Transformed Dynamic Quantile Regression on Censored Data

Chi Wing Chu*, Tony Sit† and Gongjun Xu‡

*Department of Statistics, Columbia University

†Department of Statistics, The Chinese University of Hong Kong

‡Department of Statistics, University of Michigan

cc4160@columbia.edu tonysit@sta.cuhk.edu.hk gongjun@umich.edu

Abstract

We propose a class of power-transformed linear quantile regression models for time-to-event observations subject to censoring. By introducing a process of power transformation with different transformation parameters at individual quantile levels, our framework relaxes the assumption of logarithmic transformation on survival times and provides dynamic estimation of various quantile levels. With such formulation, our proposal no longer requires the potentially restrictive global linearity assumption imposed on a class of existing inference procedures for censored quantile regression. Uniform consistency and weak convergence of the proposed estimator as a process of quantile levels are established via the martingale-based argument. Numerical studies are presented to illustrate the outperformance of the proposed estimator over existing contenders under various settings.

Keywords: Censored quantile regression, Power transformation, Empirical process, Survival analysis

Chi Wing Chu is PhD student, Department of Statistics, Columbia University, New York, NY 10027 (E-mail: *cc4160@columbia.edu*).

Tony Sit is Assistant Professor, Department of Statistics, The Chinese University of Hong Kong, Hong Kong SAR (E-mail: *tonysit@sta.cuhk.edu.hk*). Gongjun Xu is Assistant Professor, Department of Statistics, University of Michigan, Ann Arbor, MI 48109 (E-mail: *gongjun@umich.edu*).

1 Introduction

Quantile regression (QR) has become a popular technique to model the entire conditional distribution of the response variable since Koenker and Bassett (1978). Whilst the ordinary least squares regression models the conditional mean response given the regressors, the quantile regression relates the conditional quantile function of the response of interest with the explanatory variables, which in turns offers a more comprehensive description that relates the quantity of interest with the collected covariates. Given its wide applicability and modeling flexibility, the model has attracted enormous interest in various fields in the literature. Examples include economics (Koenker and Hallock, 2001; Engle and Maganelli, 2004), growth chart (Wei et al., 2006; Wei and He, 2006) and modeling of time-to-event data (Portnoy, 2003; Peng and Huang, 2008; Wang and Wang, 2009; Wu and Yin, 2013, amongst others); see also Koenker (2005) and, more recently, Koenker et al. (2017) for extensive literature reviews on the corresponding development.

When the data are completely observed, we denote $\{(T_i, \mathbf{Z}_i), i = 1, \ldots, n\}$ a random sample from the target population, where T_i is a scalar response while $\mathbf{Z}_i$ is a $p \times 1$ vector of explanatory variables. The linear quantile regression model stipulates that, for a fixed $\tau \in (0, 1)$,

$$Q_{T_i}(\tau \mid \mathbf{Z}_i) = \mathbf{Z}_i^\top \boldsymbol{\beta}_0(\tau), \tag{1.1}$$

where $Q_T(\tau \mid \mathbf{Z})$ denotes the τ-th conditional quantile of T given $\mathbf{Z}$, *i.e.* $\Pr\{T \le Q_T(\tau \mid \mathbf{Z}) \mid \mathbf{Z}\} = \tau$. Solution to (1.1) can be obtained via the least absolute deviation approach which solves $\hat{\boldsymbol{\beta}}(\tau) = \arg\min_{\boldsymbol{\beta} \in \mathbb{R}^p} \sum_{i=1}^n \rho(T_i - \mathbf{Z}_i^\top \boldsymbol{\beta}; \tau)$, where $\rho(u; \tau) = u\{\tau - I(u < 0)\}$ corresponds to the check function.

Quantile regression has also been a useful alternative to the hazard-based semiparametric models. The most prominent feature of quantile regression for censored data is its ability to accommodate heterogeneous effects of the covariates, which can influence not only the location but also the shape of the survival time distribution. It is known that the heterogeneity in covariate effects cannot be easily incorporated in either the celebrated Cox proportional hazards model (Cox, 1972, 1975) or the accelerated failure time (AFT) model (Tsiatis, 1990; Ying, 1993; Jin et al., 2003). Furthermore, the conditional quantile of the survival time is easier to interpret than the hazard function and is often of direct interest.

To tackle the extra complication due to censoring, Powell (1984, 1986) modified the least absolute deviation procedure to handle censored observations; Ying et al. (1995) considered a semiparametric procedure for median regression under the independent censoring assumption. In view of the rather restrictive independent censoring assumption in real applications, Portnoy (2003) further relaxed this requirement and proposed a recursively reweighted inference procedure based on the principle of the Kaplan-Meier (KM) estimate's self-consistency, which was later also investigated in Wang and Wang (2009) in which a locally weighted censored quantile regression approach based on the local Kaplan-Meier estimator was proposed. Upon the conditional independence assumption, Peng and Huang (2008) exploited the martingale representation of the Nelson-Aalen estimator for the cumulative hazard function and proposed a recursive series of estimating equations for a sequence of quantiles under the global linear assumption, *i.e.* (1.1) holds for all $\tau \in (0, \tau_u] \subset (0, 1)$, where $\tau_u \in (0, 1)$ denotes a fixed constant that can ensure identifiability of the model under conditional independent censoring. Huang (2010) provided a numerically stable and computationally efficient algorithm for the aforementioned framework. More recently, De Backer et al. (2017) considered an extended version of the check function which can handle the effect of censoring at various level with an appropriate correction term. Different censored quantile regression models have also been developed to tackle variations in the model and sampling schemes, such as the cure rate quantile regression model (Wu and Yin, 2013), variable selection issues for censored data

(Wang et al., 2013; Peng et al., 2014; Zheng et al., 2017), and proper treatments for biased censored observations (Wang and Wang, 2014).

Quantile regression models for survival data often assume (1.1) by applying logarithmic transformation on survival time over all quantile values. Such mathematically convenient log-transformation, however, may not be adequate in practice and consequently may lead to estimation bias. In view of this limitation, quantile regression via power transformation has been recently studied. In particular, Mu and He (2007) was the first to introduce a Box-Cox transformation (Box and Cox, 1964) at a specific quantile level for complete data. In particular, for the failure time T, its Box-Cox transformation is defined as

$$h_\gamma(T) = \begin{cases} \frac{T^\gamma - 1}{\gamma} & \text{if } \gamma \neq 0 \\ \log(T) & \text{if } \gamma = 0 \end{cases}. \tag{1.2}$$

The extension of Mu and He (2007) to censored data are non-trivial. Yin et al. (2008) first studied power-transformed linear quantile regression with independent censored data. However, the independence censoring assumption required can be restrictive for many real data applications. Leng and Tong (2014) relaxed this assumption to the conditionally independent censoring assumption given covariates. Nevertheless, the estimation procedure involves estimating the unknown conditional distribution function through a non-parametric kernel-based local Kaplan-Meier estimator, which can potentially be unstable when either the covariate dimension or the censoring rate is high.

The main contribution of the paper is two-fold: Firstly, our proposal provides flexible and simultaneous estimation for quantiles under transformations by introducing a transformation parameter process for individual quantile levels. Despite the fact that Box-Cox transformation in quantile regression has demonstrated its effectiveness in Yin et al. (2008) and Leng and Tong (2014), their proposals may suffer the aforementioned limitations. In contrast, our proposal can offer a solution that can handle conditional independent censoring without using kernel estimation. Secondly, our method relaxes the global linear assumption as required in Peng and Huang (2008); this also adds extra flexibility to the censored quantile regression models when the transformation function is specified only at certain locations, if not the entire piece of information is missing. In addition to the large-sample properties of the proposed estimator developed, our numerical results also make evidence that the new estimator outperforms the existing contender under various settings.

The remainder of the paper is organized as follows: Section 2 establishes the corresponding set of estimating equations for our model parameters. The large-sample properties of the proposed estimators are discussed in section 3. Simulation results are presented in Section 4 followed by a data analysis of the HMO dataset on HIV positive subjects, which is illustrated in Section 5. Concluding remarks are given in Section 6. All the technical proofs are deferred to the Appendix.

2 Methodology

Consider the following censored transformation linear quantile regression model

$$Q_{h_{\gamma_0(\tau)}(T)}(\tau|\mathbf{Z}) = \tilde{\mathbf{Z}}^\top \tilde{\boldsymbol{\beta}}_0(\tau) + Q_\epsilon(\tau|\tilde{\mathbf{Z}}) := \mathbf{Z}^\top \boldsymbol{\beta}_0(\tau), \quad \tau \in (0,1), \tag{2.1}$$

where $\tilde{\mathbf{Z}}$ is the covariate vector (without the intercept), $\tilde{\boldsymbol{\beta}}_0(\tau)$ is the unknown regression coefficient vector and ϵ is the error term. The distribution of ϵ is unspecified and may depend on $\tilde{\mathbf{Z}}$. We denote $\mathbf{Z} = (1, \tilde{\mathbf{Z}}^\top)^\top$ and $\boldsymbol{\beta}_0(\tau) = \{Q_\epsilon(\tau|\mathbf{Z}), \tilde{\boldsymbol{\beta}}_0(\tau)^\top\}^\top$ by including the quantile of the error into the regression intercept. Meanwhile, $h_{\gamma_0(\tau)}(T)$ is the failure time under a monotonic transformation with an unknown transformation parameter $\gamma_0(\tau)$ for a given $\tau \in (0,1)$. For instance, we may consider the Box-Cox transform in (1.2) as an example. Since $h_{\gamma(\tau)}(\cdot)$ is monotone for any given $\gamma(\tau)$, using the equivariance property of quantile regression, we have

$$Q_{h_{\gamma(\tau)}(T)}(\tau|\mathbf{Z}) = h_{\gamma(\tau)}\big(Q_T(\tau|\mathbf{Z})\big), \quad \tau \in (0,1). \tag{2.2}$$

It is easy to see that (2.1) is equivalent to the following model on the quantile function of T,

$$Q_T(\tau|\mathbf{Z}) = h^{-1}_{\gamma_0(\tau)}\left(\mathbf{Z}^\top \boldsymbol{\beta}_0(\tau)\right), \quad \tau \in (0,1). \tag{2.3}$$

Let C be the censoring random variable which is conditionally independent of T given $\mathbf{Z}$. For subjects $i = 1, \cdots, n$, we have observed data $\{(\tilde{T}_i, \delta_i, \mathbf{Z}_i), i = 1, \ldots, n\}$, where $\tilde{T}_i = T_i \wedge C_i$ and $\delta_i = I(T_i \leq C_i)$. Our goal is to estimate $\{\gamma_0(\tau), \boldsymbol{\beta}_0(\tau)\}$ for $\tau \in (0,1)$. Instead of estimating $\{\gamma_0(\tau), \boldsymbol{\beta}_0(\tau)\}$ for all $\tau \in (0,1)$, we confine our attention to $\tau \in (0, \tau_U]$, where $\tau_U \in (0,1)$ in order to avoid the identifiability issues due to conditional independent censoring.

We adopt a martingale based estimation framework considered in Peng and Huang (2008). Our proposed method does not require the global linear assumption because of the additional transformation process considered. We define $F_T(t \mid \mathbf{Z}) = P(T \leq t \mid \mathbf{Z})$, $\Lambda_T(t \mid \mathbf{Z}) = -\log\{1 - F_T(t \mid \mathbf{Z})\}$, $N(t) = I(\tilde{T} \leq t, \delta = 1)$, $Y(t) = I(\tilde{T} \geq t)$ and $H(x) = -\log(1-x)$. Note that $\Lambda_T(t|\mathbf{Z})$ is the cumulative hazard function of T conditional on $\mathbf{Z}$, so based on the conditional independent censoring assumption (see, for example, Page 20 of Fleming and Harrington, 2005), we obtain $E\{N(t) - \Lambda_T(t \wedge \tilde{T}|\mathbf{Z})|\mathbf{Z}\} = 0$. Substitution of t with $Q_T(\tau|\mathbf{Z})$ gives

$$E[N\{Q_T(\tau|\mathbf{Z})\} - \Lambda_T\{Q_T(\tau|\mathbf{Z}) \wedge \tilde{T}|\mathbf{Z}\}|\mathbf{Z}] = 0. \tag{2.4}$$

Due to the fact that $F_T\{Q_T(\tau|\mathbf{Z})|\mathbf{Z}\} = \tau$, we have

$$\begin{aligned} \Lambda_T\{Q_T(\tau|\mathbf{Z}) \wedge \tilde{T}|\mathbf{Z}\} &= H(F_T\{Q_T(\tau|\mathbf{Z})|\mathbf{Z}\}) \wedge H(F_T\{\tilde{T}|\mathbf{Z}\}) \\ &= \int_0^\tau I[u \leq F_T\{\tilde{T}|\mathbf{Z}\}] dH(u) = \int_0^\tau Y\{Q_T(u|\mathbf{Z})\} dH(u). \end{aligned} \tag{2.5}$$

The estimating equation due to (2.3), (2.4) and (2.5),

$$\mathbf{S}(\gamma_0, \boldsymbol{\beta}_0, \tau) = E\left(\mathbf{Z}\left[N\left\{h^{-1}_{\gamma_0(\tau)}\left(\mathbf{Z}^\top\boldsymbol{\beta}_0(\tau)\right)\right\} - \int_0^\tau Y\left\{h^{-1}_{\gamma_0(u)}\left(\mathbf{Z}^\top\boldsymbol{\beta}_0(u)\right)\right\}dH(u)\right]\right) = 0 \tag{2.6}$$

can thus be established. Our estimator $\{\{\hat\gamma(\tau), \hat{\boldsymbol{\beta}}(\tau)\}, \tau \in (0, \tau_U]\}$ is a right-continuous step function with jumps on a grid $\mathcal{S}_{L(n)} = \{0 = \tau_0 < \tau_1 < \cdots < \tau_{L(n)} = \tau_U < 1\}$ and denote $\|\mathcal{S}_{L(n)}\| = \sup_{1\le j\le L(n)} |\tau_j - \tau_{j-1}|$. First note that the definition of conditional quantile implies $0 = Q_{T_i}(0|\mathbf{Z}_i) = h^{-1}_{\gamma_0(0)}(\mathbf{Z}_i^\top\boldsymbol{\beta}_0(0))$, so we can set $h^{-1}_{\hat\gamma(0)}(\mathbf{Z}_i^\top\hat{\boldsymbol{\beta}}(0)) = 0$. Subsequently for each τ_j, and for any $\gamma(\tau_j)$, we obtain $\hat{\boldsymbol{\beta}}(\gamma, \tau_j)$ by solving the following monotone estimating equation, which is the empirical counterpart of (2.6), for $\boldsymbol{\beta}(\tau_j)$ as shown below:

$$\begin{aligned}\hat{\mathbf{S}}(\gamma, \boldsymbol{\beta}, \tau) = n^{-1}\sum_{i=1}^n \mathbf{Z}_i \Bigg[N_i\left\{h^{-1}_{\gamma(\tau_j)}\left(\mathbf{Z}_i^\top\boldsymbol{\beta}(\tau_j)\right)\right\} \\ - \sum_{k=1}^j Y_i\left\{h^{-1}_{\hat\gamma(\tau_{k-1})}\left(\mathbf{Z}_i^\top\hat{\boldsymbol{\beta}}(\tau_{k-1})\right)\right\}\{H(\tau_k) - H(\tau_{k-1})\}\Bigg] = 0.\end{aligned} \tag{2.7}$$

An exact root for (2.7) may not exist since it is not continuous. Therefore, we define $\hat{\boldsymbol{\beta}}(\gamma, \tau_j)$ to be the generalized solutions in the sense of Fygenson and Ritov (1994). Recall $\boldsymbol{\beta}$ is a generalized solution of $\hat{\mathbf{S}}(\gamma, \boldsymbol{\beta}, \tau) = 0$ if slight perturbations of any of its components change the sign of $\hat{\mathbf{S}}$. Observe that (2.7) is monotone, we can transform the above problem to finding a minimizer for a convex objective function of $\boldsymbol{\beta}(\tau_j)$, which implies

$$\begin{aligned}\hat{\boldsymbol{\beta}}(\gamma, \tau_j) &= \arg\min_{\boldsymbol{\beta}(\tau_j)} \hat{L}(\gamma, \boldsymbol{\beta}, \tau) \\ &= \arg\min_{\boldsymbol{\beta}(\tau_j)} n^{-1}\sum_{i=1}^n \left[h_{\gamma(\tau_j)}(\tilde{T}_i) - \mathbf{Z}_i^\top\boldsymbol{\beta}(\tau_j)\right]\Bigg[-N_i\left\{h^{-1}_{\gamma(\tau_j)}\left(\mathbf{Z}_i^\top\boldsymbol{\beta}(\tau_j)\right)\right\} \\ &\qquad + \sum_{k=1}^j Y_i\left\{h^{-1}_{\hat\gamma(\tau_{k-1})}\left(\mathbf{Z}_i^\top\hat{\boldsymbol{\beta}}(\tau_{k-1})\right)\right\}\{H(\tau_k) - H(\tau_{k-1})\}\Bigg].\end{aligned} \tag{2.8}$$

Remark 1. *Equation* (2.8) *is an L_1-type problem of the conditional quantile of T given* $\mathbf{Z}$*. To illustrate the connection of the equation to the check function in quantile regression,* *recall that the check function is given by $\rho(u;\tau) = u\{\tau - I(u < 0)\}$. If we view the residual $h_{\gamma(\tau_j)}(\tilde{T}_i) - \mathbf{Z}_i^\top \hat{\boldsymbol{\beta}}(\gamma, \tau_j)$ as u and $\sum_{k=1}^{j} Y_i \left\{ h^{-1}_{\hat{\gamma}(\tau_{k-1})} \left(\mathbf{Z}_i^\top \hat{\boldsymbol{\beta}}(\tau_{k-1}) \right) \right\} \{H(\tau_k) - H(\tau_{k-1})\}$ as a function of τ_j, say $g_i(\tau_j) = g(\tau_j, \tilde{T}_i, \delta_i, \mathbf{Z}_i)$, then we can write the objective function as $\sum_{i=1}^{n} \rho(h_{\gamma(\tau_j)}(\tilde{T}_i) - \mathbf{Z}_i^\top \hat{\boldsymbol{\beta}}(\gamma, \tau_j); g_i(\tau_j))$, which is a comparable form to the solution of the ordinary linear quantile regression model given by (1.1). Note that we have subject specific $g_i(\tau_j)$ instead of a fixed constant τ_j in the check function, because the subject specific cumulative hazard function estimate has to be adjusted for complications due to censoring. As a result, the estimation approach based on (2.8) can be considered as an optimization problem with a generalized "check" function which depends on the values of the observations. Also, note that the mean of*

$$\sum_{k=1}^{j} Y_i \left\{ h^{-1}_{\gamma_0(\tau_{k-1})} \left(\mathbf{Z}_i^\top \boldsymbol{\beta}_0(\tau_{k-1}) \right) \right\} \{H(\tau_k) - H(\tau_{k-1})\}, \quad i = 1, \cdots, n$$

approximates $E \int_0^{\tau_j} Y\{Q_T(u)\} dH(u) = \int_0^{\tau_j} P\{T \geq Q_T(u)\} dH(u) = \int_0^{\tau_j} (1-u) d\{-\log(1-u)\} = \tau_j$ in the absence of censoring. By altering the proposed "check" function, this estimation approach is a generalization of the check function from complete data to censored data by treating the censoring with a martingale-based approach.

To estimate the scalar $\gamma(\tau_j)$, the standard grid search algorithm can be adopted. Consequently, for a reasonable set of $\gamma(\tau_j)$, we obtain $\hat{\boldsymbol{\beta}}(\gamma, \tau_j)$ by (2.8). Motivated by the two-stage estimator adopted by Chamberlain (1994) and Buckinsky (1995), we obtain $\hat{\gamma}(\tau_j)$ via

$$\begin{aligned} \hat{\gamma}(\tau_j) &= \arg\min_{\gamma(\tau_j)} \hat{R}(\gamma, \hat{\boldsymbol{\beta}}(\gamma), \tau) \\ &= \arg\min_{\gamma(\tau_j)} n^{-1} \sum_{i=1}^{n} \left[\tilde{T}_i - h^{-1}_{\gamma(\tau_j)} \left(\mathbf{Z}_i^\top \hat{\boldsymbol{\beta}}(\gamma, \tau_j) \right) \right] \left[-N_i \left\{ h^{-1}_{\gamma(\tau_j)} \left(\mathbf{Z}_i^\top \hat{\boldsymbol{\beta}}(\gamma, \tau_j) \right) \right\} \right. \\ &\qquad \left. + \sum_{k=1}^{j} Y_i \left\{ h^{-1}_{\hat{\gamma}(\tau_{k-1})} \left(\mathbf{Z}_i^\top \hat{\boldsymbol{\beta}}(\tau_{k-1}) \right) \right\} \{H(\tau_k) - H(\tau_{k-1})\} \right]. \end{aligned} \tag{2.9}$$

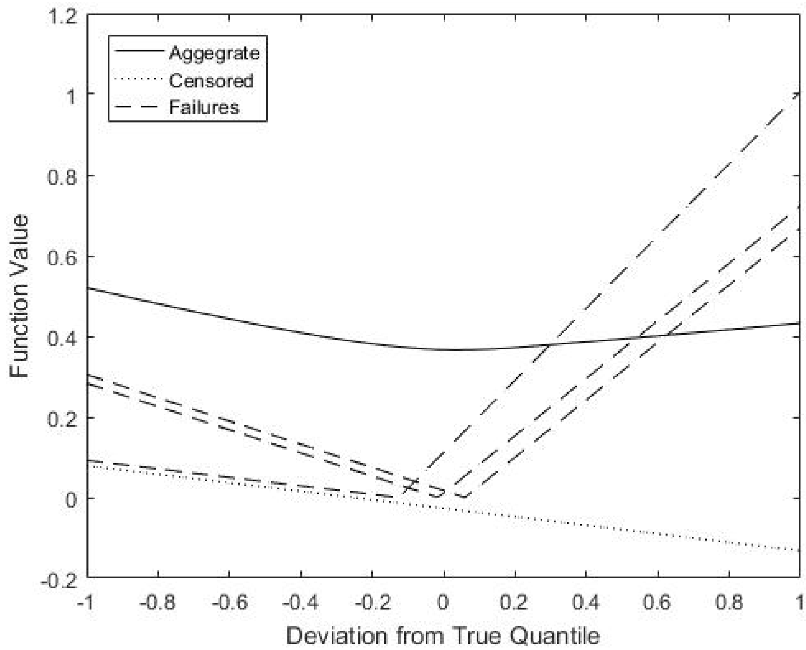
After obtaining $\hat{\gamma}(\tau_j)$, we take $\hat{\boldsymbol{\beta}}(\tau_j)$ to be $\hat{\boldsymbol{\beta}}(\hat{\gamma}, \tau_j)$. Note that the minimization of (2.9) is justified since we observe, for any $\mathbf{Z}$-measurable function $\mathbf{W}$, we have from (2.3), (2.4) and (2.5) that

$$E\left(\mathbf{W}\left[N\left\{h_{\gamma(\tau)}^{-1}\left(\mathbf{Z}^{\top}\boldsymbol{\beta}(\tau)\right)\right\}-\int_{0}^{\tau}Y\left\{h_{\gamma(u)}^{-1}\left(\mathbf{Z}^{\top}\boldsymbol{\beta}(u)\right)\right\}dH(u)\right]\right)=0. \tag{2.10}$$

If $\mathbf{W}$ is taken to be the gradient of $h_{\gamma(\tau_j)}^{-1}\left(\mathbf{Z}^{\top}\boldsymbol{\beta}(\tau_j)\right)$ with respect to $\gamma(\tau_j)$, which can be shown to be $\mathbf{Z}$-measurable, then the gradient of (2.9) with respect to $\gamma(\tau_j)$ corresponds to the empirical counterpart of (2.10). Though any choice of the weight $\mathbf{W}$ that is $\mathbf{Z}$-measurable can be used to set up a legitimate estimating equation, our particular choice of $\mathbf{W}$ here is not arbitrary. The current weight corresponds to a minimization problem (2.9), which allows easier numerical implementation; meanwhile, it captures the rate of change of the conditional quantile with respect to a change of the transformation parameter $\gamma(\tau_j)$. This particular choice of weight may account for the smaller standard errors of the estimates obtained by the proposed method compared with existing approaches; see Section 4 for details.

Similarly to (2.8), (2.9) is also an L_1-type problem using a "pseudo-check" function. Another merit of this pseudo-check function is its ability to handle heterogeneity. Instead of a universal τ_j across all observations, $g_i(\tau_j)=\sum_{k=1}^{j}Y_i\left\{h_{\hat{\gamma}(\tau_{k-1})}^{-1}\left(\mathbf{Z}_i^{\top}\hat{\boldsymbol{\beta}}(\tau_{k-1})\right)\right\}\{H(\tau_k)-H(\tau_{k-1})\}$, the term which serves as a role of τ_j in the ordinary check function, is tailored for each subject and hence could be heterogeneous. For an illustration of the linkage between the proposed objective function and the ordinary check function, Figure 1 presents the graph of (2.9) as a function of $h_{\gamma(\tau_j)}^{-1}\left(\mathbf{Z}_i^{\top}\boldsymbol{\beta}(\tau_j)\right)$, the quantity through which (2.9) depends on the parameter estimates at τ_j. In particular, we consider a simulated sample from the transformed quantile regression model described in Example I of our simulation studies. The objective function value is plotted against the deviation of the quantity $h_{\gamma(\tau_j)}^{-1}\left(\mathbf{Z}_i^{\top}\boldsymbol{\beta}(\tau_j)\right)$ from the true quantile $h_{\gamma_0(\tau_j)}^{-1}\left(\mathbf{Z}_i^{\top}\boldsymbol{\beta}_0(\tau_j)\right)$. We show the graphs for the aggregate objective function of 500 data and for some selected data points. The dashed lines suggest that the function behaves as if a

Figure 1: Pseudo check function against quantile value



check function in case of failures, exhibiting a turnaround near the true quantile value. For a censored case, the function crosses zero without a turning point. The aggregate objective function is now the mean of 500 individual check functions and still convex (in quantile value) in nature.

One practical concern for the estimation is the choice of the grid size in the sequential procedure. From the proof of weak convergence, we shall choose a grid size of order $o(n^{-1/2})$ to ensure weak convergence. In our numerical study, we adopt an equally spaced grid with 0.05 grid size, where the performance of the estimator is satisfactory and the computation efforts involved are reasonable.

Remark 2. *Apart from* (2.9)*, it is also natural to extend the concept used by Mu and He (2007) to estimate* $\gamma(\tau_j)$*, where* $\gamma(\tau_j)$ *is the minimizer of a cumsum process of residuals. To* *this extent, based on (2.7), we could define*

$$V_n\{\mathbf{z},\gamma(\tau_j)\} = n^{-1}\sum_{i=1}^{n} I(\mathbf{Z}_i \le \mathbf{z})\left[N_i\left\{h^{-1}_{\gamma(\tau_j)}\left(\mathbf{Z}_i^\top\hat{\boldsymbol{\beta}}(\tau_j)\right)\right\} - \sum_{k=1}^{j} Y_i\left\{h^{-1}_{\hat{\gamma}(\tau_{k-1})}\left(\mathbf{Z}_i^\top\hat{\boldsymbol{\beta}}(\tau_{k-1})\right)\right\}\{H(\tau_k)-H(\tau_{k-1})\}\right], \quad (2.11)$$

where $\mathbf{Z}_i \le \mathbf{z}$ denotes componentwise inequality. Then $\tilde{\gamma}(\tau_j)$ is obtained as

$$\tilde{\gamma}(\tau_j) = \underset{\gamma(\tau_j)}{\arg\min}\, V_n^*\{\gamma(\tau_j)\} = \underset{\gamma(\tau_j)}{\arg\min}\int_{-\infty}^{\infty}[V_n\{\mathbf{z},\gamma(\tau_j)\}]^2\, d\hat{F}_{\mathbf{Z}}(\mathbf{z}) = \underset{\gamma(\tau_j)}{\arg\min}\, n^{-1}\sum_{i=1}^{n}[V_n\{\mathbf{z}_i,\gamma(\tau_j)\}]^2, \quad (2.12)$$

where $\hat{F}_{\mathbf{Z}}(\mathbf{z}) = n^{-1}\sum_{i=1}^{n} I(\mathbf{Z}_i \le \mathbf{z})$ defines the empirical distribution function of $\{\mathbf{Z}_1,\ldots,\mathbf{Z}_n\}$. From our numerical studies, the two stage estimator appears to perform better in the sense that estimates with smaller standard errors are obtained. One possible reason for the reduced variance obtained in our simulation is suggested by the choice of the weight, which is the gradient of the conditional quantile with respect to the transformation parameter. Moreover, as we require a sequential estimation procedure, the performance of the estimator in higher quantiles depend on the lower ones. In view of this, it may be more beneficial to consider a procedure that produces estimates with smaller standard errors in which case the whole process of estimator would be more stable.

3 Asymptotic Results

In this section, we establish the uniform consistency and weak convergence of the proposed estimator $\hat{\gamma}(\tau)$ and $\hat{\boldsymbol{\beta}}(\tau)$.

Theorem 1. *Under Conditions (C1)–(C7) given in the appendix, if $\lim_{n\to\infty}\|\mathcal{S}_{L(n)}\| = 0$, then $\sup_{\tau\in[v,\tau_U]}|\hat{\gamma}(\tau)-\gamma_0(\tau)| + \left\|\hat{\boldsymbol{\beta}}(\hat{\gamma},\tau)-\boldsymbol{\beta}_0(\tau)\right\| \to_p 0$, where $0 < v < \tau_U$.*

Theorem 2. *Under Conditions (C1)–(C7), if $\lim_{n\to\infty} n^{1/2}\|\mathcal{S}_{L(n)}\| = 0$, then $n^{1/2}\binom{\hat{\gamma}(\tau)-\gamma_0(\tau)}{\hat{\boldsymbol{\beta}}(\hat{\gamma},\tau)-\boldsymbol{\beta}_0(\tau)}$ converges weakly to a Gaussian process for $\tau \in [v,\tau_U]$, where $0 < v < \tau_U$.*

The regularity conditions and proofs are provided in the appendix. A key assumption of the uniform consistency and weak convergence is the Lipschitz continuity of $\gamma_0(\tau)$ in τ. When $\gamma_0(\tau)$ is not Lipschitz and there are finite discontinuity jump points of $\gamma_0(\tau)$ as in the simulation examples, we can obtain from the proof that the consistency result of $|\hat{\gamma}(\tau) - \gamma_0(\tau)| + \left\|\hat{\boldsymbol{\beta}}(\hat{\gamma}, \tau) - \boldsymbol{\beta}_0(\tau)\right\| \to_p 0$ and asymptotic normality of $n^{1/2}[\hat{\gamma}(\tau) - \gamma_0(\tau), \{\hat{\boldsymbol{\beta}}(\hat{\gamma}, \tau) - \boldsymbol{\beta}_0(\tau)\}^\top]^\top$ still hold almost everywhere for τ except the jump points.

Note that the uniform consistency and weak convergence results only hold for τ outside the neighborhood of 0. In particular, recall that we start the iterations in our estimation procedure by taking $h^{-1}_{\hat{\gamma}(0)}(\mathbf{Z}_i^\top \hat{\boldsymbol{\beta}}(0)) = h^{-1}_{\gamma_0(0)}(\mathbf{Z}_i^\top \boldsymbol{\beta}_0(0)) = 0$. This suggests that $\hat{\gamma}(0)$ and $\hat{\boldsymbol{\beta}}(0)$ are indeed not unique, therefore the uniform consistency and weak convergence results cannot hold for τ in the neighbourhood of 0. However, as discussed in Peng and Huang (2008), the numerical instability of the estimation at small τ has a small impact on the estimation at larger $\tau < \tau_U$.

Noteworthy, the covariance matrix of the limiting process of $n^{1/2}[\hat{\gamma}(\tau) - \gamma_0(\tau), \{\hat{\boldsymbol{\beta}}(\hat{\gamma}, \tau) - \boldsymbol{\beta}_0(\tau)\}^\top]^\top$, as shown in the proof of Theorem 2, involves the conditional density functions of the survival time, which are unknown and our quantities of interest. In order to make inference on the parameters, we propose to use a simple resampling approach by perturbing the minimand as adopted in Jin et al. (2001) and Peng and Huang (2008).

Let $\zeta_1, \ldots, \zeta_n$ be i.i.d. random variables from a nonnegative known distribution with mean 1 and variance 1, such as exponential distribution with rate 1. For a data set with sample size n, we fix the data values and generate $\{\zeta_1, \ldots, \zeta_n\}$, then we obtain $\gamma^*(\tau_j)$ and $\boldsymbol{\beta}^*(\tau_j)$ sequentially for each $j \in \{1, \ldots, L\}$ by solving the following two perturbed objective functions

$$\boldsymbol{\beta}^*(\gamma, \tau_j) = \underset{\boldsymbol{\beta}(\tau_j)}{\arg\min}\, n^{-1} \sum_{i=1}^{n} \zeta_i \left[h_{\gamma(\tau_j)}(\tilde{T}_i) - \mathbf{Z}_i^\top \boldsymbol{\beta}(\tau_j)\right] \left[-\delta_i I\left\{h_{\gamma(\tau_j)}(\tilde{T}_i) \le \mathbf{Z}_i^\top \boldsymbol{\beta}(\tau_j)\right\}\right.$$
$$\left. + \sum_{k=1}^{j} I\left\{\tilde{T}_i \ge h^{-1}_{\gamma^*(\tau_{k-1})}\left(\mathbf{Z}_i^\top \boldsymbol{\beta}^*(\tau_{k-1})\right)\right\}\{H(\tau_k) - H(\tau_{k-1})\}\right];$$

$$\gamma^*(\tau_j) = \arg\min_{\gamma(\tau_j)} n^{-1} \sum_{i=1}^{n} \zeta_i \left[\tilde{T}_i - h^{-1}_{\gamma(\tau_j)}\left(\mathbf{Z}_i^\top \boldsymbol{\beta}^*(\gamma, \tau_j)\right)\right] \left[-\delta_i I\left\{h_{\gamma(\tau_j)}(\tilde{T}_i) \le \mathbf{Z}_i^\top \boldsymbol{\beta}^*(\gamma, \tau_j)\right\} + \sum_{k=1}^{j} I\left\{\tilde{T}_i \ge h^{-1}_{\gamma^*(\tau_{k-1})}\left(\mathbf{Z}_i^\top \boldsymbol{\beta}^*(\tau_{k-1})\right)\right\}\{H(\tau_k) - H(\tau_{k-1})\}\right].$$

Again $h^{-1}_{\gamma^*(0)}(\mathbf{Z}_i^\top \boldsymbol{\beta}^*(0))$ are set to be 0, and $\boldsymbol{\beta}^*(\tau_j)$ are taken to be $\boldsymbol{\beta}^*(\gamma^*, \tau_j)$. Then $\gamma^*(\tau)$ and $\boldsymbol{\beta}^*(\tau)$ are defined to be a piecewise-constant function that jumps only at τ_j, $j \in \{1, \ldots, L\}$. The above procedure is repeated for B times such that for each $r \in \{1, \ldots, B\}$, we generate a set of variates $\{\zeta_1, \ldots, \zeta_n\}$ and obtain B realizations of $\gamma^*_r(\tau)$ and $\boldsymbol{\beta}^*_r(\tau)$. Consequently, the confidence intervals for $\gamma(\tau)$ and $\boldsymbol{\beta}(\tau)$ can be constructed using the percentiles of $\gamma^*_r(\tau)$ and $\boldsymbol{\beta}^*_r(\tau)$ respectively or by normal approximation. In order to justify the above resampling method, we present the following theorem.

Theorem 3. *Under Conditions (C1)–(C7), $n^{1/2}[\gamma^*(\tau) - \hat{\gamma}(\tau), \{\boldsymbol{\beta}^*(\gamma^*, \tau) - \hat{\boldsymbol{\beta}}(\hat{\gamma}, \tau)\}^\top]^\top$ given the observed data converges weakly to the same limiting process of $n^{1/2}[\hat{\gamma}(\tau) - \gamma_0(\tau), \{\hat{\boldsymbol{\beta}}(\hat{\gamma}, \tau) - \boldsymbol{\beta}_0(\tau)\}^\top]^\top$, for $\tau \in [v, \tau_U]$, where $0 < v < \tau_U$.*

The proof of Theorem 3 is given in the appendix.

It is often of general interest to question whether or not the transformation process is indeed a zero process so Peng and Huang (2008)'s approach can be adopted directly. To this end, we consider a class of null hypothesis in the form of $H_{10} : \gamma(\tau) = r_0(\tau)$ for $\tau \in [l, u] \subset (0, \tau_U]$, where $r_0(\tau)$ is a known process of hypothesized value. Inspired by Peng and Huang (2008)'s inference procedure, we exploit the asymptotic normality of the proposed estimator and consider an integral test statistic, which is defined by $\mathcal{T}_1 = n^{1/2} \int_l^u |\hat{\gamma}(u) - r_0(u)| \Xi_0(u) du$, where $\Xi_0(u)$ is a non-negative weight function that can be chosen to capture the deviation from H_{10} and achieve a reasonable power. Note that $\mathcal{T}_1$ is asymptotically mean zero by Theorem 2 and the continuous mapping theorem. In line with the above resampling scheme, we define $\mathcal{T}_1^* = n^{1/2} \int_l^u |\gamma^*(u) - \hat{\gamma}(u)| \Xi_0(u) du$. Due to Theorem 3, the conditional distribution of $\mathcal{T}_1^*$ given the observed data is equivalent to the unconditional distribution of $\mathcal{T}_1$. Therefore, we may construct a size α test of H_{10} that rejects when $\mathcal{T}_1 > c_{1,1-\alpha}$ where $c_{1,1-\alpha}$ is the $(1-\alpha)$th empirical quantile of $\mathcal{T}_1^*$, which is obtained from the B realizations of resample.

Meanwhile, one may also be interested in assessing whether a varying quantile effect exists in term of a non-constant transformation process. Mathematically, it can be expressed as the test of $H_{20} : \gamma(\tau) = r$, where r is an unspecified constant. Note that a natural estimator of the average transformation parameter over a range of quantiles could be $\hat{\gamma}_{avg} = \int_l^u \gamma(\tau)d\tau$, where $[l, u] \subset (0, \tau_U]$. When H_{20} is true, $\hat{\gamma}_{avg}$ can be interpreted as an estimate of the unknown constant r. Accordingly, we consider an integral test statistic for H_{20} that exploits the departure of the transformation parameter estimate from the average, $\mathcal{T}_2 = |n^{1/2} \int_l^u \{\hat{\gamma}(u) - \hat{\gamma}_{avg}\}\Xi_1(u)du|$. The null distribution of $\mathcal{T}_2$ can be approximated by the conditional distribution of $\mathcal{T}_2^* = |n^{1/2} \int_l^u [\{\gamma^*(u) - \hat{\gamma}(u)\} - \{\gamma^*_{avg} - \hat{\gamma}_{avg}\}]\Xi_1(u)du|$. As a result, by obtaining B realizations of $\mathcal{T}_2^*$ via the above resampling method, a size α test of H_{20} can be constructed by rejecting H_{20} when $\mathcal{T}_2^* > c_{2,1-\alpha}$ where $c_{2,1-\alpha}$ is the $(1 - \alpha)$th empirical quantile of $\mathcal{T}_2^*$.

4 Simulations

This section examines the finite-sample performance of the proposed methods through Monte Carlo simulations. We evaluate the proposed method using six examples. First, we consider two Box-Cox transformation quantile regression models in Examples I and II and compare the proposed method with several existing transformation quantile regression estimation methods, including Leng and Tong (2014) and Yin et al. (2008). Second, we compare in Example III and IV the performance of the proposed method with a couple of quantile regression estimation methods that are fitted under the true transformation, including Peng and Huang (2008) and Wang and Wang (2009). Third, we demonstrate the robustness of the proposed method in Example V and VI under the scenarios of a misspecified transformed quantile model except the interested quantile level and of a moderately high number of covariates. For each method, we report the empirical bias, the sample standard deviations, and the empirical mean squared errors based on 500 simulated data sets of sample size

Table 1: Simulation under a transformed quantile regression model

		$\tau = .25$					$\tau = .5$					$\tau = .75$				
		$\hat\gamma$	$\hat\beta^{(0)}$	$\hat\beta^{(1)}$	$\hat\beta^{(2)}$	$\hat\beta^{(3)}$	$\hat\gamma$	$\hat\beta^{(0)}$	$\hat\beta^{(1)}$	$\hat\beta^{(2)}$	$\hat\beta^{(3)}$	$\hat\gamma$	$\hat\beta^{(0)}$	$\hat\beta^{(1)}$	$\hat\beta^{(2)}$	$\hat\beta^{(3)}$
								$n = 200$ & 40% censoring								
Proposed	Bias	-.090	.011	-.044	-.041	-.042	.068	.077	.095	.072	.076	-.009	.033	-.000	-.001	-.003
	SD	.368	.127	.253	.220	.215	.341	.258	.446	.370	.390	.165	.127	.245	.215	.213
	MSE	.144	.016	.066	.050	.048	.121	.073	.208	.142	.158	.027	.017	.060	.046	.045
	SE	.368	.172	.298	.256	.257	.325	.245	.393	.339	.341	.238	.173	.302	.279	.279
	CP	.924	.940	.930	.928	.930	.942	.962	.936	.926	.934	.954	.966	.940	.950	.960
CS	Bias	-.059	.022	-.021	-.017	-.016	.082	.093	.115	.094	.100	-.002	.043	.016	.015	.013
	SD	.350	.133	.259	.223	.224	.370	.284	.492	.421	.426	.201	.160	.287	.258	.259
	MSE	.126	.018	.068	.050	.050	.143	.089	.255	.186	.192	.040	.027	.083	.067	.067
PH	Bias	-	-.244	-.538	-.533	-.532	-	-.200	-.447	-.451	-.447	-	-.233	-.477	-.476	-.476
	SD	-	.052	.094	.072	.072	-	.069	.105	.058	.064	-	.043	.064	.043	.042
	MSE	-	.062	.299	.289	.289	-	.045	.211	.207	.204	-	.056	.231	.228	.228
LT	Bias	-.039	-.004	.022	.027	.026	.075	.174	.198	.190	.195	-.062	-.041	-.031	-.030	-.030
	SD	.351	.123	.278	.240	.236	.350	.298	.633	.564	.591	.198	.144	.271	.248	.248
	MSE	.124	.015	.078	.058	.056	.128	.119	.440	.354	.387	.043	.022	.074	.063	.062
YZL	Bias	-.178	.178	.869	.859	.867	.058	.083	.372	.360	.372	-.006	.005	.132	.131	.129
	SD	.323	.314	.628	.570	.566	.347	.352	.815	.752	.773	.180	.169	.290	.252	.258
	MSE	.136	.130	1.150	1.063	1.073	.124	.131	.802	.694	.736	.032	.029	.102	.080	.083
re-YZL	Bias	-.053	1.076	.853	.934	.878	.093	.602	.225	.339	.336	.041	.271	.076	.100	.108
	SD	.373	.730	1.261	.932	.913	.407	.945	1.202	1.061	1.287	.205	.375	.637	.479	.478
	MSE	.142	1.691	2.317	1.741	1.605	.174	1.256	1.495	1.241	1.769	.044	.214	.411	.239	.240
								$n = 200$ & 20% censoring								
Proposed	Bias	-.074	-.002	-.036	-.037	-.039	.050	.049	.062	.052	.052	-.000	.025	.013	.009	.009
	SD	.287	.101	.204	.180	.175	.260	.200	.312	.281	.284	.142	.114	.201	.186	.185
	MSE	.088	.010	.043	.034	.032	.070	.042	.101	.082	.083	.020	.014	.041	.035	.034
	SE	.327	.140	.248	.215	.217	.272	.195	.311	.272	.274	.182	.137	.237	.215	.215
	CP	.948	.952	.942	.944	.948	.964	.980	.966	.964	.958	.948	.944	.926	.948	.936
CS	Bias	-.051	.006	-.017	-.018	-.019	.061	.060	.073	.064	.066	-.002	.030	.018	.013	.017
	SD	.289	.105	.218	.194	.191	.285	.212	.339	.301	.308	.183	.144	.247	.232	.234
	MSE	.086	.011	.048	.038	.037	.085	.048	.120	.094	.099	.034	.021	.061	.054	.055
PH	Bias	-	-.261	-.519	-.515	-.516	-	-.216	-.432	-.435	-.434	-	-.248	-.461	-.462	-.461
	SD	-	.044	.083	.063	.064	-	.066	.102	.052	.055	-	.036	.057	.039	.038
	MSE	-	.070	.276	.269	.271	-	.051	.197	.192	.191	-	.063	.216	.215	.214
LT	Bias	-.052	.009	-.007	-.008	-.009	.040	.163	.101	.091	.095	-.089	-.036	-.072	-.082	-.077
	SD	.285	.105	.225	.196	.190	.258	.259	.490	.442	.463	.182	.128	.234	.214	.218
	MSE	.084	.011	.051	.038	.036	.068	.094	.250	.204	.223	.041	.018	.060	.053	.054
YZL	Bias	-.184	.540	.613	.596	.589	.047	.267	.209	.189	.185	-.012	.138	.063	.052	.056
	SD	.266	.315	.474	.445	.432	.274	.373	.641	.570	.511	.173	.176	.268	.232	.235
	MSE	.105	.391	.600	.553	.533	.077	.211	.455	.361	.296	.030	.050	.076	.057	.059
re-YZL	Bias	-.065	1.219	.665	.814	.786	.060	.566	.112	.179	.171	.008	.302	.012	.045	.024
	SD	.309	.777	1.100	.811	.852	.309	.621	1.286	.783	.866	.194	.335	.527	.403	.406
	MSE	.100	2.090	1.652	1.320	1.344	.099	.706	1.667	.646	.780	.038	.203	.278	.164	.165

*Approaches CS, PH, LT, YZL and re-YZL denote the proposed method with cumsum minimand for $\gamma(\cdot)$, Peng and Huang (2008)'s, Leng and Tong (2014)'s, Yin et al. (2008)'s and the revised Yin et al. (2008)'s (described in Remark 3) proposals, respectively. Refer to Remark 4 for details of the underlined figures.

Table 2: Simulation under a transformed quantile regression model with heteroscedastic errors

		$\tau = .25$					$\tau = .5$					$\tau = .75$				
		$\hat{\gamma}$	$\hat{\beta}^{(0)}$	$\hat{\beta}^{(1)}$	$\hat{\beta}^{(2)}$	$\hat{\beta}^{(3)}$	$\hat{\gamma}$	$\hat{\beta}^{(0)}$	$\hat{\beta}^{(1)}$	$\hat{\beta}^{(2)}$	$\hat{\beta}^{(3)}$	$\hat{\gamma}$	$\hat{\beta}^{(0)}$	$\hat{\beta}^{(1)}$	$\hat{\beta}^{(2)}$	$\hat{\beta}^{(3)}$
$n = 200$ & 40% censoring																
Proposed	Bias	-.099	.019	-.024	-.038	-.036	.054	.075	.133	.065	.069	-.005	.050	.034	.013	.010
	SD	.433	.168	.287	.264	.268	.407	.299	.570	.420	.432	.221	.178	.387	.289	.284
	MSE	.197	.029	.083	.071	.073	.169	.095	.342	.181	.191	.049	.034	.151	.084	.081
	SE	.429	.212	.343	.302	.307	.387	.293	.523	.413	.413	.339	.258	.562	.417	.414
	CP	.926	.942	.922	.926	.920	.910	.946	.938	.928	.918	.982	.982	.966	.962	.970
CS	Bias	-.074	.026	-.009	-.018	-.015	.062	.085	.156	.079	.084	.021	.067	.089	.048	.050
	SD	.421	.176	.304	.281	.293	.436	.320	.663	.465	.481	.240	.202	.435	.317	.315
	MSE	.182	.032	.092	.080	.086	.194	.110	.463	.223	.238	.058	.045	.197	.103	.102
PH	Bias	-	-.250	-.439	-.515	-.514	-	-.206	-.430	-.448	-.444	-	-.225	-.567	-.483	-.485
	SD	-	.059	.111	.083	.086	-	.078	.127	.072	.079	-	.054	.090	.060	.058
	MSE	-	.066	.205	.272	.272	-	.048	.201	.206	.203	-	.054	.329	.237	.238
LT	Bias	-.054	-.003	.007	.043	.046	.058	.170	.311	.224	.231	-.048	-.028	-.041	.026	.031
	SD	.427	.162	.301	.292	.297	.406	.371	.839	.689	.715	.250	.199	.417	.338	.348
	MSE	.185	.026	.090	.087	.090	.168	.166	.800	.525	.564	.065	.041	.175	.115	.122
YZL	Bias	-.185	.151	1.260	.994	1.008	.027	.036	.590	.404	.428	-.013	-.016	.234	.173	.176
	SD	.390	.373	1.010	.803	.843	.399	.371	1.174	.841	.878	.217	.211	.466	.335	.350
	MSE	.187	.162	2.607	1.633	1.727	.160	.139	1.726	.871	.954	.047	.045	.272	.142	.153
re-YZL	Bias	-.062	1.201	1.601	1.111	1.071	.062	.625	.616	.463	.342	.033	.336	.211	.155	.123
	SD	.448	1.099	2.130	1.637	1.479	.453	1.130	2.236	1.830	1.206	.261	.557	.948	.678	.615
	MSE	.204	2.650	7.100	3.914	3.335	.209	1.668	5.381	3.562	1.571	.069	.423	.943	.484	.393
$n = 200$ & 20% censoring																
Proposed	Bias	-.104	-.001	-.029	-.042	-.043	.048	.049	.092	.055	.055	-.004	.033	.037	.013	.014
	SD	.392	.146	.252	.233	.244	.315	.241	.386	.328	.342	.196	.156	.326	.258	.257
	MSE	.165	.021	.064	.056	.062	.102	.060	.157	.110	.120	.038	.025	.108	.067	.066
	SE	.389	.176	.287	.260	.264	.326	.235	.409	.336	.338	.262	.197	.419	.317	.317
	CP	.922	.934	.912	.916	.912	.946	.966	.960	.950	.952	.958	.970	.946	.942	.948
CS	Bias	-.079	.005	-.018	-.028	-.027	.054	.059	.104	.065	.068	-.009	.030	.036	.008	.011
	SD	.369	.147	.250	.233	.237	.342	.264	.436	.361	.383	.209	.164	.347	.263	.265
	MSE	.142	.022	.063	.055	.057	.120	.073	.201	.135	.152	.044	.028	.121	.069	.070
PH	Bias	-	-.265	-.426	-.500	-.499	-	-.219	-.418	-.431	-.431	-	-.242	-.548	-.469	-.469
	SD	-	.051	.098	.074	.077	-	.071	.117	.065	.066	-	.043	.078	.051	.051
	MSE	-	.073	.191	.255	.255	-	.053	.188	.190	.190	-	.060	.306	.223	.222
LT	Bias	-.084	.007	-.006	-.014	-.011	.043	.180	.245	.127	.131	-.074	-.017	-.057	-.048	-.043
	SD	.375	.149	.260	.239	.248	.320	.341	.631	.539	.566	.232	.169	.355	.282	.283
	MSE	.148	.022	.068	.057	.062	.104	.149	.459	.307	.338	.059	.029	.129	.082	.082
YZL	Bias	-.184	.548	1.183	.707	.706	.029	.255	.491	.225	.232	-.018	.129	.243	.077	.078
	SD	.364	.400	.951	.646	.666	.321	.490	.849	.678	.672	.212	.212	.453	.301	.301
	MSE	.166	.460	2.305	.917	.941	.104	.305	.961	.510	.506	.045	.061	.264	.096	.097
re-YZL	Bias	-.097	1.272	1.502	.831	.795	.049	.681	.442	.279	.178	.004	.362	.182	.054	.040
	SD	.401	.997	1.811	1.211	1.138	.358	1.020	1.518	1.329	1.113	.222	.491	.968	.559	.583
	MSE	.170	2.611	5.534	2.157	1.927	.130	1.505	2.498	1.843	1.270	.049	.372	.971	.315	.342

*Approaches CS, PH, LT, YZL and re-YZL denote the proposed method with cumsum minimand for $\gamma(\cdot)$, Peng and Huang (2008)'s, Leng and Tong (2014)'s, Yin et al. (2008)'s and the revised Yin et al. (2008)'s (described in Remark 3) proposals, respectively. Refer to Remark 4 for details of the underlined figures.

$n = 200$ or $n = 500$ each. Due to space concern, we present here only the numerical results for $n = 200$; corresponding results for $n = 500$ are included in Section D of the supplementary materials for comparison and reference.

Example I In the first example, event times are generated from the following Box-Cox transformation quantile regression model

$$Q_T(\tau|\mathbf{Z}) = h^{-1}_{\gamma_0(\tau)}\left(\mathbf{Z}^\top \boldsymbol{\beta}_0(\tau)\right) = h^{-1}_{\gamma_0(\tau)}\Bigg(b_0 + Z_1 b_1 + Z_2 b_2 + Z_3 b_3 + Q_\epsilon(\tau|\mathbf{Z})\Bigg).$$

where ϵ follows $N(0, 0.25^2)$. Under model (2.3), the corresponding regression quantile given $\mathbf{Z} = (1, Z_1, Z_2, Z_3)^\top$ is $\boldsymbol{\beta}_0(\tau) = (Q_\epsilon(\tau) + b_0, b_1, b_2, b_3)^\top$. The covariates Z_1, Z_2 and Z_3 are $Unif(0, 1)$, $N(0, 0.5^2)$ and $N(0, 0.5^2)$ respectively. We set $b_0 = b_1 = b_2 = b_3 = 1$, so that the first entry of $\boldsymbol{\beta}_0(\tau)$ will be τ-dependent, while the remainings are constant. The transformation coefficient $\gamma_0(\tau)$ is 1 for $\tau \leq 0.4$ and 0.5 for $\tau > 0.4$ in which case $h^{-1}_{\gamma_0(\tau)}(x) = (\gamma_0(\tau)x + 1)^{1/\gamma_0(\tau)}$. The censoring distribution is generated by $Unif(0, c)$ with $c = \exp(c_0 + Z_1 + Z_2 + Z_3)$, where c_0 is taken to be 1.6 and 2.3 to yield the target censoring rates of 40% and 20%, respectively. Under this setting, we consider 500 simulated data sets of sample size $n = 200$ or $n = 500$. The resampling size is taken to be 200, while the perturbation variable ζ is generated from $exponential(1)$. We adopt an equally spaced grid with 0.05 grid size for the implementation of the proposed methods presented in section 2. The optimal $\hat{\gamma}(\tau)$ is located through grid search in the interval $[-2, 2]$. In particular, we first locate a preliminary estimate by using only the uncensored observations in our procedure (or equivalently, by Chamberlain (1994)'s method). The final estimate of $\hat{\gamma}(\tau)$ is then obtained by searching locally around this initial value using the proposed method with all observations in sake of higher numerical stability. However, one has to be aware that the search range should be wide enough so that the optimal $\hat{\gamma}(\tau)$ does not fall on the boundaries. In our simulation studies, we use a range of ± 0.2 for the local search such that the run time is reasonable while maintaining the empirical proportions of $\hat{\gamma}(\tau)$ occurring on the boundaries negligible.

In Table 1, we compare the numerical results at $\tau = 0.25, 0.50$ and 0.75 for five different approaches, namely the proposed method, the alternative method using a cumsum minimand for $\hat{\gamma}(\tau)$ suggested by (2.11) and (2.12), Peng and Huang (2008)'s, Leng and Tong (2014)'s, Yin et al. (2008)'s and a modification of Yin et al. (2008)'s (see Remark 3) proposals. We report the empirical bias (Bias), the sample standard deviations (SD) and the mean squared errors (MSE) for comparison. For the proposed method, we also give the average of the estimated standard errors (SE) based on the resampling method, and the coverage probabilities (CP) of the 95% confidence intervals constructed by the empirical distribution of the estimates. We observe that the proposed method gives smaller biases and standard deviations than the cumsum alternative and Leng and Tong (2014) in almost all cases. Indeed, the proposed method dominates its counterparts if we consider the mean squared errors. On the other hand, since the global linear assumption required by Peng and Huang (2008) and the unconditional independence assumption required by Yin et al. (2008) are clearly violated, their methods do not provide reasonable estimates in the current setup. It is also shown that the resampling-based standard errors are close to the empirical standard deviations, and the coverage probabilities are close to 95%, which justify the use of the proposed resampling method.

Remark 3. *One may consider modifying Yin et al. (2008) so that the inverse probability weight scheme can be extended for the conditional independent censoring case. To achieve this goal, one has to provide a reliable estimate for the conditional survival function of the censoring time given covariates, i.e.* $\widehat{\widehat{G}}(\cdot \mid \mathbf{Z})$*, which is often approximated by an appropriate kernel estimator. To examine its effectiveness, we run a set of simulation using the local Kaplan-Meier estimator to estimate* $\widehat{\widehat{G}}(\cdot \mid \mathbf{Z})$*, with the bandwidth taken to be* $n^{-1/3+0.01}$*, as described in Leng and Tong (2014). The revised method, which is regarded as "re-YZL" approach, however, does not provide satisfactory numerical results compared with our proposal. This could be possibly due to the fact that the structure of Yin et al. (2008)'s estimating equa-*

tions can sometimes be numerically unstable when the inverse probability weights are close to zero. Nevertheless, given the neat structure of Yin et al. (2008)'s estimating equations, it is potentially an interesting problem for further research on how to extend the scope so that conditional independence censoring can also be covered under their framework.

Remark 4. *We found that the estimation approach suggested by Leng and Tong (2014) may not converge properly for high quantile levels. The reported simulation results of their proposal in the case of ($\tau = .75$) adopt a slight modification by considering the absolute value of their minimization function. We suspect this phenomenon can possibly be explained by the unsatisfactory performance of the locally weighted kernel estimator for the conditional cumulative hazard function for high quantile levels.*

Example II In the second set of simulation, we consider a Box-Cox transformation quantile regression model with heteroscedastic errors. The event times are generated from

$$Q_T(\tau|\mathbf{Z}) = h^{-1}_{\gamma_0(\tau)}\left(\mathbf{Z}^\top \boldsymbol{\beta}_0(\tau)\right) = h^{-1}_{\gamma_0(\tau)}\Big(b_0 + Z_1 b_1 + Z_2 b_2 + Z_3 b_3 + (1+Z_1)Q_\epsilon(\tau|\mathbf{Z})\Big).$$

where ϵ follows $N(0, 0.25^2)$. The corresponding regression quantile given $\mathbf{Z} = (1, Z_1, Z_2, Z_3)^\top$ under model (2.1) is $\boldsymbol{\beta}_0(\tau) = (Q_\epsilon(\tau) + b_0, Q_\epsilon(\tau) + b_1, b_2, b_3)^\top$. The covariates Z_1, Z_2 and Z_3 are again generated from $Unif(0,1)$, $N(0, 0.5^2)$ and $N(0, 0.5^2)$ respectively. Again we set $b_0 = b_1 = b_2 = b_3 = 1$, but the first two entries of $\boldsymbol{\beta}_0(\tau)$ are now both τ-dependent, while the last two are still constant. The transformation coefficient $\gamma_0(\tau)$ is 1 for $\tau \leq 0.4$ and 0.5 for $\tau > 0.4$. The censoring distribution is generated in the same fashion as the homogeneous case. Again we generate 500 simulated data sets of sample size 200 or 500, with resampling size 200. Table 2 reports results in the same format as in Table 1. We can see that the proposed method performs well in this heteroscedastic error case as in the first example. The standard error estimates are close to the standard deviations, and the coverage probabilities are satisfactory.

Table 3: Simulation under a log-transformed quantile regression model

		$\tau = .25$					$\tau = .5$					$\tau = .75$				
		$\hat{\gamma}$	$\hat{\beta}^{(0)}$	$\hat{\beta}^{(1)}$	$\hat{\beta}^{(2)}$	$\hat{\beta}^{(3)}$	$\hat{\gamma}$	$\hat{\beta}^{(0)}$	$\hat{\beta}^{(1)}$	$\hat{\beta}^{(2)}$	$\hat{\beta}^{(3)}$	$\hat{\gamma}$	$\hat{\beta}^{(0)}$	$\hat{\beta}^{(1)}$	$\hat{\beta}^{(2)}$	$\hat{\beta}^{(3)}$
								$n = 200$ & 40% censoring								
Proposed	Bias	-.015	.014	-.011	-.014	-.008	-.001	.023	.007	.008	.010	.002	.037	.024	.022	.017
	SD	.099	.078	.171	.139	.144	.095	.085	.171	.152	.156	.122	.131	.230	.209	.206
	MSE	.010	.006	.029	.020	.021	.009	.008	.029	.023	.024	.015	.019	.053	.044	.043
PH	Bias	-	.019	-.001	-.001	.003	-	.017	-.000	.002	.003	-	.018	.009	.002	.002
	SD	-	.059	.100	.050	.057	-	.055	.093	.048	.057	-	.067	.111	.058	.066
	MSE	-	.004	.010	.002	.003	-	.003	.009	.002	.003	-	.005	.012	.003	.004
								$n = 200$ & 20% censoring								
Proposed	Bias	-.011	.006	-.007	-.011	-.008	.002	.016	.011	.009	.008	.009	.029	.033	.026	.023
	SD	.083	.066	.147	.121	.126	.079	.073	.151	.127	.132	.093	.097	.195	.168	.169
	MSE	.007	.004	.022	.015	.016	.006	.006	.023	.016	.017	.009	.010	.039	.029	.029
PH	Bias	-	.011	.001	-.001	.000	-	.010	.002	.000	.000	-	.012	.010	.002	-.000
	SD	-	.051	.087	.045	.051	-	.049	.086	.042	.048	-	.056	.098	.050	.056
	MSE	-	.003	.008	.002	.003	-	.003	.007	.002	.002	-	.003	.010	.002	.003

*Approach PH denotes Peng and Huang (2008)'s proposal.

Table 4: Empirical rejection rate at level $\alpha = .1$ for testing if the transformation process is zero or constant

	$n = 200$ & 40% censoring			$n = 200$ & 20% censoring			$n = 500$ & 40% censoring			$n = 500$ & 20% censoring		
	I	II	III	I	II	III	I	II	III	I	II	III
$\mathcal{T}_1$	.952	.844	.060	.990	.902	.070	1.000	.996	.074	1.000	1.000	.108
$\mathcal{T}_2$	.448	.324	.034	.568	.406	.056	.904	.780	.042	.942	.854	.062

*Transformation process in Examples I and II is 1 for $\tau \leq 0.4$ and 0.5 for $\tau > 0.4$, while it is a zero process in Example III.

Example III The third example aims to show that our proposed estimator can also yield reasonable estimates when the global linear assumption holds. Event times are generated from the following log transformation quantile regression model, i.e. a Box-Cox transformed model with transformation coefficient $\gamma_0(\tau) = 0$ for all τ,

$$Q_T(\tau|\mathbf{Z}) = \exp\left(\mathbf{Z}^\top \boldsymbol{\beta}_0(\tau)\right) = \exp\Big(b_0 + Z_1 b_1 + Z_2 b_2 + Z_3 b_3 + Q_\epsilon(\tau|\mathbf{Z})\Big).$$

where $b_0 = b_1 = b_2 = b_3 = 1$ and ϵ follows $N(0, 0.25^2)$. The covariates Z_1, Z_2 and Z_3 are $Unif(0,1)$, $Unif(-1,1)$ and $N(0, 0.5^2)$ respectively. The censoring distribution is generated by $Unif(0,c)$ with $c = \exp(c_0 + Z_1 + Z_2 + Z_3)$, where c_0 is taken to be 2 and 2.7 respectively to yield a censoring rate of 40% or 20%. 500 simulated data sets of sample size 200 or 500 are generated. Table 3 shows that, despite the fact that our procedure needs to estimate additional transformation parameters, the proposed estimator performs comparably with respect to Peng and Huang (2008) method in terms of empirical bias and standard deviations.

Table 5: Simulation under a transformed quantile regression model with known transformation

		$\tau = .25$					$\tau = .5$					$\tau = .75$				
		$\hat{\gamma}$	$\hat{\beta}^{(0)}$	$\hat{\beta}^{(1)}$	$\hat{\beta}^{(2)}$	$\hat{\beta}^{(3)}$	$\hat{\gamma}$	$\hat{\beta}^{(0)}$	$\hat{\beta}^{(1)}$	$\hat{\beta}^{(2)}$	$\hat{\beta}^{(3)}$	$\hat{\gamma}$	$\hat{\beta}^{(0)}$	$\hat{\beta}^{(1)}$	$\hat{\beta}^{(2)}$	$\hat{\beta}^{(3)}$
								$n = 200$ & 40% censoring								
Proposed	Bias	-	.028	-.006	-.006	-.006	-	.015	-.013	-.019	-.016	-	.028	-.008	-.006	-.009
	SD	-	.074	.129	.073	.075	-	.074	.130	.092	.097	-	.064	.109	.071	.064
	MSE	-	.006	.017	.005	.006	-	.006	.017	.009	.010	-	.005	.012	.005	.004
WW	Bias	-	-.000	-.002	.005	.003	-	-.042	-.004	-.021	-.011	-	-.023	.005	.009	.010
	SD	-	.062	.104	.062	.063	-	.095	.163	.111	.114	-	.060	.103	.061	.058
	MSE	-	.004	.011	.004	.004	-	.011	.027	.013	.013	-	.004	.011	.004	.003
								$n = 200$ & 20% censoring								
Proposed	Bias	-	.014	-.001	-.003	-.003	-	.007	-.004	-.008	-.008	-	.018	-.001	-.003	-.002
	SD	-	.056	.099	.057	.057	-	.065	.109	.074	.079	-	.053	.093	.058	.054
	MSE	-	.003	.010	.003	.003	-	.004	.012	.006	.006	-	.003	.009	.003	.003
WW	Bias	-	.001	.002	.000	.001	-	-.021	-.006	-.011	-.009	-	-.010	.003	.002	.004
	SD	-	.055	.096	.055	.054	-	.076	.133	.089	.090	-	.051	.091	.056	.054
	MSE	-	.003	.009	.003	.003	-	.006	.018	.008	.008	-	.003	.008	.003	.003

* Approach WW denotes Wang and Wang (2009)'s proposal.

Regarding the testing procedure described in Section 3 on whether the transformation process is zero or constant, we compute the test statistic for each Monte Carlo data set, and then obtain the empirical rejection rate (ERR) by averaging over all Monte Carlo trials. In line with Peng and Huang (2008), we choose a weight function $\Xi_0(u) = 1$ for test statistic $\mathcal{T}_1$ with a null value $r_0(u) = 0$, and a weight function $\Xi_1(u) = I\{u \geq (l+u)/2\}$ for test statistic $\mathcal{T}_2$, where $l = 0.05$ and $u = 0.8$. Table 4 displays the ERR of $\mathcal{T}_1$ and $\mathcal{T}_2$ at level $\alpha = 0.1$ for Example I and II (where the transformation process is non-zero and non-constant) and for Example III (where the transformation process is a zero process). We observe that the empirical type I errors of both $\mathcal{T}_1$ and $\mathcal{T}_2$ are kept below the significance level 0.1. Similar to the results presented in Mu and He (2007), although the tests tend to be conservative when the sample size is small or the censoring is high, we observe a tendency of Type I errors getting closer to their nominal values of 0.1 as the sample size grows or as the censoring rate reduces. Meanwhile, the empirical powers in case of a misspecified null are also fairly reasonable. This can justify the use of $\mathcal{T}_1$ and $\mathcal{T}_2$ in testing respectively for a zero and a constant transformation process is proper.

Example IV The proposed method can be easily modified if the transformation parameter is known in advance, in which the regression coefficients can be estimated directly based on

Table 6: Simulation when the transformed quantile regression model is only true for $\tau = 0.5$

		$\tau = .25$					$\tau = .5$					$\tau = .75$				
		$\hat{\gamma}$	$\hat{\beta}^{(0)}$	$\hat{\beta}^{(1)}$	$\hat{\beta}^{(2)}$	$\hat{\beta}^{(3)}$	$\hat{\gamma}$	$\hat{\beta}^{(0)}$	$\hat{\beta}^{(1)}$	$\hat{\beta}^{(2)}$	$\hat{\beta}^{(3)}$	$\hat{\gamma}$	$\hat{\beta}^{(0)}$	$\hat{\beta}^{(1)}$	$\hat{\beta}^{(2)}$	$\hat{\beta}^{(3)}$
		$n = 200$ & 40% censoring														
Proposed	Bias	.007	.063	-.138	.005	.009	-.012	.025	.013	-.007	-.002	-.024	.007	.178	-.006	-.008
	SD	.242	.121	.223	.202	.211	.221	.126	.235	.211	.209	.268	.172	.346	.271	.264
	MSE	.059	.019	.069	.041	.045	.049	.016	.055	.044	.043	.073	.029	.152	.073	.070
LT	Bias	.062	.055	-.077	.109	.113	.044	.110	.187	.109	.110	-.014	-.023	.197	.062	.059
	SD	.259	.126	.260	.255	.263	.239	.150	.329	.276	.273	.276	.178	.376	.300	.297
	MSE	.071	.019	.073	.077	.082	.059	.035	.143	.088	.087	.076	.032	.180	.094	.091
		$n = 200$ & 20% censoring														
Proposed	Bias	-.021	.036	-.157	-.012	-.010	-.015	.010	.007	-.008	-.005	-.031	-.013	.166	-.013	-.014
	SD	.199	.099	.185	.164	.162	.182	.106	.193	.167	.171	.231	.148	.285	.233	.230
	MSE	.040	.011	.059	.027	.026	.033	.011	.037	.028	.029	.054	.022	.109	.054	.053
LT	Bias	.009	.052	-.118	.028	.029	-.005	.104	.127	.021	.021	-.067	-.035	.137	-.031	-.030
	SD	.217	.105	.205	.185	.184	.204	.131	.250	.210	.209	.250	.153	.301	.241	.241
	MSE	.047	.014	.056	.035	.035	.041	.028	.078	.044	.044	.067	.025	.109	.059	.059

*Approach LT denotes Leng and Tong (2014)'s proposal. Also see Remark 4 for their results in the case of $\tau = .75$.

the given transformation. The fourth set of simulation illustrates the numerical performance of the modified procedure for situations in which only the transformation parameter at the interested quantile level is known in advance. Specifically, we compare the estimation results of the proposed method and Wang and Wang (2009) under the transformed quantile regression model specified in the first example. Results summarized in Table 5 show that although the proposed method has to make additional estimation on the transformation process prior to the interested quantile levels, the empirical bias and standard deviations are still competitive against the benchmark made by Wang and Wang (2009)'s method.

Example V Although the proposed method relaxes the global linear quantile assumption by Peng and Huang (2008), it nevertheless assumes that a transformed linear quantile model holds up to the interested quantile level τ. Therefore, we investigate the performance of our proposal under the scenario that the quantile model only holds for the specific τ. Accordingly, we consider a transformed quantile regression model similar to that in Example II except

$$Q_T(\tau|\mathbf{Z}) = h^{-1}_{\gamma_0(\tau)}\left(\mathbf{Z}^\top \boldsymbol{\beta}_0(\tau)\right) = h^{-1}_{\gamma_0(\tau)}\Bigg(b_0 + Z_1 b_1 + Z_2 b_2 + Z_3 b_3 + (1 + Z_1^2) Q_\epsilon(\tau|\mathbf{Z})\Bigg).$$

Again $b_0 = b_1 = b_2 = b_3 = 1$, ϵ follows $N(0, 0.25^2)$, and the covariates Z_1, Z_2 and Z_3 are $Unif(0, 1)$, $N(0, 0.5^2)$ and $N(0, 0.5^2)$ respectively. The transformation coefficient $\gamma_0(\tau)$ is 1 for $\tau \leq 0.4$ and 0.5 for $\tau > 0.4$. Now, note that since we have not considered Z_1^2 in the covariate vector, the transformed linear quantile model does not hold here unless $\tau = 0.5$, where we have $Q_\varepsilon(\tau) = 0$ so that the last term in the last display involving Z_1^2 vanishes. We generate censoring times from $Unif(0, c)$ with $c = \exp(c_0 + Z_1^2 + Z_2 + Z_3)$, where c_0 is taken to be 1.6 and 2.3 to yield the target censoring rates of 40% and 20%, respectively. With 500 simulated data sets of sample size 200 or 500 each, numerical results obtained by our proposal are tabulated in Table 6 together with those by Leng and Tong (2014)'s method. Unsurprisingly, the proposed method produces some bias for $\tau = 0.25$ and $\tau = 0.75$ where the transformed quantile model is misspecified, specifically for the regression coefficients associated with Z_1, but the results are quite satisfactory for $\tau = 0.5$, i.e. the only quantile level for which the transformed quantile model is true. In some sense, despite the recursive nature of our proposal, this demonstrates its robustness when previous quantiles are not correctly estimated. On the other hand, although Leng and Tong (2014) only assumes the transformed quantile model at the specific quantile, this flexibility does not save them from a worse numerical performance through using a kernel-based estimator.

Example VI This example demonstrates the performance of the proposed estimator in a scenario of moderately high number of covariates. The simulation setup is similar to the first example, where the true transformation process $\gamma_0(\tau)$ equals 1 for $\tau \leq 0.4$ and 0.5 for $\tau > 0.4$. But the covariate process $\mathbf{Z} = (1, Z_1, Z_2, \cdots, Z_9)^\top$ is now 10-dimensional where $Z_1, \cdots, Z_9$ are each independent $N(0, 0.5^2)$, while the corresponding regression coefficient is $\boldsymbol{\beta}_0(\tau) = (Q_\varepsilon(\tau) + 1, 1, \cdots, 1)^\top$ and ϵ follows $N(0, 0.25^2)$. The censoring distribution is again generated by $Unif(0, c)$ where $c = \exp(c_0 + Z_1 + \cdots + Z_9)$ with c_0 equals 1.6 or 2.3 to attain a censoring rate of 40% and 20% respectively. We simulate 500 Monte Carlo data sets of sample size 200 or 500 each, and the estimation results are tabulated in Table 7.

Because of space limitation, we summarize here only the minimum, mean, and maximum values of the empirical bias, standard deviations, and mean squared errors for the $\hat{\boldsymbol{\beta}}(\tau)$ vector. Readers may refer to the supplementary materials for a complete version of Table 7 which present all the individual entries as in the previous tables. As shown, performance of the proposed estimator in terms of empirical bias and standard deviations is still fairly reasonable comparing to an existing alternative when the number of covariates increases.

One practical concern on the proposed methodology is the computation efforts required. Indeed, it is easy to see that the runtime required for the estimation of the regression coefficients given a fixed transformation parameter will be similar to that of Peng and Huang (2008), where the transformation parameter is fixed at zero for all quantile levels. Therefore, the aggregate computation effort of our proposed algorithm would be of an order of the number of grid nodes in the grid search procedure for the transformation parameter. Noteworthy, the grid search algorithm can be done by the parallel computing technique, which could linearly reduce the runtime due to the grid search step. Meanwhile, as we observe that the step size of τ in the stepwise procedure does not significantly influence the numerical performance of the proposed estimator, we adopt a wider grid size of 0.05 instead of 0.01 or 0.02 as chosen in Peng and Huang (2008) in order to reduce the computation efforts. Despite the grid search procedure we adopted on top of Peng and Huang (2008)'s type recursive algorithm, our methodology requires runtime of approximately 20 seconds to produce all parameter estimates from $\tau = 0$ to $\tau = 0.8$ for a sample size of 500 without resampling on a standard computer with i7-8550U CPU and 16GB RAM.

5 Data Analysis

As an illustration, we apply the proposed method to analyze the HMO data as in Leng and Tong (2014). Details about the study can be found in Hosmer and Lemeshow (1999). In this study, 100 HIV positive subjects were followed until death due to AIDS or related factors, until the study end or lost to follow-up. Their survival time and censoring status are reported.

Table 7: Simulation with moderately high number of parameters

		$\tau = .25$				$\tau = .5$				$\tau = .75$			
			$\{\hat{\beta}^{(i)}, i = 0, 1, \cdots, 9\}$				$\{\hat{\beta}^{(i)}, i = 0, 1, \cdots, 9\}$				$\{\hat{\beta}^{(i)}, i = 0, 1, \cdots, 9\}$		
		$\hat{\gamma}$	min	mean	max	$\hat{\gamma}$	min	mean	max	$\hat{\gamma}$	min	mean	max
$n = 200$ & 40% censoring													
Proposed	Bias	-.136	-.094	-.078	.001	-.015	-.055	-.044	-.013	-.139	-.127	-.117	-.049
	SD	.211	.110	.120	.141	.086	.093	.099	.104	.281	.185	.192	.196
	MSE	.063	.020	.021	.024	.008	.009	.012	.014	.098	.036	.051	.054
LT	Bias	-.253	-.117	-.099	-.090	-.161	-.102	-.096	-.074	-.664	-.478	-.390	-.376
	SD	.270	.154	.171	.214	.193	.134	.156	.223	.236	.115	.144	.152
	MSE	.137	.033	.039	.060	.063	.028	.034	.055	.497	.162	.174	.241
$n = 200$ & 20% censoring													
Proposed	Bias	-.136	-.094	-.078	.001	-.015	-.055	-.044	-.013	-.139	-.127	-.117	-.049
	SD	.211	.110	.120	.141	.086	.093	.099	.104	.281	.185	.192	.196
	MSE	.049	.014	.016	.017	.006	.006	.009	.010	.079	.030	.038	.039
LT	Bias	-.253	-.117	-.099	-.090	-.161	-.102	-.096	-.074	-.664	-.478	-.390	-.376
	SD	.270	.154	.171	.214	.193	.134	.156	.223	.236	.115	.144	.152
	MSE	.131	.037	.041	.045	.047	.019	.024	.026	.548	.197	.200	.208

*Approach LT denotes Leng and Tong (2014)'s proposal. Also see Remark 4 for their results in the case of $\tau = .75$.

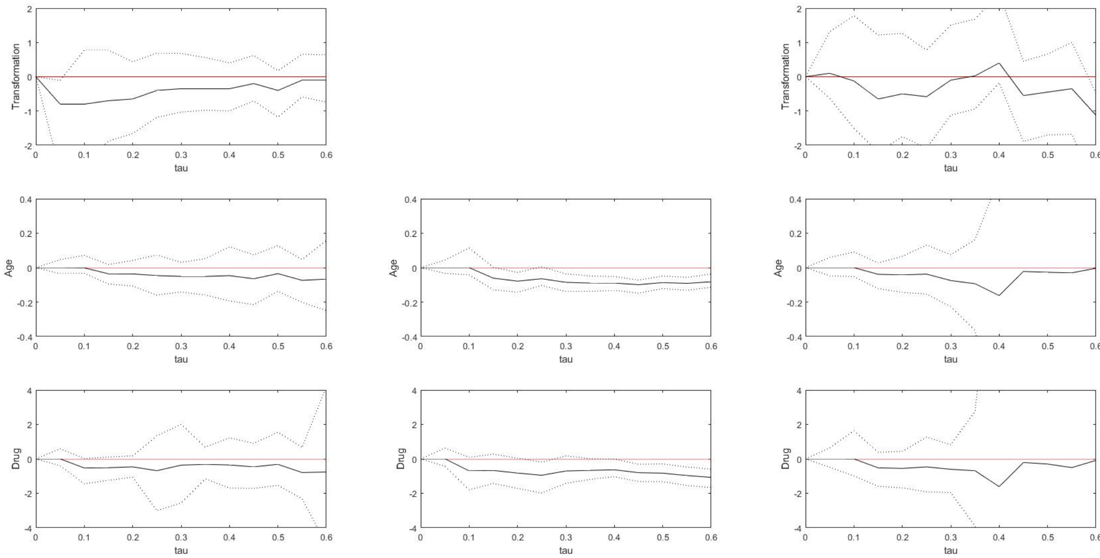
Covariates include AGE, which records patient's age in years; and DRUG, which equals 1 if the patient has a history of IV drug use and 0 otherwise. The Cox proportional hazard model and results from Leng and Tong (2014) suggest that both covariates are significant.

We examine the effects of the covariates on different quantiles of the survival time by considering a series of quantiles ranging from 0.05 to 0.6 with an increment of 0.05. We set the size of the resampling scheme to be 500, with perturbation generated from $exponential(1)$. We plot the estimated transformation parameter $\gamma(\tau)$ and the quantile regression coefficients $\boldsymbol{\beta}(\tau)$ in Figure 2. The corresponding pointwise confidence intervals computed from the proposed resampling method are also given. Plots of the parameter estimates and the confidence intervals using Peng and Huang (2008)'s and Leng and Tong (2014)'s methods are produced for comparison. We can observe that the transformation parameter estimate is approximately zero for all quantiles. It implies that a log transformation is appropriate and it agrees with the results obtained by Leng and Tong (2014). More formally, we use the test statistics $\mathcal{T}_1$ and $\mathcal{T}_2$ described in the end of Section 3 to test whether the transformation process is zero or constant. We again consider $\mathcal{T}_1$ with a null value $r_0(u) = 0$ and a weight function $\Xi_0(u) = 1$, as well as $\mathcal{T}_2$ with a weight function $\Xi_1(u) = I\{u \geq (l+u)/2\}$, where $l = 0.05$ and $u = 0.6$. The significance level is set to be $\alpha = 0.1$. The empirical p-value of the test $\mathcal{T}_1$, i.e. the empirical proportion of the 500 resampling-based $\mathcal{T}_1^*$ being greater than the actual $\mathcal{T}_1$, is 0.216, suggesting a zero transformation process. Meanwhile, since the empirical p-value of $\mathcal{T}_2$ is 0.320, it is also evident that the transformation process is constant.

Despite the need for estimating the unknown transformation parameter, our proposal still provides comparable confidence intervals with Peng and Huang (2008)'s method, which requires the assumption that $\gamma(\tau)$ is fixed at zero across all quantile levels (the global linear assumption). On the contrary, Leng and Tong (2014) obtains similar parameter point estimates but with much wider confidence intervals. It is noteworthy that the figures we obtained based on Leng and Tong (2014)'s approach are not entirely the same as the corresponding figures presented in Leng and Tong (2014). In their numerical illustration, $\gamma(\tau)$ is assumed to be zero upon which the estimation of $\boldsymbol{\beta}(\tau)$ is performed. In contrast, we estimate both transformation and effect parameters without assuming a known transformation. Recall that, in their approach, estimates of $\boldsymbol{\beta}(\tau)$ also depend on that of $\gamma(\tau)$. When $\gamma(\tau)$ is now unknown and has to be estimated, there will be extra (non-linear) variation in the estimates of $\boldsymbol{\beta}(\tau)$; the phenomenon is particularly prominent in the resampling estimation, which results in much wider confidence intervals for $\boldsymbol{\beta}(\tau)$ compared to the case where $\gamma(\tau)$ is fixed. From this point of view, Figure 2 suggests that our estimate of $\gamma(\tau)$ is more stable in the sense that narrower confidence intervals are obtained. Our proposal is thus more preferable over the two alternatives because of our flexibility to accommodate an unknown transformation process while maintaining relatively stable confidence intervals.

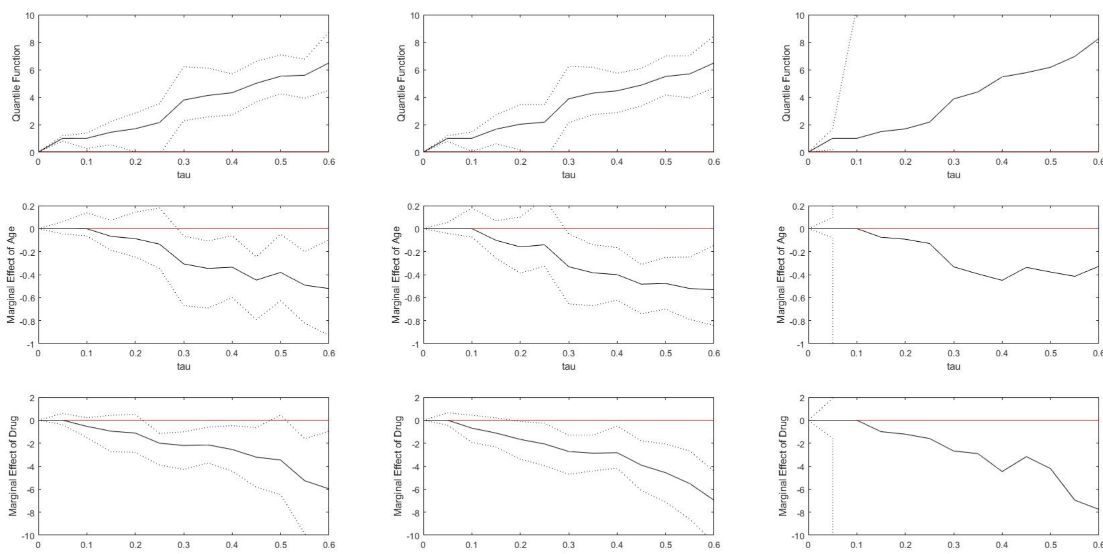
Another observation is that the quantile regression coefficients for both covariates appear to be insignificant, which contradicts to the conclusion of the Cox model. The reason is because we do not assume a log transformation of the survival time, hence the unconditional marginal effect of the covariates is not given by $\boldsymbol{\beta}(\tau)$ only, but a function of both $\gamma(\tau)$ and $\boldsymbol{\beta}(\tau)$. In particular, we follow Mu and He (2007) to assess the marginal effect by

$$\frac{\partial}{\partial \mathbf{Z}} Q_T(\tau|\mathbf{Z}) = \frac{\partial}{\partial \mathbf{Z}} h^{-1}_{\gamma(\tau)}\left(\mathbf{Z}^\top \boldsymbol{\beta}(\tau)\right) = \boldsymbol{\beta}(\tau) \cdot Q_T(\tau|\mathbf{Z})^{1-\gamma(\tau)}.$$



(a) Our Proposal (b) Peng and Huang (2008) (c) Leng and Tong (2014)

Figure 2: Estimates and Pointwise Confidence Intervals for the Transformation Parameter and the Covariates associated with Age and Prior Drug Use



(a) Our Proposal (b) Peng and Huang (2008) (c) Leng and Tong (2014)

Figure 3: Estimates and Pointwise Confidence Intervals for the Quantile Function and Marginal Effects for an Individual Aged 35 with Prior Drug Use

In Figure 3, we plot the estimated quantile function of the survival time and marginal effects of the two covariates for an individual aged 35 with prior drug use. Again we use the perturbation bootstrap method to produce corresponding pointwise confidence intervals. We observe that the quantile function and marginal effects are significant for most quantiles. The confidence intervals are also fairly stable across different quantiles and are closed to Peng and Huang (2008)'s results. Similarly, Leng and Tong (2014)'s proposal gives comparable point estimates but significantly wider confidence intervals, because of their more volatile estimates of $\gamma(\tau)$ and $\boldsymbol{\beta}(\tau)$. As a result, our method can be viewed as a better alternative through incorporating an unknown transformation process while retaining consistent numerical results.

6 Conclusion

We have proposed a class of dynamic transformed linear quantile regression models for survival data subjected to conditionally independent censoring. Using martingale theory, we construct unbiased estimating equations for the unknown transformation parameters and regression coefficients. Our algorithm involves a stepwise two-stage procedure where estimates of the transformation parameter and the regression coefficients at each subsequent quantiles are computed recursively through minimizing two L_1-type objective functions respectively. By incorporating different transformation parameters across individual quantiles, the proposed methodology relaxes the global linear assumption required in some existing literature and thus enjoys greater flexibility under the conditional independent censoring setting.

With empirical process theory, we have established the uniform consistency and weak convergence of the proposed estimator as a process of the quantile level. On the other hand, numerical studies suggest that the proposed method performs competitively desirable compared to existing proposals for sample sizes that are of practical interest. The real example presented illustrates how the simultaneous confidence intervals for the transformation parameter and the regression coefficients can be constructed. As discussed in Mu and He (2007), it is worth noted that the regression coefficients shall not be directly comparable because of the different transformation parameter estimated across various quantiles. For a meaningful comparison, it would be reasonable to consider the marginal effects of the covariates, which combine the influence of the transformation parameter and the regression coefficients.

In this work, we have focused our discussion on ordinary right-censored time-to-event data. In practical study, however, lifetime data may be biased sampled in nature due to study design or data collecting mechanism. Special treatments are required to tackle data with various bias sampling scheme. Kim et al. (2013) and Kim et al. (2016) discuss the semiparametric transformation model and the accelerated failure time model under general bias sampling scheme respectively. From the perspective of censored quantile regression, Xu et al. (2017) developed a martingale based estimating procedure for various types of biased data. However, a similar global linear assumption as in Peng and Huang (2008) has to be included, which limits the model flexibility. This direction of research merits further investigations.

Acknowlwedgement

The authors would like to acknowledge the editor, the associate editor and the anonymous referees whose constructive and valuable comments have substantially improved the manuscript. Sit's work was partially supported by Hong Kong Research Grant Council RGC-14301618 and RGC-14317716. Xu's work was partially supported by SES-1659328, SES-1846747 and DMS-1712717.

References

Box, G. E. P. and Cox, D. R. (1964), "An analysis of transformations," *Journal of the American Statistical Association*, 26, 211–52.

Buckinsky, M. (1995), "Quantile Regression, Box-Cox Transformation Model, and the U.S. Wage Structure," *Journal of Econometrics*, 65, 109–154.

Chamberlain, G. (1994), "Quantile Regression, Censoring and the Structure of Wages," in *Advances in Econometrics, New York: Cambridge University Press*, ed. Sims, C., pp. 171–209.

Cox, D. R. (1972), "Regression models and life-tables," *Journal of the Royal Statistical Society: Series B*, 34, 187–220.

— (1975), "Partial likelihood," *Biometrika*, 62, 269–76.

De Backer, M., El Ghouch, A., and Van Keilegom, I. (2017), "An Adapted Loss Function for Censored Quantile Regression," ArXiv:1703.07975.

Engle, R. F. and Maganelli, S. (2004), "CAViaR: Conditional autoregressive Value at Risk by regression quantiles," *Journal of Business and Economics Statistics*, 22, 367–81.

Fleming, T. R. and Harrington, D. P. (2005), *Counting Processes and Survival Analysis*, John Wiley & Sons.

Fygenson, M. and Ritov, Y. (1994), "Monotone Estimating Equations for Censored Data," *The Annals of Statistics*, 22, 732–46.

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," *Annals of Statistics*, 38, 1607–1637.

Jin, Z., Lin, D. Y., Wei, L. J., and Ying, Z. (2003), "Rank-based inference for the accelerated failure time model," *Biometrika*, 90, 341–53.

Jin, Z., Ying, Z., and Wei, L. J. (2001), "A Simple Resampling Method by Perturbing the Minimand," *Biometrika*, 88, 381–390.

Kim, J. P., Lu, W., Sit, T., and Ying, Z. (2013), "A unified approach to semiparametric transformation models under general biased sampling schemes," *Journal of the American Statistical Association*, 108, 217–27.

Kim, J. P., Sit, T., and Ying, Z. (2016), "Accelerated failure time model under general biased sampling scheme," *Biostatistics*, 17, 576–88.

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

Koenker, R. and Bassett, G. (1978), "Regression quantiles," *Econometrica*, 46, 33–50.

Koenker, R., Chernozhukov, V., He, X., and Peng, L. (eds.) (2017), *Handbook of Quantile Regression*, Chapman & Hall/CRC Handbooks of Modern Statistical Methods.

Koenker, R. and Hallock, K. (2001), "Quantile regression," *Journal of Economic Perspectives*, 15, 143–56.

Leng, C. and Tong, X. (2014), "Censored quantile regression via Box-Cox transformation under conditional independence," *Statistica Sinica*, 24, 221–49.

Mu, Y. and He, X. (2007), "Power transformation toward a linear regression quantile," *Journal of the American Statistical Association*, 102, 269–79.

Peng, L. and Huang, Y. (2008), "Survival analysis with quantile regression models," *Journal of the American Statistical Association*, 103, 637–49.

Peng, L., Xu, J., and Kutner, N. (2014), "Shrinkage estimation of varying covariate effects based on quantile regression," *Statistics and Computing*, 24, 853–69.

Portnoy, S. (2003), "Censored regression quantiles," *Journal of the American Statistical Association*, 98, 1001–12.

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

— (1986), "Censored regression quantiles," *Journal of Econometrics*, 32, 143–55.

Tsiatis, A. A. (1990), "Estimating regression parameters using linear rank tests for censored data," *The Annals of Statistics*, 18, 354–72.

Wang, H. and Wang, L. (2014), "Quantile regression analysis of length-biased survival data," *Stat*, 3, 31–47.

Wang, H., Zhou, J., and Li, Y. (2013), "Variable selection for censored quantile regression," *Statisica Sinica*, 23, 145–67.

Wang, H. J. and Wang, L. (2009), "Locally weighted censored quantile regression," *Journal of the American Statistical Association*, 104, 1117–28.

Wei, Y. and He, X. (2006), "Conditional growth charts (with discussions)," *Annals of Statistics*, 34, 2069–97 and 2126–31.

Wei, Y., Pere, A., Koenker, R., and He, X. (2006), "Quantile regression methods for reference growth charts," *Statistics in Medicine*, 25, 1369–82.

Wu, Y. and Yin, G. (2013), "Cure rate quantile regression for censored data with a survival fraction," *Journal of the American Statistical Association*, 108, 1517–31.

Xu, G., Sit, T., Wang, L., and Huang, C.-Y. (2017), "Estimation and Inference of Quantile Regression for Survival Data under Biased Sampling," *Journal of the American Statistical Association*, To appear.

Yin, G., Zeng, D., and Li, H. (2008), "Power-transformed linear quantile regression with censored data," *Journal of the American Statistical Association*, 103, 1214–24.

Ying, Z. (1993), "A large sample study of rank estimation for censored regression data," *The Annals of Statistics*, 21, 76–99.

Ying, Z., Jung, S. H., and Wei, L. J. (1995), "Survival analysis with median regression models," *Journal of the American Statistical Association*, 90, 178–84.

Zheng, Q., Peng, L., and He, X. (2017), "High dimensional censored quantile regression," *Annals of Statistics*, To appear.