(Z'\beta_0)|Z)(1 - G_0(\exp(Z'\beta_0)|Z)),$$

which is positive definite by C7. This means that (2.2) in Chen, Linton, and van Keilegom (2003) holds. By routine Taylor expansions, we can verify their (2.3). Therefore, $n^{1/2}(\hat\beta_n - \beta_0) \xrightarrow{d} N(0, \Gamma_1^{-1} V_1 \Gamma_1^{-1})$.

Lemma A.2. *For all positive sequences $\xi_n = o(1)$, we have*

$$\sup_{\|\beta-\beta_0\|\le\xi_n, \|F-F_0\|_\infty\le\xi_n} \|U_n(\beta, F) - U(\beta, F) - U_n(\beta_0, F_0)\| = o_p(n^{-1/2}).$$

Proof. Let $\eta_1 = \sup_{z\in\mathcal{Z}} \|z\|^2$ and $\eta_2 = \sup_{F\in\mathcal{F}, z\in\mathcal{Z}, t\le\mathcal{T}}(f(t|z) + g_0(t|z))$. For any $(\beta, F) \in \mathcal{B}\times\mathcal{F}$ and $(\beta^*, F^*) \in \mathcal{B}\times\mathcal{F}$, we have $\|u(\beta, F) - u(\beta^*, F^*)\|^2 \le 2(B_1 + B_2)$, where

$$\begin{aligned} B_1 &= \|Z\delta\{I(Y \ge \exp(Z'\beta)) - I(Y \ge \exp(Z'\beta^*))\}\|^2 \\ &\le \eta_1 |I(Y \ge \exp(Z'\beta)) - I(Y \ge \exp(Z'\beta^*))|, \\ B_2 &= \|Z(H(\tau \wedge F(Y|Z)) - H(\tau \wedge F^*(Y|Z))\|^2 \\ &\le \frac{\eta_1}{(1-\tau)^2}\|(\tau \wedge F(Y|Z)) - (\tau \wedge F^*(Y|Z))\|_\infty^2 \le \frac{\eta_1}{(1-\tau)^2}\|F - F^*\|_\infty^2. \end{aligned}$$

Direct calculation yields that $E(\sup_{\|\beta-\beta^*\|\le\xi_n} B_1) \le \eta_3\xi_n$ and $\sup_{\|\beta-\beta^*\|\le\xi_n, \|F-F^*\|_\infty\le\xi_n} \|U(\beta, F) - U(\beta^*, F^*)\|^2 \le \eta_4\xi_n$ for some positive constants, η_3, η_4, when n is sufficiently large. Therefore, (3.2) in Chen, Linton, and van Keilegom (2003) holds with $r = 2$ and $s_j = 1/2$. Condition (3.1) in

Chen, Linton, and van Keilegom (2003) follows clearly if we set the term in their (3.1) to 0.

Now we verify their (3.3). Let $N(\eta, \mathcal{F}, \|\cdot\|_\infty)$ be the covering numbers (van der Vaart and Wellner (1996, p.83)) for the function class $\mathcal{F}$ under the metrics $\|\cdot\|_\infty$. An application of Theorem 2.7.1 in van der Vaart and Wellner (1996, p.159) and the compactness of $\mathcal{Z}$ give that the logarithm of the covering number is bounded by $O(\eta^{-1/2})$ for $\eta \le 1$ and $= 0$ for $\eta > 1$, which yields that

$$\int_0^\infty \{\log N(\eta^2, \mathcal{F}, \|\cdot\|_\infty)\}^{1/2} d\eta \le O(1) \int_0^1 \eta^{-1/2} d\eta < \infty.$$

It then follows easily from Theorem 3 of Chen, Linton, and van Keilegom (2003) that Lemma A.2 holds.

Lemma A.3. *Assume that the conditions in Theorem* 2 *hold, then*

$$n^{1/2}(U_n(\beta_0, F_0) + \Gamma_2(\beta_0, F_0)[\hat{F} - F_0]) \xrightarrow{d} N(0, V_1),$$

where $V_1 = \mathrm{Cov}\,(v_i)$ *with*

$$v_i = u_i(\beta_0, F_0) - Z_i f_Z(Z_i) \int_0^{\exp(Z_i'\beta_0)} \psi(Y_i, \delta_i; t, Z_i) \Big\{ \frac{1 - G_0(t|Z_i)}{1 - F_0(t|Z_i)} f_0(t|Z_i) + g_0(t|Z_i) \Big\} dt,$$

$$\psi(Y_i, \delta_i; t, z) = \{1 - F_0(t|z)\} \Bigg[\int_0^{Y_i \wedge t} \frac{-f_0(s|z) ds}{\{1 - F_0(s|z)\}^2 \{1 - G_0(s|z)\}} + \frac{\delta_i I(Y_i \le t)}{\{1 - F_0(Y_i|z)\}\{1 - G_0(Y_i|z)\}} \Bigg].$$

Proof. By the definition of Γ_2, a direct calculation yields that

$$\Gamma_2(\beta_0, F_0)[F - F_0] = -EZI(F_0(Y|Z) \le \tau) \frac{F(Y|Z) - F_0(Y|Z)}{1 - F_0(Y|Z)}. \tag{A.4}$$

For any function $\hbar$, the conditional expectation of $\hbar(Y)$ given $Z = z$ takes the form

$$\begin{aligned} E\{\hbar(Y)|z\} &= E\{\hbar(T) I(T \le C) + \hbar(C) I(T \ge C)|z\} \\ &= \int \hbar(t)\{(1 - G_0(t|z)) f_0(t|z) + (1 - F_0(t|z)) g_0(t|z)\} dt. \end{aligned} \tag{A.5}$$

From Theorem 2.3 of Gonzalez-Manteiga and Cadarso-Suarez (1994) and the proof of Theorem 2 in Wang and Wang (2009), under Assumptions C3−C7, we have that

$$\hat{F}(t|z) - F_0(t|z) = \frac{1}{nh_n} \sum_{i=1}^n K\Big(\frac{z - Z_i}{h_n}\Big) \psi(Y_i, \delta_i; t, z) + O_p\Big(\Big\{\frac{\log n}{nh_n}\Big\}^{3/4} + h_n^2\Big). \tag{A.6}$$

This result holds only when there is a single covariate. When $1/4 < v < 1/3$, the residual term on the right hand side of (A.6) is $o_p(n^{-1/2})$. Plugging (A.6) into (A.4) through (A.5), and using standard change of variables and Taylor expansion arguments, and the assumption that $\int K(u)du = 1$, one obtains

$$\begin{aligned}
&\Gamma_2(\beta_0, F_0)[\hat{F} - F_0] \\
&\quad = -\frac{1}{n}\sum_{i=1}^{n} Z_i f_Z(Z_i) \int_0^{\exp(Z_i'\beta_0)} \psi(Y_i, \delta_i; t, Z_i)\left\{\frac{1 - G_0(t|Z_i)}{1-F_0(t|Z_i)} f_0(t|Z_i) + g_0(t|Z_i)\right\} dt \\
&\qquad + o_p(n^{-1/2}).
\end{aligned}$$

Therefore, $n^{1/2}(U_n(\beta_0, F_0) + \Gamma_2(\beta_0, F_0)[\hat{F} - F_0]) = n^{-1/2}\sum_{i=1}^{n} v_i + o_p(1)$. An application of the Central Limit Theorem gives this lemma.

Appendix C: Proof of Theorem 3.

We introduce the following lemma from Mu and He (2007) that specifies the identifiability condition.

Lemma A.4. *The model parameter is identifiable in the sense that if* $\rho_{\gamma^*}^{(-1)}(Z'\beta^*) = \rho_{\gamma_0}^{(-1)}(Z'\beta_0)$ *almost surely, then* $\beta^* = \beta_0$, $\gamma^* = \gamma_0$.

To facilitate the proof, we set $V(\beta;\gamma) = \delta I(\rho_\gamma(Y) - Z'\beta \le 0) - H_\tau(F_0(Y|Z))$, $V_\beta(Z) = \partial E\{V(\beta;\gamma)|Z\}/\partial\beta|_{\beta=\beta_0,\gamma=\gamma_0}$, and $V_\gamma = \partial E\{V(\beta;\gamma)|Z\}/\partial\gamma|_{\beta=\beta_0,\gamma=\gamma_0}$. Note that $E\{ZV_\beta(Z)'\} = E[ZZ'\dot{\rho}_{\gamma_0}^{(-1)}(Z'\beta_0) f_0(\rho_{\gamma_0}^{(-1)}(Z'\beta_0)|Z)(1 - G_0(\rho_{\gamma_0}^{(-1)}(Z'\beta_0)|Z))]$ is positive definite from assumption (C7$'$). Let $\widetilde{R}_n(\beta;\gamma) = (1/n)\sum_{i=1}^{n} Z_i'\beta\{\delta_i I(\rho_\gamma(Y_i) - Z_i'\beta \le 0) - H_\tau(\hat{F}(Y_i|Z_i))\}$ and $\widetilde{R}(\beta;\gamma)$ be the limit of $\widetilde{R}_n(\beta;\gamma)$ as n tends to infinity.

Proof of consistency. Since the minimization of $R_n^*(\gamma)$ is taken over Υ and Υ is compact, there exists a subsequence indexed by n, such that $\hat{\gamma}_n \to \gamma^*$.

Step 1. We prove that $\sup_n \|\hat{\beta}_n(\hat{\gamma}_n)\| < \infty$. Otherwise, for a subsequence, still indexed by n, $\|\hat{\beta}_n(\hat{\gamma}_n)\| \to \infty$. Let

$$\hat{\beta}_n^* = (1 - \frac{1}{\|\hat{\beta}_n(\hat{\gamma}_n) - \beta(\gamma^*)\|})\beta(\gamma^*) + \frac{1}{\|\hat{\beta}_n(\hat{\gamma}_n) - \beta(\gamma^*)\|}\hat{\beta}_n(\hat{\gamma}_n).$$

Then one can easily verify that $\|\hat{\beta}_n^* - \beta(\gamma^*)\| = 1$ and $\|\hat{\beta}_n^*\| \le \|\hat{\beta}_n^* - \beta(\gamma^*)\| + \|\beta(\gamma^*)\| = 1 + \|\beta(\gamma^*)\|$ is bounded. Therefore, there exists a subsequence, still indexed by n, such that $\hat{\beta}_n^* \to \beta^*$. By convexity,

$$\begin{aligned}
\widetilde{R}_n(\hat{\beta}_n^*, \hat{\gamma}_n) &\leq (1 - \frac{1}{\|\hat{\beta}_n(\hat{\gamma}_n) - \beta(\gamma^*)\|})\widetilde{R}_n(\beta(\gamma^*), \hat{\gamma}_n) \\
&\quad + \frac{1}{\|\hat{\beta}_n(\hat{\gamma}_n) - \beta(\gamma^*)\|}\widetilde{R}_n(\hat{\beta}_n(\hat{\gamma}_n), \hat{\gamma}_n) \\
&= \widetilde{R}_n(\beta(\gamma^*), \hat{\gamma}_n) - \frac{1}{\|\hat{\beta}_n(\hat{\gamma}_n) - \beta(\gamma^*)\|}(\widetilde{R}_n(\beta(\gamma^*), \hat{\gamma}_n) - \widetilde{R}_n(\hat{\beta}_n(\hat{\gamma}_n), \hat{\gamma}_n)) \\
&\leq \widetilde{R}_n(\beta(\gamma^*), \hat{\gamma}_n),
\end{aligned}$$

since $\hat{\beta}_n(\hat{\gamma}_n)$ minimizes $\widetilde{R}_n(\beta, \hat{\gamma}_n)$. Note that the class of function $\{Z'\beta(\delta I(\rho_\gamma(Y) - Z'\beta \leq 0) - H_\tau(F_0(Y|Z))) : \gamma \in \Upsilon, \|\beta - \beta(\gamma^*)\| = 1, z \in \mathcal{Z}\}$ is a Glivenko-Cantelli class. This, together with Lemma A.1, yields that $\widetilde{R}(\beta^*, \gamma^*) \leq \widetilde{R}(\beta(\gamma^*), \gamma^*)$. This contradicts with the fact that $\beta(\gamma^*)$ is the unique maximizer of $\widetilde{R}(\beta, \gamma)$ for $\gamma = \gamma^*$, since $\|\beta^* - \beta(\gamma^*)\| = 1$ and thus $\beta^* \neq \beta(\gamma^*)$, so $\hat{\beta}_n$ must be bounded. Therefore, there exists a subsequence, indexed by n, for which $\hat{\beta}_n \to \beta^*$.

Step 2. We prove that $\gamma^* = \gamma_0$ and $\beta^* = \beta_0$. Note that the class of function

$$\{I(Z \leq z)(\delta I(\rho_\gamma(Y) - Z'\beta \leq 0) - H_\tau(F_0(Y|Z))) : \gamma \in \Upsilon, z \in \mathcal{Z}, \beta \in \mathcal{B}\}$$

is a Glivenko-Cantelli class. An application of Glivenko-Cantelli Theorem gives

$$R_n^*(\hat{\gamma}_n) \to E(\|E[I(Z_i \leq z)(\delta_i I(\rho_{\gamma^*}(Y_i) - Z_i'\beta(\gamma^*) \leq 0) - H_\tau(F_0(Y_i|Z_i)))]\|^2|_{z=Z})$$

and $R_n^*(\gamma_0) \to 0$. Since $\hat{\gamma}_n$ minimizes $R_n^*(\gamma)$, we have $R_n^*(\gamma_0) \geq R_n^*(\hat{\gamma}_n)$. Letting $n \to \infty$, one obtains

$$E[I(Z_i \leq z)(\delta_i I(\rho_{\gamma^*}(Y_i) - Z_i'\beta(\gamma^*) \leq 0) - H_\tau(F_0(Y_i|Z_i)))]|_{z=Z} = 0, \text{ a.s. for all } Z.$$

This yields that $\rho_{\gamma^*}^{(-1)}(Z'\beta(\gamma^*)) = \rho_{\gamma_0}^{(-1)}(Z'\beta(\gamma_0))$ almost surely and thus $\gamma^* = \gamma_0$ and $\beta(\gamma^*) = \beta_0$ derived from Lemma A.4. The consistency of $\hat{\gamma}_n$ and $\hat{\beta}_n(\hat{\gamma}_n)$ follows by an application of Helly's selection theorem.

Proof of asymptotic normality. Using the inverse mapping theorem, one has

$$\beta(\gamma) = \beta_0 - \{E(ZV_\beta(Z))\}^{-1}E\{ZV_\gamma(Z)\}(\gamma - \gamma_0) + o(|\gamma - \gamma_0|). \tag{A.7}$$

Note that

$$\begin{aligned}
&\{I(Z \leq z)\{\delta I(\rho_\gamma(Y) \leq Z'\beta) - H_\tau(F(Y|Z))\} : \\
&\qquad \gamma \in \text{neighborhood of } \gamma_0,\ F \in \mathcal{F},\ z \in \mathcal{Z},\ \beta \in \mathcal{B}\}
\end{aligned}$$

is a Donsker class. Therefore

$$\sup_{z,\gamma\in\Upsilon} |R_n(z;\gamma) - E[I(Z \le z)\{\delta I(\rho_\gamma(Y) \le Z'\hat{\beta}_n(\gamma)) - H_\tau(\hat{F}(Y|Z))\}] \\ -(\mathcal{P}_n - \mathcal{P})[I(Z \le z)\{\delta I(\rho_\gamma(Y) \le Z'\beta(\gamma)) - H_\tau(F_0(Y|Z))\}]| = o_p(n^{-1/2}),$$

where $\mathcal{P}$ is the expectation operator, $\mathcal{P}_n$ is the empirical measure. Applying (A.6), after some basic calculations, one has

$$\begin{aligned} R_n(z;\gamma) = {} & E\{I(Z \le z)(\delta I(\rho_\gamma(Y) \le Z'\beta(\gamma)) - H_\tau(F_0(Y|Z)))\} \\ & + E\{I(Z \le z)V_\beta'(Z)\}(\hat{\beta}_n(\gamma) - \beta(\gamma)) + (\mathcal{P}_n - \mathcal{P})V_1^*(Y,\delta,Z;\gamma,z) \\ & + o_p(\|\hat{\beta}_n(\gamma) - \beta(\gamma)\| + |\gamma - \gamma_0| + n^{-1/2}), \end{aligned}$$

where

$$\begin{aligned} V_1^*(Y,\delta,Z;\gamma,z) = {} & I(Z \le z)\{\delta I(\rho_\gamma(Y) \le Z'\beta(\gamma)) - H_\tau(F_0(Y|Z))\} \\ & - \widetilde{E}\Big\{\frac{f_Z(Z)I(Z \le z)I(F_0(\tilde{Y}|Z) \le \tau)\psi(Y,\delta,Z;\tilde{Y},Z)}{1 - F_0(\tilde{Y}|Z)}\Big\}, \end{aligned}$$

$(\tilde{Y},\tilde{\delta},\tilde{Z})$ is an i.i.d. copy of (Y,δ,Z) and $\tilde{E}$ is the expectation with respect to the joint distribution of $(\tilde{Y},\tilde{\delta},\tilde{Z})$. Let $V_2^*(z) = E\{I(Z \le z)V_\gamma(Z)\} - E\{I(Z \le z)V_\beta(Z)'\}\{E(ZV_\beta(Z)')\}^{-1}E\{ZV_\gamma(Z)\}$, $V_3^*(z) = E\{I(Z \le z)V_\beta(Z)'\}$, $V_4^* = \tilde{E}\{V_2^*(\tilde{Z})'V_1^*(Y,\delta,Z;\gamma_0,\tilde{Z})\}$, $V_5^* = E\{V_2^*(Z)'V_3^*(Z)\}$ and $V_6^* = EV_2^*(Z)'V_2^*(Z) + V_5^*\{E(ZV_\beta(Z)')\}^{-1}E\{ZV_\gamma(Z)\}$.

Using the arguments in Yin, Zeng, and Li (2008), for any γ in neighborhood of γ_0, $R_n^*(\gamma)$ has a quadratic expansion around γ_0 and

$$\begin{aligned} & E\{V_2^*(Z)'V_2^*(Z)\}(\hat{\gamma}_n - \gamma_0) + E\{V_2^*(Z)'V_3^*(Z)\}(\hat{\beta}_n(\hat{\gamma}_n) - \beta(\hat{\gamma}_n)) \\ & \quad = -(\mathcal{P}_n - \mathcal{P})V_4^* + o_p(\|\hat{\beta}_n(\hat{\gamma}_n) - \beta(\hat{\gamma}_n)\| + |\hat{\gamma}_n - \gamma_0| + n^{-1/2}), \end{aligned}$$

since $\hat{\gamma}_n$ minimizes $R_n^*(\gamma)$. Therefore,

$$V_5^*(\hat{\beta}_n(\hat{\gamma}_n) - \beta_0) + V_6^*(\hat{\gamma}_n - \gamma_0) = -(\mathcal{P}_n - \mathcal{P})V_4^* + o_p(\|\hat{\beta}_n(\hat{\gamma}_n) - \beta_0\| + |\hat{\gamma}_n - \gamma_0| + n^{-1/2}). \tag{A.8}$$

Noting that $\{Z\{\delta I(\rho_\gamma(Y) \le Z'\beta) - H_\tau(F(Y|Z))\} : \gamma \in$ neighborhood of γ_0, $F \in \mathcal{F}$, $\beta \in \mathcal{B}\}$ is a Donsker class, one has

$$\begin{aligned} & E\{ZV_\beta(Z)'\}(\hat{\beta}_n(\hat{\gamma}_n) - \beta_0) + E\{ZV_\gamma(Z)\}(\hat{\gamma}_n - \gamma_0) \\ & \quad = -(\mathcal{P}_n - \mathcal{P})V_7^* + o_p(\|\hat{\beta}_n(\hat{\gamma}_n) - \beta_0\| + |\hat{\gamma}_n - \gamma_0| + n^{-1/2}) \end{aligned} \tag{A.9}$$

and $V_7^* = Z\{\delta I(\rho_\gamma(Y) \le Z'\beta(\gamma)) - H_\tau(F_0(Y|Z))\} - Zf_Z(Z)\widetilde{E}\{I(F_0(\tilde{Y}|Z) \le \tau)\psi(Y,\delta,Z;\tilde{Y},Z)/(1 - F_0(\tilde{Y}|Z))\}$. Combining (A.8) and (A.9), one obtains that

$$\Gamma_2\sqrt{n}(\hat{\theta}_n - \theta_0) = \sqrt{n}(\mathcal{P}_n - \mathcal{P})V^*(Y,\delta,Z) + o_p(1),$$

where

$$\Gamma_2 = \begin{pmatrix} E\{ZV_\beta(Z)'\} & E\{ZV_\gamma(Z)\} \\ V_5^* & E\{V_2^*(Z)'V_2^*(Z)\} + V_5^*\{E(ZV_\beta(Z))\}^{-1}E\{ZV_\gamma(Z)\} \end{pmatrix},$$

and $\hat{\theta}_n = (\hat{\beta}_n(\hat{\gamma}_n)', \hat{\gamma}_n)'$, $\theta_0 = (\beta_0', \gamma_0)'$, $V^*(Y, \delta, Z) = -(V_7^{*'}, V_4^*)'$. Note that

$$\Gamma_2 = \begin{pmatrix} E\{ZV_\beta(Z)'\} & 0 \\ E\{V_2^*(Z)'V_3^*(Z)\} & EV_2^*(Z)'EV_2^*(Z) \end{pmatrix} \begin{pmatrix} I_p & \{E(ZV_\beta(Z)')\}^{-1}E\{ZV_\gamma(Z)\} \\ 0 & 1 \end{pmatrix}$$

and thus Γ_2 is invertible. Using the Multivariate Central Limit Theorem, we have proved Theorem 3 with $V_2 = \mathrm{Cov}\,(V^*)$.

Appendix D: Multiple covariates

We extend the asymptotic results to higher dimensional covariates. The difficulty in deriving good properties of the proposed estimator for $p > 1$ lies in the fact that the estimation of the distribution function $F(\cdot|z)$ may have larger bias and slower convergence rate. Note that the results in Wang and Wang (2009) only apply for one dimensional problems. To our knowledge, there are few papers that study the properties of the local Kaplan-Meier estimates with multiple covariates. Dabrowska (1989) considered this problem and obtained uniform consistency. Lopez (2009) considered the nonparametric estimation of the multivariate distribution and assumed a known function $g : R^p \to R$, such that T and C are conditionally independent given $g(Z)$, and the dependence of T and C on Z is only through $g(Z)$. This method effectively reduces the dimensionality at the expense of the need to determine g.

Consistency. In the proof of Theorem 1, (A.3) is needed to prove the consistency. And (A.2) holds regardless of p. Thus, a sufficient condition for (A.1) is

$$\sup_{t,z} |\hat{F}(t|z) - F_0(t|z)| = o_p(1). \tag{A.10}$$

To prove the consistency of the proposed estimator for general p-dimensional covariates, one only needs have (A.10).

Dabrowska (1989) considered the kernel conditional Kaplan-Meier estimate of the distribution function in presence of right censoring and obtained uniform consistency. We assume that the following regularity conditions:

C5$'$. The bandwidth $h_n = O(n^{-v})$, with $0 < v < 1/p$.

C6$'$. (i) K is a kernel with bounded compact support and total variation. (ii) The kernel K satisfies $\int z_j K(z)dz = 0$, $j = 1, \ldots, p$.

By Corollary 2.1 in Dabrowska (1989), (A.10) holds. Therefore, the asymptotic consistency in Theorem 1 holds when the conditions C5 and C6 are replaced by C5′ and C6′ for multi-dimensional problems.

Asymptotic Normality. From the proofs of Theorems 2 and 3, the key condition for establishing the asymptotic normality is (A.6), which holds only for $p = 1$ with a single continuous covariate. For the multi-dimensional case, (A.6) does not hold. However, a similar representation of $\hat{F}(t|z) - F_0(t|z)$ has (Liang, de Uña-Álavarez, and Iglesias-Pérez (2010))

$$\hat{F}(t|z) - F_0(t|z) = \frac{1}{nh_n^p} \sum_{i=1}^{n} K\left(\frac{z - Z_i}{h_n}\right) \psi(Y_i, \delta_i; t, z) + r_n(t, z) \tag{A.11}$$

and, according to the proof of Theorem 2, we need to ensure that $r_n(t, z)$ satisfies

$$\sup_{t,z} |r_n(t, z)| = o_p(n^{-1/2}). \tag{A.12}$$

For the one dimensional case, under some regularity conditions, the remainder term is $r_n(t, z) = O_p((\log n/nh_n)^{3/4} + h_n^2)$. The condition $h_n = O(n^{-v})$ with $v \in (1/4, 1/3)$ ensures (A.12). However, this condition on h_n is no longer sufficient. To this end, we turn to higher order kernels (Müller (1988)).

C6″. (i) K is a bounded kernel function with bounded compact support in R^p. (ii) The kernel K has order q, which satisfies $\int K(z)dz = 1$ and

$$\int z_1^{u_1} \cdots z_p^{u_p} K(z_1, \ldots, z_p) dz \begin{cases} = 0, & \text{if } \ 0 \neq \sum_{j=1}^{p} u_j < q, \\ \neq 0, & \text{if } \ \sum_{j=1}^{p} u_j = q. \end{cases}$$

C4″. The first q partial derivatives with respect to z of the density function $f_Z(z)$ are uniformly bounded for $z \in \mathcal{Z}$, and $f_0(t|z)$ and $g_0(t|z)$ are uniformly bounded away from infinity and have bounded (uniformly in t) first q order partial derivatives with respect to z.

C5″. The bandwidth satisfies $h_n = O(n^{-v})$ with $1/2q < v < 1/3p$.

A kernel function K that satisfies C6″ is a higher order kernel with order q. If $q = 2$, the kernel K is the familiar second-order kernel (Silverman (1986); Dabrowska (1997)). Liang, de Uña-Álavarez, and Iglesias-Pérez (2010) studied the asymptotic properties of the conditional distribution estimator under truncated, censored, and dependent data for a general p. We can apply their Theorem 2.

Corollary 1. *Under conditions* C1, C2, C3 (*or* C3′), C4″-C6″, (A.11), *and* (A.12) *hold with*

$$\sup_{t,z} |r_n(t,z)| = O_p\left(\left[\frac{\log n}{nh_n^p}\right]^{3/4} + h_n^q\right) = o_p(n^{-1/2}).$$

Theorem 4. *Under* C1, C2, C3, C4″−C6″, *and* C7*, Theorem* 2 *holds; under* C1, C2, C3′, C4″−C6″, C7′*, and* C8*, Theorem* 3 *holds.*

Remark 4. From Theorem 4, one can see that: When $p = 1$ and $q = 2$, we need $1/4 < v < 1/3$ for Theorems 2 and 3; The order of the kernel needed for asymptotic normality increases with the dimensionality of the covariates. In the simulation, we have used higher order kernels with $q = 4$ for $p = 2$, $q = 5$ for $p = 3$, and $q = 7$ for $p = 4$ (Müller (1988)).

References

Box, G. E. P. and Cox, D. R. (1964). An analysis of transformation (with discussion). *J. Roy. Statist. Soc. Ser. B* **26**, 211-252.

Cai, T., Tian, L. and Wei, L. J. (2005). Semiparametric Box-Cox power transformation models for censored survival observations. *Biometrika* **92**, 619-632.

Chen, X., Linton, O. and van Keilegom, I. (2003). Estimation of semiparametric models when the criterion function is not smooth. *Econometrica* **71**, 1591-1608.

Dabrowska, D. M. (1989). Uniform consistency of the kernel conditional Kaplan-Meier estimate. *Ann. Statist.* **17**, 1157-1167.

Dabrowska, D. M. (1997). Smoothed Cox regression. *Ann. Statist.* **25**, 1510-1540.

Gonzalez-Manteiga, W. and Cadarso-Suarez, C. (1994). Asymptotic properties of a generalized Kaplan-Meier estimator with some applications. *J. Nonparametr. Stat.* **4**, 65-78.

Hall, P. (1992). *The Bootstrap and Edgeworth Expansion.* Springer, New York.

He, X. and Zhu, L. X. (2003). A lack of fit test for quantile regression. *J. Amer. Statist. Assoc.* **98**, 1013-1022.

Hosmer, D. W. and Lemeshow, S. (1999). *Applied Survival Analysis: Regression Modeling of Time to Event Data.* Wiley, New York.

Huang, Y. (2010). Quantile calculus and censored regression. *Ann. Statist.* **38**, 1607-1637.

Kalbfleisch, J. and Prentice, R. (2002). *The Statistical Analysis of Failure Time Data.* Wiley, New York.

Koenker, R. (2005). *Quantile Regression.* Cambridge University Press, Cambridge, UK.

Koenker, R. and Bassett, G. (1978). Regression quantiles. *Econometrica* **50**, 1577-1584.

Koenker, R. and Geling, O. (2001). Reappraising medfly longevity: A quantile regression survival analysis. *J. Amer. Statist. Assoc.* **96**, 458-468.

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

Liang, H., de Uña-Álavarez, J., and Iglesias-Pérez, M. (2010). Asymptotic properties of conditional distribution estimator with truncated, censored and dependent data. *Submitted*.

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

Lin, D. Y., Wei, L. J. and Ying, Z. (2002). Model-Checking techniques based on cumulative residuals. *Biometrics*, **58**, 1-12.

Lopez, O. (2009). Nonparametric estimation of the multivariate distribution function in a censored regression model with applications. *Comm. Statist. Theory Methods*, to appear.

Mu, Y. and He, X. (2007). Power transformation toward a linear regression quantile. *J. Amer. Statist. Assoc.* **102**, 269-279.

Müller, H.G. (1988). *Nonparametric Regression Analysis of Longitudinal Data.* Springer-Verlag, New York.

Peng, L. and Huang, Y. (2008). Survival analysis with quantile regression models. *J. Amer. Statist. Assoc.* **103**, 637-649.

Portnoy, S. (2003). Censored regression quantiles. *J. Amer. Statist. Assoc.* **98**, 1001-1012.

Portnoy, S. and Koenker, R. (1997). The Gaussian hare and the Laplacian tortoise: Computability of squared-error vs. absolute-error estimators. *Statist. Sci.* **12**, 279-300.

Powell, J. (1986). Censored regression quantiles. *J. Econom.* **32**, 143-155.

Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis.* Chapman and Hall, London.

van der Vaart, A. W. and Wellner, J. (1996). *Weak Convergence and Empirical Processes: with Applications to Statistics.* Springer, New York.

van Keilegom, I. (1998). Nonparametric estimation of the conditional distribution in regression with censored data. Ph.D. Thesis.

Wang, H. and Wang, L. (2009). Locally weighted censored quantile regression. *J. Amer. Statist. Assoc.* **104**, 1117-1128.

Wang, H. J., Zhou, J. and Li, Y. (2011). Variable selection for censored quantile regression. *Statist. Sinica* **23**, 145-167.

Yang, S. (1999). Censored median regression using weighted empirical survival and hazard function. *J. Amer. Statist. Assoc.* **94**, 137-145.

Yin, G., Zeng, D. and Li, H. (2008). Power-transformed linear quantile regression with censored data. *J. Amer. Statist. Assoc.* **103**, 1214-1224.

Ying, Z., Jung, S. and Wei, L. J. (1995). Survival analysis with median regression models. *J. Amer. Statist. Assoc.* **90**, 178-184.

Department of Statistics, University of Warwick, Coventry, CV4 7AL, United Kingdom.

E-mail: C.Leng@warwick.ac.uk

School of Mathematical Sciences, Beijing Normal University, Beijing, 100875, China.

E-mail: xweitong@bnu.edu.cn

(Received March 2012; accepted December 2012)