Efficient Estimation for Censored Quantile Regression

Sze Ming Lee*, Tony Sit* and Gongjun Xu†

*Department of Statistics, The Chinese University of Hong Kong

†Department of Statistics, University of Michigan

1155049067@link.cuhk.edu.hk tonysit@sta.cuhk.edu.hk gongjun@umich.edu

Abstract

Censored quantile regression (CQR) has received growing attention in survival analysis because of its flexibility in modeling heterogeneous effect of covariates. Advances have been made in developing various inferential procedures under different assumptions and settings. Under the conditional independence assumption, many existing CQR methods can be characterized either by stochastic integral-based estimating equations (see, for example, Peng and Huang, 2008, *JASA* **103,** 637–49) or by locally weighted approaches to adjust for the censored observations (see, for instance, Wang and Wang, 2009, *JASA* **104,** 1117–28). While there have been proposals of different apparently dissimilar strategies in terms of formulations and the techniques applied for CQR, the inter-relationships amongst these methods are rarely discussed in the literature. In addition, given the complicated structure of the asymptotic variance, there has been limited investigation on improving the estimation efficiency for censored quantile regression models. This article addresses these open questions by proposing a unified framework under which many conventional approaches for CQR are covered as special cases. The new formulation also facilitates the construction of the most efficient estimator for the parameters of interest amongst a general class of estimating functions. Asymptotic properties including consistency and weak convergence of the proposed estimator are established via the martingale-based argument. Numerical studies are presented to illustrate the promising performance of the proposed estimator as compared to existing contenders under various settings.

Keywords: Survival analysis; Estimation efficiency; Check function; Martingale; Kernel estimation.

Sze Ming Lee is M.Phil. student, Department of Statistics, The Chinese University of Hong Kong, Shatin, NT, Hong Kong SAR (E-mail: *1155049067@link.cuhk.edu.hk*). Tony Sit is Associate Professor, Department of Statistics, The Chinese University of Hong Kong, Shatin, NT, Hong Kong SAR (E-mail: *tonysit@sta.cuhk.edu.hk*). Gongjun Xu is Associate Professor, Department of Statistics, University of Michigan, Ann Arbor, MI 48109 (E-mail: *gongjun@umich.edu*).

1 Introduction

Quantile regression (QR) has become a powerful and popular technique to model the conditional distribution given the covariates since its introduction in Koenker and Bassett (1978); readers are referred to Koenker (2005) and, more recently, Koenker et al. (2017) for an extensive elaboration and literature review on the corresponding development. Given its versatility, quantile regression model has been a powerful tool for solving problems in various disciplines, including longitudinal study (Wang and Fygenson, 2009; Ma and Wei, 2012), growth chart (Wei and He, 2006; Wei, 2008), risk management (Engle and Maganelli, 2004) amongst others.

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

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

where $Q_\tau(Y \mid \mathbf{Z})$ denotes the τ-th conditional quantile of Y given $\mathbf{Z}$. Given the observations $\{(\mathbf{Z}_i, Y_i)\}_{i=1,\ldots,n}$, an estimate of $\boldsymbol{\beta}(\tau)$ in (1.1) can be obtained via the least absolute deviation approach which minimizes

$$\hat{\boldsymbol{\beta}}(\tau) = \arg\min_{\mathbf{b}\in\mathbb{R}^p} \sum_{i=1}^{n} \rho_\tau(Y_i - \mathbf{Z}_i^\top \mathbf{b}),$$

where $\rho_\tau(u) = u\{\tau - I(u < 0)\}$.

In addition to its applications in health statistics and economics, the quantile regres-

sion model has also been a valuable and effective tool in survival analysis with censored observations. The most prominent advantage of censored quantile regression is its ability to accommodate heterogeneous effects of 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; Wei et al., 1990; Jin et al., 2003). Furthermore, the conditional quantile of the survival time is easier to interpret than the hazard function and is often of the ultimate interest of end users. Under the independent censoring assumption, Ying et al. (1995) proposed a semi-parametric procedure for median regression. To relax the independent censoring assumption and impose a less restrictive conditional independent censoring assumption instead, Portnoy (2003) proposed a recursively reweighted inference procedure based on the principle of the Kaplan-Meier (KM) estimate's self-consistency, which can be interpreted as shifting masses of censored data points to the right as in the sense of Efron (1967). Since then, many interesting proposals have been discussed to address various challenges for inference.

Typically, these proposals establish mean-zero estimation equations either via adopting a counting process approach or by introducing a weight adjustment, which depends on an estimate of the unknown underlying distributions, to incorporate the partial information due to censoring. For instance, Peng and Huang (2008) insightfully exploited the martingale representation of the Nelson-Aalen estimator for the cumulative hazard function and proposed a recursive series of estimating equations for not only one point but instead a sequence of quantiles under the global linear assumption, i.e. (1.1) holds for all $\tau \in (0, \tau_u] \subset (0, 1)$. Huang (2010) provided a numerically stable and computationally efficient algorithm for the aforementioned framework. Leng and Tong (2014) considered a kernel-based estimator for the distribution function instead of the cumulative hazard function so as to avoid imposing the global linear assumption.

Another promising approach is the introduction of local weight functions which can cor-

rect the bias due to censored data. Wang and Wang (2009) proposed a novel locally weighted approach for censored quantile regression (CQR). This wise formulation leverages the efficient check function proposed in Koenker and Bassett (1978) for uncensored data while for the censored observations, the kernel-based local weight proposed adjusts for the redistribution of mass due to censoring. More recently, De Becker et al. (2019) modified the check function with a proper adjustment based on an integral of the survival distribution of the censoring time.

Despite the fact that there has been tremendous effort on solving the related inference problems for CQR, there has been little discussion in the literature on how these individual estimation methods are related to each other, and the inter-relationships amongst them remain unclear. Furthermore, given the convoluted structure of the asymptotic variance, the study of estimation efficiency of CQR has so far been limited; the question of how to construct an efficient estimator for CQR remains largely unexplored. This work addresses these open problems and our main contribution is two-fold:

1. Firstly, we aim to provide a unified framework under which many existing methodologies for CQR that adopt either the martingale (Peng and Huang, 2008; Leng and Tong, 2014) or the weight adjustment approach (Wang and Wang, 2009) are covered. Indeed, the connection between these two classes of approaches is reasonable because both classes of inference procedures are carried out via optimization with respect to the unweighted check function, or its equivalent that originates from the corresponding counting process evaluated at the target quantile value, for the observed failures. Meanwhile, partial information conveyed in the censored observations are usually handled by suitable adjustments whose values depend on the underlying distributions of the failure and the censoring time distributions, which can be translated into the compensator of the aforementioned counting process after suitable transformation and simplification.

2. Secondly, we further devise the most efficient weight to improve statistical efficiency amongst this proposed unified framework. The optimal weight is shown to be depend

on the unknown conditional density functions of the survival and censoring times. To ensure that our methodology is practically feasible, we propose a kernel-based estimator for the optimal weight. We further show that the estimator adopting the estimated optimal weight theoretically attains the minimum variance and numerically outperforms existing methods under various settings in simulation studies as well as our data analysis. To the best of our knowledge, it is the first attempt to study the estimation efficiency issue for censored quantile regression in the literature.

The rest of the paper is organized as follows. In Section 2, we first construct a set of martingale-based estimation equations to motivate our new quantile regression estimator. We also discuss its connection with its predecessors. We then present the consistency and asymptotic results in Section 3 for the case where the optimal weight is incorporated. Our numerical simulations reported in Section 4 also agree with our theoretical development. The results presented also demonstrate that our proposal works decently as compared with other contenders. Section 5 illustrates the effectiveness of our method via an analysis of gastric adenocarcinoma patient data. We conclude with final remarks in Section 6. All the technical proofs are presented in the Supplementary Materials.

2 Methodology and Model Setup

2.1 Estimation

Let T be the failure time of interest, C the censoring time, and $\mathbf{Z}$ the p-vector of covariates. With right censoring, we only observe $\tilde{T} = \min(T, C)$ and we denote the censoring indicator as $\Delta = I(T \leq C)$. The observed data consist of n i.i.d. replicates of $(\tilde{T}, \Delta, \mathbf{Z})$, namely $\{(\tilde{T}_i, \Delta_i, \mathbf{Z}_i)\}_{i=1,\ldots,n}$. Instead of imposing the global linear assumption to the following censored linear quantile model for all quantile levels simultaneously, we only assume that it holds at a specific quantile level of interest $\tau \in (0, 1)$:

$$Q_\tau\left(\log(T) \mid \mathbf{Z}\right) = \boldsymbol{\beta}_0(\tau)^\top \mathbf{Z}, \tag{2.1}$$

where $\boldsymbol{\beta}_0(\tau)$ is a $p \times 1$ vector of unknown regression parameter. In the sequel, we suppress the τ in $\boldsymbol{\beta}_0(\tau)$ whenever there is no ambiguity. The above quantile regression model (2.1) also implies that, for this fixed τ,

$$\log(T) = \boldsymbol{\beta}_0(\tau)^\top \mathbf{Z} + \epsilon,$$

where ϵ is a random error whose τth conditional quantile given $\mathbf{Z}$ is zero. This expression indicates that the quantile regression can be regarded as a generalization of the AFT model with a potentially heterogeneous error. Since the conditional distribution of the failure time T given $\mathbf{Z}$ is a function of the error's distribution, it is natural to model the residual directly for (2.1). As a result, for $i = 1, \ldots, n$, we denote $\epsilon_i(\mathbf{b}) = \log(T_i) - \mathbf{b}^\top \mathbf{Z}_i$ and $e_i(\mathbf{b}) = \log(\tilde{T}_i) - \mathbf{b}^\top \mathbf{Z}_i$ the true error and the observed residual given model parameter estimates $\mathbf{b} \in \mathbb{R}^p$, respectively. Based on the observed residuals $e_i(\mathbf{b})$, we further define the counting process and the at-risk process as $N_i(\mathbf{b}, t) = \Delta_i I(e_i(b) \leq t) = \Delta_i I\left(\log(\tilde{T}_i) - \mathbf{b}^\top \mathbf{Z}_i \leq t\right)$ and $Y_i(\mathbf{b}, t) = I\left(\log(\tilde{T}_i) - \mathbf{b}^\top \mathbf{Z}_i \geq t\right)$, respectively. To emphasize that these quantities are evaluated at the true parameter values $\mathbf{b} = \boldsymbol{\beta}_0$, we suppress the notation $\mathbf{b}$ and reserve ϵ_i, e_i, $N_i(t)$ and $Y_i(t)$, respectively for simplicity.

With $\Lambda_0(t \mid \mathbf{Z}_i)$ as the conditional cumulative hazard function of ϵ_i given $\mathbf{Z}_i$, we can construct the following mean zero martingale process; see, for example, Fleming and Harrington (2005) and Peng and Huang (2008):

$$M_i(t) = N_i(t) - \int_{-\infty}^{t} Y_i(u) d\Lambda_0(u \mid \mathbf{Z}_i).$$

Based on the martingale property of $M(\cdot)$ as well as the quantile property that $\Lambda_0(0 \mid \mathbf{Z}_i) = -\log(1-\tau)$, we propose a general family of weighted estimating equations with the weight

function $\phi\left(\mathbf{Z}, \Lambda_0(t \mid \mathbf{Z})\right)$:

$$E\left(\frac{1}{n}\sum_{i=1}^{n}\mathbf{Z}_i\left\{\int_{-\infty}^{0}\phi\left(\mathbf{Z}_i, \Lambda_0(t \mid \mathbf{Z}_i)\right)dN_i(t) - \int_{-\infty}^{0}Y_i(t)\phi\left(\mathbf{Z}_i, \Lambda_0(t \mid \mathbf{Z}_i)\right)d\Lambda_0(t \mid \mathbf{Z}_i)\right\}\right) = 0, \tag{2.2}$$

which is equivalent to

$$E\left(\frac{1}{n}\sum_{i=1}^{n}\mathbf{Z}_i\Big\{\phi(\mathbf{Z}_i, \Lambda_0(e_i \mid \mathbf{Z}_i))\Delta_i I(e_i \le 0) - \Phi\left(\mathbf{Z}_i, H_\tau(\Lambda_0(e_i \mid \mathbf{Z}_i))\right) + \Phi(\mathbf{Z}_i, 0)\Big\}\right) = 0, \tag{2.3}$$

where $\Phi(\mathbf{Z}, t) = \int_{-\infty}^{t}\phi(\mathbf{Z}, s)ds$ and $H_\tau(t) = t \wedge \{-\log(1-\tau)\}$.

It is noteworthy that the proposed weight function $\phi\left(\mathbf{Z}, \Lambda_0(t \mid \mathbf{Z})\right)$ is generic in the sense that both the covariate and the time effects are taken into account. To make the best use of the quantile property, we consider $\Lambda_0(t \mid \mathbf{Z})$, a transformed version of the time t, in this weight function instead of directly using the original domain t itself.

The above martingale-based estimation approach offers a unified framework that connects many of the existing proposals for censored quantile regression modeling, including the martingale-based approach (Peng and Huang, 2008; Leng and Tong, 2014) and the local weight adjustment approach (Wang and Wang, 2009). For the benefit of better illustrating our construction, we first assume both $\Lambda_0(t \mid \mathbf{Z})$ and $\phi(\mathbf{Z}, \Lambda_0(t \mid \mathbf{Z}))$, and hence, $\Phi(\mathbf{Z}, \Lambda_0(t \mid \mathbf{Z}))$, are known. Estimation of these unknown quantities shall be discussed in details in Section 2.3.

Connection with martingale-based approach It is easy to see that, with the unity weight, i.e. $\phi(\mathbf{Z}, \Lambda_0(t \mid \mathbf{Z})) \equiv 1$, our estimating equation (2.2) is reduced to

$$E\left(\frac{1}{n}\sum_{i=1}^{n}\mathbf{Z}_i\left\{N_i(0)-\int_{-\infty}^{0}Y_i(t)d\Lambda_0(t\mid\mathbf{Z}_i)\right\}\right)$$
$$=E\left(\frac{1}{n}\sum_{i=1}^{n}\mathbf{Z}_i\left\{N_i(0)-H_\tau\left(\Lambda_0(e_i\mid\mathbf{Z}_i)\right)\right\}\right)=0. \quad (2.4)$$

This special formulation clearly resembles the insightful formulation proposed in Peng and Huang (2008) except that we need to directly estimate the associated compensator instead of adopting an iterative approach to re-express the quantity as a function of the previous quantile values under the global linear assumption on the quantiles at different levels. In spite of structural similarities, one difference between (2.4) and the corresponding estimating equation proposed in Leng and Tong (2014) is that, instead of estimating $\Lambda_0(\cdot \mid \mathbf{Z})$, Leng and Tong (2014) proposed a kernel-based approach for evaluating $-\log(1-F_0(\cdot \mid \mathbf{Z}))$. The one-to-one correspondence between the cumulative hazard function and the distribution function of the error surprisingly does not lead to similar performance given by these two methods. As we can see later in Section 4, our numerical experience suggests that estimating the cumulative hazard function directly produces more stable and efficient finite-sample results, in particular for high quantile levels.

Connection with local weight adjustment approach To make the connection evident, it is helpful to re-examine the weighted estimating equation (2.2) with the weight $\Phi(\mathbf{Z}, t) = \exp\{\Lambda_0(t \mid \mathbf{Z})\}$. Specifically, we can write

$$E\left(\frac{1}{n}\sum_{i=1}^{n}\mathbf{Z}_i\left\{\int_{-\infty}^{0}\exp\{\Lambda_0(t\mid\mathbf{Z})\}\,dN_i(t)-\int_{-\infty}^{0}Y_i(t)\exp\{\Lambda_0(t\mid\mathbf{Z})\}\,d\Lambda_0(t\mid\mathbf{Z}_i)\right\}\right)=0,$$

which is equivalent to

$$E\left(\frac{1}{n}\sum_{i=1}^{n}u_i\right) =: E\left(\frac{1}{n}\sum_{i=1}^{n}\mathbf{Z}_i\big[\exp\{\Lambda_0(e_i \mid \mathbf{Z}_i)\}\Delta_i I(e_i \le 0) - \exp\left\{H_\tau\left(\Lambda_0(e_i \mid \mathbf{Z}_i)\right)\right\} + 1\big]\right) = 0. \tag{2.5}$$

Through straightforward algebra, one can observe that u_i in (2.5) can be simplified as

$$u_i = \begin{cases} \mathbf{Z}_i\{I(e_i \le 0) - \tau\}(1-\tau)^{-1} & , \text{ when } \Delta = 1 \\ \mathbf{Z}_i\big[1 - \exp\left\{H_\tau\left(\Lambda_0(e_i \mid \mathbf{Z}_i)\right)\right\}\big] & , \text{ otherwise} \end{cases}$$

in which case the estimating equation (2.5) becomes

$$S_n(\mathbf{b}) = \frac{1}{n}\sum_{i=1}^{n}\mathbf{Z}_i\big(\Delta_i\left\{I\left(e_i(\mathbf{b}) \le 0\right) - \tau\right\} + (1-\Delta_i)(1-\tau)\left[1 - \exp\left\{H_\tau\left(\Lambda_0(e_i \mid \mathbf{Z}_i)\right)\right\}\right]\big) = 0. \tag{2.6}$$

As a result of its monotonicity with respect to $\boldsymbol{\beta}$, solving (2.6) can further be translated into a convex optimization problems. More specifically, we can obtain the root that solves (2.6), say $\breve{\boldsymbol{\beta}}^{(1)}$, by minimizing the following weighted quantile regression objective function

$$L(\mathbf{b}, \Lambda_0) = \frac{1}{n}\sum_{i=1}^{n}\left\{\Delta_i\rho_\tau\left(e_i(\mathbf{b})\right) + (1-\Delta_i)(1-\tau) \times \rho_\tau\left(\tilde{T}^* - \left[\frac{\exp\left\{H_\tau\left(\Lambda_0(e_i \mid \mathbf{Z}_i)\right)\right\} - 1}{\tau}\mathbf{b}^\top\mathbf{Z}_i\right]\right)\right\} \tag{2.7}$$

with respect to $\mathbf{b}$, where $\tilde{T}^*$ is a large constant. It is interesting to note that for uncensored data, i.e. for cases in which all the Δ_i's values are 1, our proposed estimating equation is reduced to $n^{-1}\sum_{i=1}^{n}\rho_\tau\left(e_i(\mathbf{b})\right)$, which coincides with the well-known estimating equation proposed in Koenker and Bassett (1978).

It is also interesting to find that when $\Lambda_0(\cdot \mid \mathbf{Z}_i)$ is known, the estimator proposed by

Wang and Wang (2009) is equivalent to the above estimator $\breve{\boldsymbol{\beta}}^{(1)}$. Specifically, in addition to the obvious equivalence up to the sign for cases with $\Delta_i = 1$, one can also show that when $\Delta_i = 0$, the gradient of the estimator proposed by Wang and Wang (2009) equals

$$\mathbf{Z}_i \left\{ \tau - \frac{\tau - F_0(e_i \mid \mathbf{Z}_i)}{1 - F_0(e_i \mid \mathbf{Z}_i)} I(e_i < 0) \right\} = \mathbf{Z}_i \left\{ \frac{(1-\tau) F_0(e_i \mid \mathbf{Z}_i)}{1 - F_0(e_i \mid \mathbf{Z}_i)} I(e_i < 0) + \tau I(e_i \geq 0) \right\}. \quad (2.8)$$

On the other hand, when $\Delta_i = 0$, our proposed estimating equation can be written as

$$\begin{aligned} & \mathbf{Z}_i (1-\tau)[1 - \exp\{\Lambda_0(e_i \mid \mathbf{Z}_i) \wedge -\log(1-\tau)\}] \\ = \; & \mathbf{Z}_i (1-\tau) \left\{ 1 - \frac{1}{1 - F_0(e_i \mid \mathbf{Z}_i)} \wedge \frac{1}{1-\tau} \right\} \\ = \; & -\mathbf{Z}_i \left\{ \frac{(1-\tau) F_0(e_i \mid \mathbf{Z}_i)}{1 - F_0(e_i \mid \mathbf{Z}_i)} \wedge \tau \right\}. \end{aligned} \quad (2.9)$$

Hence (2.8) and (2.9) also differ by a negative sign as the event $\{F_0(e_i \mid \mathbf{Z}_i)(1-\tau)\{1 - F_0(e_i \mid \mathbf{Z}_i)\}^{-1} \geq \tau\}$ is equivalent to the event $\{e_i \geq 0\}$. In other words, the estimating equation proposed in Wang and Wang (2009) and u_i are equivalent when $\Lambda_0(\cdot \mid \mathbf{Z}_i)$ is known. For cases where the cumulative hazard function has to be estimated, we can also compare our vanilla estimator, say $\tilde{\boldsymbol{\beta}}^{(1)}$, which solves (2.14) with $\phi(\mathbf{Z}) = 1$, to those considered in Leng and Tong (2014) and De Becker et al. (2019). One can show that the three estimators share the same asymptotic variance in which case these formulations can also be considered as asymptotically equivalent. A more elaborated discussion on this generalization will be included in Remark 2 in Section 3.

2.2 Efficient Censored Quantile Regression

The previous section considers specific choices of the weight function that unify various censored quantile regression approaches into one framework. We also demonstrate that our formulation can be regarded as a generalization of Koenker and Bassett (1978) for censored observations. This subsection is devoted to the construction of the most efficient weight

for the censored quantile regression (2.1) amongst a generic class of candidate weight functions. Given an arbitrary weight function $\phi(\mathbf{Z}, \Lambda_0(t \mid \mathbf{Z}))$, we denote generically $\breve{\boldsymbol{\beta}}^{(\phi(\mathbf{Z},\Lambda_0))}$ as a solution that solves (2.3) while assuming the cumulative hazard function $\Lambda_0(e \mid \mathbf{Z})$ is known, which is indeed not possible in real practice. This unknown quantity can, however, be replaced by a consistent estimator. In this work, we focus on a general family of kernel estimators, denoted by $\tilde{\Lambda}(e_i \mid \mathbf{Z}_i)$, which will be defined in (2.12) in Section 2.3. Correspondingly, we denote $\tilde{\boldsymbol{\beta}}^{(\phi(\mathbf{Z},\tilde{\Lambda}))}$ as a solution to (2.3) where the estimator $\tilde{\Lambda}(e_i \mid \mathbf{Z}_i)$ is used in lieu of $\Lambda_0(\cdot \mid \mathbf{Z}_i)$. Under this setting, we first present the following proposition which states that the asymptotic variance of $\tilde{\boldsymbol{\beta}}^{(\phi(\mathbf{Z},\tilde{\Lambda}))}$ only depends on the value of $\phi(\mathbf{Z}, -\log(1-\tau))$, a property that also holds in the ideal case when $\Lambda_0(e \mid \mathbf{Z})$ is known. In other words, the component $\tilde{\Lambda}(e_i \mid \mathbf{Z}_i)$ in the weight function has no effect on the asymptotic variance of the estimator.

Proposition 1. *Assume the regularity conditions in Section 3 hold. For any weight function of the form* $\phi(\mathbf{Z}, \tilde{\Lambda})$, $n^{1/2}\left(\tilde{\boldsymbol{\beta}}^{(\phi(\mathbf{Z},\tilde{\Lambda}))} - \boldsymbol{\beta}_0\right)$ *has the same limiting normal distribution as that obtained with the weight* $\phi(\mathbf{Z}, -\log(1-\tau))$.

Readers are referred to Section F in the Supplementary Material for the related proof. The result above first provides us with an insight that $\tilde{\Lambda}$ in the weight function has no effect on the variance of the limiting distribution of $n^{1/2}\left(\tilde{\boldsymbol{\beta}}^{(\phi(\mathbf{Z},\tilde{\Lambda}))} - \boldsymbol{\beta}_0\right)$. A more extensive argument for its asymptotic normality is established in Theorem 2 in Section 3. As a result of Proposition 1, we are motivated to consider applying $\mathbf{Z}$-measurable weight $\phi(\mathbf{Z})$ in (2.3) and further evaluate an optimal weight via careful examination of the structure of the asymptotic variance associated with $\phi(\mathbf{Z})$. In the sequel, we consider the estimator $\tilde{\boldsymbol{\beta}}^{(\phi(\mathbf{Z}))}$ only, which refers to the estimator derived based on the weight $\phi(\mathbf{Z})$ instead of $\phi(\mathbf{Z}, \Lambda(t \mid \mathbf{Z}))$.

For uncensored data, weighted quantile regression can bring in efficiency improvement if the conditional densities of the response are heterogeneous; see pp.160 of Koenker (2005). With censored observations, the asymptotic variance form suggests that adding weight can improve the efficiency when the conditional density of either ϵ or $\log(C) - \boldsymbol{\beta}_0^\top \mathbf{Z}$ is heteroge-

neous. In the sequel, we define $\phi^{opt}(\mathbf{Z})$ as the optimal weight that minimizes the asymptotic variance. Let $f_0(t \mid \mathbf{Z}_i)$ and $\lambda_0(t \mid \mathbf{Z}_i)$ be the conditional density and hazard function of ϵ given $\mathbf{Z}$ respectively. Similarly, we adopt $G_0(t \mid \mathbf{Z})$ to denote the conditional distribution functions of $\log(C) - \boldsymbol{\beta}_0^\top \mathbf{Z}$ given $\mathbf{Z}$. Based upon the following estimating equation,

$$E\left[\phi(\mathbf{Z}_i)\mathbf{Z}_i\left\{\Delta_i\{I(e_i \le 0) - \tau\} + (1-\Delta_i)(1-\tau)\left(1 - \exp[H_\tau\{\tilde{\Lambda}(e_i \mid \mathbf{Z}_i)\}]\right)\right\}\right] = 0 \tag{2.10}$$

and by invoking a multivariate generalization of the Cauchy-Schwarz inequality due to Tripathi (1999), one can deduce that the asymptotic variance of $\tilde{\boldsymbol{\beta}}^{(\phi(\mathbf{Z}))}$ is minimized by taking

$$\phi^{opt}(\mathbf{Z}) = \frac{p(0,\mathbf{Z})}{P(0,\mathbf{Z})}, \tag{2.11}$$

where $p(u,\mathbf{Z}) = \lambda_0(u \mid \mathbf{Z})[\{1 - F_0(u \mid \mathbf{Z})\}\{1 - G_0(u \mid \mathbf{Z})\}]^{-1}$ and $P(0,\mathbf{Z}) = \int_{-\infty}^{0} p(u,\mathbf{Z})du$; see Theorem 3 in Section 3. It should be noted that for uncensored data, one can show that $p(0,\mathbf{Z}) = f_0(0 \mid \mathbf{Z})(1-\tau)^{-2}$ and $\int_{-\infty}^{0} p(u,\mathbf{Z})du = \tau(1-\tau)^{-1}$, in which case $\phi^{opt}(\mathbf{Z})$ defined in (2.11) is equivalent to $f_0(0 \mid \mathbf{Z})$, which coincides with the most efficient $\mathbf{Z}$-weight discussed in Koenker (2005). This optimal weight also resembles the optimal weight function for the AFT model (see, for example, Tsiatis, 1990 and Lin and Chen, 2013), but both the numerator and the denominator of $\phi^{opt}(\mathbf{Z})$ involve the conditional distribution functions $F_0(\cdot \mid \mathbf{Z})$ and $G_0(\cdot \mid \mathbf{Z})$ of ϵ and $\log C - \boldsymbol{\beta}_0^\top \mathbf{Z}$, respectively, in addition to the conditional hazard rate function $\lambda_0(\cdot \mid \mathbf{Z})$. This is natural because for the quantile regression model, we no longer enjoy the homogeneity of the residual distribution across all quantile levels. The form of the optimal weight (2.11) concurs with our intuition that it should be quantile level τ specific.

2.3 Computation Issues

The proposed weighted estimating equation (2.10) with (2.11) as the optimal weight is feasible for implementation only when the unknown quantities including both the true cumulative hazard of the error $\Lambda_0(\cdot \mid \mathbf{Z})$ and the optimal weight function $\phi^{opt}(\mathbf{Z})$ can be evaluated. In

this subsection, we discuss how these estimates can be obtained.

Estimation of $\Lambda_0(\cdot \mid \mathbf{Z})$ Inspired by the strategies adopted in Wang and Wang (2009), we propose the use of kernel to estimate $\Lambda_0(e_i \mid \mathbf{Z}_i)$ nonparametrically using a localized version of the Nelson-Aalen estimator

$$\tilde{\Lambda}(e_i \mid \mathbf{Z}_i) = \sum_{j=1}^{n} \frac{B_{h,j}(\mathbf{Z}_i)\Delta_j I(\log \tilde{T}_j \leq \log \tilde{T}_i)}{\sum_{r=1}^{n} B_{h,r}(\mathbf{Z}_i) I(\log \tilde{T}_r \geq \log \tilde{T}_j)}, \tag{2.12}$$

where $B_{h,j}(\mathbf{Z})$ is a sequence of weights adding up to 1. Noteworthy, when $B_{h,j}(\mathbf{Z}) = n^{-1}$ for all j, $\tilde{\Lambda}(e_i \mid \mathbf{Z}_i)$ reduces to the classical Nelson-Aalen estimator. More specifically, we adopt the kernel weights

$$B_{h,j}(\mathbf{Z}) = \frac{K_{h,j}(\mathbf{Z})}{\sum_{k=1}^{n} K_{h,k}(\mathbf{Z})},$$

where $K_{h,j}(\mathbf{Z}) := K_1\left(h^{-1}(\mathbf{Z} - \mathbf{Z}_j)\right)$ is a density kernel function with bandwidth h that may depend on n. When there is only one continuous covariate, we may choose the biquadratic kernel $K_1(x) = (15/16)(1-x^2)^2 I(|x| \leq 1)$. When there are multiple continuous covariates in $\mathbf{Z}$, one may adopt a product kernel with a higher order kernel for each covariate as discussed in Leng and Tong (2014). For instance, when there are two continuous covariates, we use the kernel $K_1(x) = (15/32)(3 - 10x^2 + 7x^4)I(|x| \leq 1)$. Since these higher order kernels can give $\tilde{\Lambda}(e_i \mid \mathbf{Z}_i) < 0$, we set $\tilde{\Lambda}(e_i \mid \mathbf{Z}_i) = 0$ whenever necessary. The resulting estimating equation with weight function $\phi(\mathbf{Z})$ then becomes

$$S_n^{\phi}(\mathbf{b}) = \frac{1}{n}\sum_{i=1}^{n} \phi(\mathbf{Z}_i)\Bigg(\Delta_i \left\{ I\left(e_i(\mathbf{b}) \leq 0\right) - \tau \right\} + (1-\Delta_i)(1-\tau)\left[1 - \exp\left\{ H_\tau\left(\tilde{\Lambda}(e_i \mid \mathbf{Z}_i)\right)\right\}\right] \Bigg) = 0. \tag{2.13}$$

Likewise, the above root-solving procedure can be translated into minimizing the convex loss function

$$L^{\phi}(\mathbf{b}, \tilde{\Lambda}) = \frac{1}{n}\sum_{i=1}^{n}\phi(\mathbf{Z}_i)\Bigg\{\Delta_i\rho_\tau\left(e_i(\mathbf{b})\right) + (1-\Delta_i)(1-\tau)\rho_\tau\left(\tilde{T}^* - \left[\frac{\exp\left\{H_\tau\left(\tilde{\Lambda}(e_i \mid \mathbf{Z}_i)\right)\right\}-1}{\tau}\mathbf{b}^\top\mathbf{Z}_i\right]\right)\Bigg\} \quad (2.14)$$

with respect to $\mathbf{b}$ but with $\Lambda_0(e_i \mid \mathbf{Z}_i)$ replaced by $\tilde{\Lambda}(e_i \mid \mathbf{Z}_i)$ and $\tilde{T}^* > \max_i\{\tau^{-1}(\exp[H_\tau\{\tilde{\Lambda}(e_i \mid \mathbf{Z}_i)\}] - 1)\mathbf{b}^\top\mathbf{Z}_i\}$. We take $\tilde{T}^* = \max_{i\in\{1,\ldots,n\}}\{\log(\tilde{T}_i)\} + 100$ for simulation studies and data analysis. The computation of $\tilde{\boldsymbol{\beta}}^{(\phi(\mathbf{Z}))}$, the minimizer of (2.14), is simple to implement with currently available software. After we have estimated the conditional cumulative hazard function $\Lambda_0(e_i \mid \mathbf{Z}_i)$, we consider the augmented data set $\left\{\log\tilde{T}_i, \mathbf{Z}_i\right\}_{i=1,\ldots,n}$ and $\left\{\tilde{T}^*, \tau^{-1}(\exp[H_\tau\{\tilde{\Lambda}(e_i \mid \mathbf{Z}_i)\}] - 1)\mathbf{Z}_i\right\}_{i=1}^{n}$. Then $\tilde{\boldsymbol{\beta}}^{(\phi(\mathbf{Z}))}$ is computed using the function `rq` in `R` package `quantreg` by regressing the augmented data set with weights $\left\{\phi(\mathbf{Z}_i)\Delta_i\right\}_{i=1,\ldots,n}$ for $\left\{\log\tilde{T}_i, \mathbf{Z}_i\right\}_{i=1,\ldots,n}$ and $\left\{\phi(\mathbf{Z}_i)(1-\Delta_i)(1-\tau)\right\}_{i=1,\ldots,n}$ for $\left\{\tilde{T}^*, \tau^{-1}\left[\exp\left\{H_\tau\left(\tilde{\Lambda}(e_i \mid \mathbf{Z}_i)\right)\right\} - 1\right]\mathbf{Z}_i\right\}_{i=1,\ldots,n}$. The extra effort needed to implement our approach is minimal.

Estimation of efficient weight $\phi^{opt}(\mathbf{Z})$ Similar to the previous concern about the unknown true conditional cumulative hazard function $\Lambda_0(\cdot \mid \mathbf{Z})$, the optimal weight function $\phi^{opt}(\mathbf{Z})$ also involves unknown quantities which need to be estimated. We let $1 - H(u \mid \mathbf{Z}) = \{1 - F_0(u \mid \mathbf{Z})\}\{1 - G_0(u \mid \mathbf{Z})\}$ so that $\phi^{opt}(\mathbf{Z}) = \lambda_0(0 \mid \mathbf{Z})[\{1 - H(0 \mid \mathbf{Z})\}P(0, \mathbf{Z})]^{-1}$. We propose to estimate $\lambda_0(0 \mid \mathbf{Z}), \{1 - H(0 \mid \mathbf{Z})\}$ and $P(0, \mathbf{Z})$ separately by kernel estimation. Note that an initial estimator of $\boldsymbol{\beta}_0$ is needed for the estimation of these three terms. Hence we first obtain $\tilde{\boldsymbol{\beta}}^{(1)}$ by solving $L(\mathbf{b}, \tilde{\Lambda})$ in (2.14) with $\phi\left(\mathbf{Z}_i\right) \equiv 1$ upon which we estimate the aforementioned three quantities. The estimate $\hat{\phi}^{opt}(\mathbf{Z}, \tilde{\boldsymbol{\beta}}^{(1)})$ for $\phi^{opt}(\mathbf{Z})$ is obtained by carefully combining these esimates. Let $\epsilon_* = \exp(\epsilon)$ and let $\lambda_*(\cdot \mid \mathbf{Z})$ be the hazard function of ϵ_* given $\mathbf{Z}$. Denote $F_*(\cdot \mid \mathbf{Z})$ and $G_*(\cdot \mid \mathbf{Z})$ as the cumulative distribution functions of ϵ_*

and $\exp\{\log C - \boldsymbol{\beta}_0^\top \mathbf{Z}\}$ given $\mathbf{Z}$ respectively. Let

$$B_{d,j}(\mathbf{Z}) = \frac{K_{d,j}(\mathbf{Z})}{\sum_{k=1}^n K_{d,k}(\mathbf{Z})},$$

where $K_{d,j}(\mathbf{Z}) := K_2\left\{d^{-1}(\mathbf{Z} - \mathbf{Z}_j)\right\}$ is a density kernel function with bandwidth d that may depend on n. Denote $K_b(s) = b^{-1}K_3(s/b)$, where $K_3(s/b)$ is a kernel function with support $[-1, 1]$ and bandwidth b that may depend on n. Since $\lambda_0(t \mid \mathbf{Z}) = \exp(t)\lambda_*\{\exp(t) \mid \mathbf{Z}\}$, we estimate $\lambda_0(0 \mid \mathbf{Z}_i) = \lambda_*(1 \mid \mathbf{Z}_i)$ by $\hat{\lambda}_*(1, \tilde{\boldsymbol{\beta}}^{(1)} \mid \mathbf{Z}_i)$, where

$$\hat{\lambda}_*(1, \tilde{\boldsymbol{\beta}}^{(1)} \mid \mathbf{Z}_i) = \sum_{j=1}^n \frac{B_{d,j}(\mathbf{Z}_i)\Delta_j K_b\left(\exp\{e_j(\tilde{\boldsymbol{\beta}}^{(1)})\} - 1\right)}{\sum_{r=1}^n B_{d,r}(\mathbf{Z}_i) I[\exp\{e_r(\tilde{\boldsymbol{\beta}}^{(1)})\} \geq \exp\{e_j(\tilde{\boldsymbol{\beta}}^{(1)})\}]}.$$

Mimicking the choice of kernel for the estimation of λ in Lin and Chen (2013) when λ is independent of $\mathbf{Z}$, we choose $K_3(\cdot)$ to be a Gaussian kernel function in simulation studies and data analysis.

Next, we estimate $1 - H_*(1 \mid \mathbf{Z}_i) := \{1 - F_*(1 \mid \mathbf{Z}_i)\}\{1 - G_*(1 \mid \mathbf{Z}_i)\}$ by $1 - \hat{H}_*(1, \tilde{\boldsymbol{\beta}}^{(1)} \mid \mathbf{Z}_i) = \sum_{j=1}^n I\left(\exp\{e_j(\tilde{\boldsymbol{\beta}}^{(1)})\} \geq 1\right) B_{dj}(\mathbf{Z}_i)$. On the other hand,

$$P(0, \mathbf{Z}_i) = \int_{-\infty}^0 \frac{d\Lambda_0(u \mid \mathbf{Z}_i)}{1 - H(u \mid \mathbf{Z}_i)} = \int_0^1 \frac{d\Lambda_*(t \mid \mathbf{Z}_i)}{1 - H_*(t \mid \mathbf{Z}_i)} =: P_*(1 \mid \mathbf{Z}_i).$$

Denote

$$\hat{\Lambda}_*(t, \tilde{\boldsymbol{\beta}}^{(1)} \mid \mathbf{Z}_i) = \sum_{j=1}^n \frac{B_{d,j}(\mathbf{Z}_i)\Delta_k I\left(\exp\{e_j(\tilde{\boldsymbol{\beta}}^{(1)})\} \leq t\right)}{\sum_{r=1}^n B_{d,r}(\mathbf{Z}_i) I\left(\exp\{e_r(\tilde{\boldsymbol{\beta}}^{(1)})\} \geq \exp\{e_j(\tilde{\boldsymbol{\beta}}^{(1)})\}\right)},$$

then one may estimate $P_*(1 \mid \mathbf{Z}_i)$ by

$$\hat{P}_*(1, \tilde{\boldsymbol{\beta}}^{(1)} \mid \mathbf{Z}_i) = \int_0^1 \frac{d\hat{\Lambda}_*(t, \tilde{\boldsymbol{\beta}}^{(1)} \mid \mathbf{Z}_i)}{1 - \hat{H}_*(t, \tilde{\boldsymbol{\beta}}^{(1)} \mid \mathbf{Z}_i)} = \sum_{j=1}^n \frac{B_{d,j}(\mathbf{Z}_i)\Delta_k I\left(\exp\{e_j(\tilde{\boldsymbol{\beta}}^{(1)})\} \leq 1\right)}{\left\{\sum_{r=1}^n B_{d,r}(\mathbf{Z}_i) I\left(\exp(e_r(\tilde{\boldsymbol{\beta}}^{(1)})) \geq \exp\{e_j(\tilde{\boldsymbol{\beta}}^{(1)})\}\right)\right\}^2}.$$

Combining the estimators above, we can write the estimate $\phi^{opt}(\mathbf{Z})$ as

$$\tilde{\phi}^{opt}(\mathbf{Z}, \tilde{\boldsymbol{\beta}}^{(1)}) = \frac{\hat{\lambda}_*(1, \tilde{\boldsymbol{\beta}}^{(1)} \mid \mathbf{Z})}{\{1 - \hat{H}_*(1, \tilde{\boldsymbol{\beta}}^{(1)} \mid \mathbf{Z}_i)\}\{\hat{P}_*(1, \tilde{\boldsymbol{\beta}}^{(1)} \mid \mathbf{Z}_i)\}}. \tag{2.15}$$

A more careful observation of our numerical procedure reveals that there are modifications which can further enhance the stability of the estimation without affecting the asymptotic properties of the estimator. Since it is possible to observe $1-\hat{H}_*(1, \tilde{\boldsymbol{\beta}}^{(1)} \mid \mathbf{Z}_i) = 0$ occasionally, we recommend a modified estimator of $1-H_*(1 \mid \mathbf{Z}_i)$ to avoid such situations so as to improve the numerical stability. The correction term is given by

$$1 - \hat{H}^1_*(1, \tilde{\boldsymbol{\beta}}^{(1)} \mid \mathbf{Z}_i) = 1 - \hat{H}_*(1, \tilde{\boldsymbol{\beta}}^{(1)} \mid \mathbf{Z}_i) + B_{di}(\mathbf{Z}_i) I\left(1 - \hat{H}_*(1, \tilde{\boldsymbol{\beta}}^{(1)} \mid \mathbf{Z}_i) < B_{di}(\mathbf{Z}_i)\right).$$

As shown in the proof of (S1) on Page 13 of the supplementary materials, one can see that the supremum of $|B_{hi}(\mathbf{Z}_i)|$ is $O((nd_n^p)^{-1})$, which decays faster than $o_p(n^{-1/4})$, thus the modification in $\hat{H}^1_*(t, \tilde{\boldsymbol{\beta}} \mid \mathbf{Z})$ does not affect the asymptotic variance of the estimator. The other modifications are based on the quantile assumption and thus having no impact to the asymptotic variance naturally, as can be seen from the proof of Theorem 4. Readers may also refer to Remark 1 for further justification of this correction term. Another modification is proposed as a result of an observation that $P_*(1, \mathbf{Z}_i) \geq \Lambda_*(1 \mid \mathbf{Z}_i) = \Lambda(0 \mid \mathbf{Z}_i) = -\log(1-\tau)$. Hence, we may consider $\hat{P}^1_*(1, \tilde{\boldsymbol{\beta}}^{(1)} \mid \mathbf{Z}_i) = \hat{P}_*(1, \tilde{\boldsymbol{\beta}}^{(1)} \mid \mathbf{Z}_i) \vee -\log(1-\tau)$ instead of $\hat{P}_*(1, \tilde{\boldsymbol{\beta}}^{(1)} \mid \mathbf{Z}_i)$. Finally, due to the fact that

$$\{1 - H_*(1 \mid \mathbf{Z})\} \int_0^1 \frac{d\Lambda_*(t \mid \mathbf{Z}_i)}{1 - H_*(t \mid \mathbf{Z}_i)} \leq -\log(1-\tau),$$

we may estimate $\phi^{opt}(\mathbf{Z})$ via

$$\hat{\phi}^{opt}(\mathbf{Z}, \tilde{\boldsymbol{\beta}}^{(1)}) = \frac{\hat{\lambda}_*(1, \tilde{\boldsymbol{\beta}}^{(1)} \mid \mathbf{Z})}{H_\tau\left(\{1 - \hat{H}^1_*(1, \tilde{\boldsymbol{\beta}}^{(1)} \mid \mathbf{Z}_i)\}\{\hat{P}^1_*(1, \tilde{\boldsymbol{\beta}}^{(1)} \mid \mathbf{Z}_i)\}\right)}. \tag{2.16}$$

Plugging (2.16) into the original convex objective function (2.14), we can obtain $\hat{\boldsymbol{\beta}}^{(\hat{\phi}^{opt})}$ by minimizing

$$L_{n,\hat{\phi}^{opt}}(\mathbf{b}, \hat{\Lambda}_{\tilde{\boldsymbol{\beta}}^{(1)}}) = \frac{1}{n}\sum_{i=1}^{n} \hat{\phi}^{opt}(\mathbf{Z}_i, \tilde{\boldsymbol{\beta}}^{(1)}) \Bigg\{ \Delta_i \rho_\tau \left(e_i(\mathbf{b})\right) + (1-\Delta_i)(1-\tau) \times \rho_\tau \left(\tilde{T}^* - \left[\frac{\exp\left\{ H_\tau \left(\hat{\Lambda}_{\tilde{\boldsymbol{\beta}}^{(1)}}(e_i \mid \mathbf{Z}_i) \right) \right\} - 1}{\tau} \mathbf{b}^\top \mathbf{Z} \right] \right) \Bigg\}, \quad (2.17)$$

where

$$\hat{\Lambda}_{\tilde{\boldsymbol{\beta}}^{(1)}}\{e_i \mid \mathbf{Z}_i\} = \sum_{j=1}^{n} \frac{B_{h,j}(\mathbf{Z}_i)\Delta_j I\{e_j(\tilde{\boldsymbol{\beta}}^{(1)}) \le e_i(\tilde{\boldsymbol{\beta}}^{(1)})\}}{\sum_{r=1}^{n} B_{h,r}(\mathbf{Z}_i) I\{e_r(\tilde{\boldsymbol{\beta}}^{(1)}) \ge e_j(\tilde{\boldsymbol{\beta}}^{(1)})\}}.$$

Similar to the computation of $\tilde{\boldsymbol{\beta}}^{(1)}$ discussed in Section 2.1, the estimate $\hat{\boldsymbol{\beta}}^{(\hat{\phi}^{opt})}$ is computed by regressing an augmented data set with weights $\left\{\hat{\phi}^{opt}(\mathbf{Z}_i, \tilde{\boldsymbol{\beta}}^{(1)})\Delta_i\right\}_{i=1,\ldots,n}$ for $\left\{ \log \tilde{T}_i, \mathbf{Z}_i \right\}_{i=1}^{n}$ and $\left\{\hat{\phi}^{opt}(\mathbf{Z}_i, \tilde{\boldsymbol{\beta}}^{(1)})(1-\Delta_i)(1-\tau)\right\}_{i=1,\ldots,n}$ for $\left\{\tilde{T}^*, \tau^{-1}\left[\exp\left\{H_\tau\left(\hat{\Lambda}_{\tilde{\boldsymbol{\beta}}^{(1)}}(e_i \mid \mathbf{Z}_i)\right)\right\} - 1\right]\mathbf{Z}_i\right\}_{i=1,\ldots,n}$. As to be shown in Theorem 4 in Section 3, the proposed estimator $\hat{\boldsymbol{\beta}}^{(\hat{\phi}^{opt})}$ and the ideal estimator $\tilde{\boldsymbol{\beta}}^{(\phi^{opt})}$, which is obtained when $\phi^{opt}(\mathbf{Z})$ is known, are asymptotically equivalent.

Remark 1. *For the modified estimator* $1-\hat{H}^1_*(1, \tilde{\boldsymbol{\beta}}^{(1)} \mid \mathbf{Z}_i)$*, when the value of* $1-\hat{H}_*(1, \tilde{\boldsymbol{\beta}}^{(1)} \mid \mathbf{Z}_i)$ *is smaller than* $B_{hi}(\mathbf{Z}_i)$*, the small value of* $B_{hi}(\mathbf{Z}_i)$ *indicates that there are relatively more data point with similar values of covariates, but yet none, or very few, of them satisfies* $e_i(\tilde{\boldsymbol{\beta}}^{(1)}) > 0$ *in which case the quantity* $1 - H_*(1 \mid \mathbf{Z}_1)$ *should also be small, and vice versa. Therefore the modified estimator can provide additional information on* $1 - H_*(1 \mid \mathbf{Z}_i)$ *while preventing the denominator of the estimate of* $\hat{\phi}^{opt}(\mathbf{Z}, \tilde{\boldsymbol{\beta}}^{(1)})$ *and* $1 - H_*(1 \mid \mathbf{Z})$ *to be zero.*

3 Large Sample Properties

To establish the asymptotic properties of our proposed estimator, we impose the following regularity assumptions:

(C1) T and C are conditionally independent given the covariate $\mathbf{Z}$.

(C2) The true value $\boldsymbol{\beta}_0$ is in the interior of a bounded convex region $\boldsymbol{B}$. The support $\mathcal{Z}$ of $\mathbf{Z}$ is bounded and compact.

(C3) $\inf_{\mathbf{Z}\in\mathcal{Z}} P(\tilde{T} \geq \mathcal{T} \mid \mathbf{Z}) \geq 1 - \eta_0 > 0$, where $\mathcal{T} = \sup_{\mathbf{Z}\in\mathcal{Z}} \sup_{\mathbf{b}\in\boldsymbol{B}} \exp(\mathbf{b}^\top \mathbf{Z})$.

(C4) Denote $q = \max(q_1, q_2)$. The first q partial derivatives with respect to $\mathbf{Z}$ of the density function $f_{\mathbf{Z}}(\mathbf{Z})$ are uniformly bounded for $\mathbf{Z} \in \mathcal{Z}$, and $f_0(t \mid \mathbf{Z})$ and $g_0(t \mid \mathbf{Z})$ are uniformly bounded away from infinity and have bounded(uniformly in t) first q order partial derivatives with respect to $\mathbf{Z}$. Moreover, $\inf_{\mathbf{Z}\in\mathcal{Z}} f_{\mathbf{Z}}(\mathbf{Z}) \geq \delta_0$ for some $\delta_0 > 0$.

(C5) The bandwidths h_n, b_n and d_n satisfy $h_n = O(n^{-v_h})$, $b_n = O(n^{-v_b})$ and $d_n = O(n^{-v_d})$ with $1/2q_1 < v_h < 1/3p$, $v_b > 1/8$, $v_d > (4q_2)^{-1}$ and $pv_d + v_b < 1/2$.

(C6) (i) The kernel functions $K_1(\cdot), K_2(\cdot)$ are Lipschitz-continuous density functions with compact support on R^p.

(ii) The intergral $\int_{R^p} z_1^{i_1} \cdots z_p^{i_p} K_j(\mathbf{z}) d\mathbf{z} = 0$ for non-negative integers $i_1, \ldots i_p$ with $i_1 + \cdots + i_p \leq q_j - 1$, $j = 1, 2$.

(C7) For $\mathbf{b}$ in the neighbourhood of $\boldsymbol{\beta}_0$, the matrix

$$E\left(\phi(\mathbf{Z})\mathbf{Z}\mathbf{Z}^\top f_0\left((\mathbf{b}-\boldsymbol{\beta}_0)^\top \mathbf{Z} \mid \mathbf{Z}\right)\left\{1 - G_0\left((\mathbf{b}-\boldsymbol{\beta}_0)^\top \mathbf{Z} \mid \mathbf{Z}\right)\right\}\right)$$

is positive definite.

(C8) The weight function $\phi(\mathbf{Z})$ is non-negative and bounded above uniformly in $\mathbf{Z}$.

Conditions (C1)-(C4) are standard assumptions imposed in analyzing failure time data. In (C3), we assume that $1 - \sup_{\mathbf{Z}\in\mathcal{Z}} H(0 \mid \mathbf{Z}) \geq 1 - \eta_0$ for some $\eta_0 > 0$. So the denominator should never be zero if H is known. This assumption was adopted in Leng and Tong (2014) and Wang and Wang (2009) to ensure identifiability; see also Liang et al. (2012). Conditions (C5) and (C6) specify the conditions on the bandwidth and kernel function. Condition (C7) ensures that the quantile regression estimator is unique and is used to establish the

asymptotic normality of the estimator. Condition (C8) ensures that the the estimate is still consistent and asymptotic normal after adding weights. A broad class of weights satisfy this condition, in particular the optimal weight $\phi^{opt}(\mathbf{Z})$ proposed in Section 2.

The following two theorems state the consistency and asymptotic normality of our estimator adopting a given $\mathbf{Z}$-measurable weight function $\tilde{\boldsymbol{\beta}}^{(\phi(\mathbf{Z}))}$ which minimizes (2.14):

Theorem 1. *(Consistency) Under conditions (C1)-(C6) and (C8), for any $\phi(\mathbf{Z})$,*

$$\tilde{\boldsymbol{\beta}}^{\phi(\mathbf{Z})} \to \boldsymbol{\beta}_0$$

in probability as $n \to \infty$.

Theorem 2. *(Asymptotic normality) Under conditions (C1)-(C8), for any $\phi(\mathbf{Z})$, we have*

$$n^{1/2}(\tilde{\boldsymbol{\beta}}^{(\phi(\mathbf{Z}))} - \boldsymbol{\beta}_0) \xrightarrow{d} N(0, {\Gamma_1^{(\phi)}}^{-1} V_1^{(\phi)} {\Gamma_1^{(\phi)}}^{-1}),$$

where

$$\begin{aligned}
\Gamma_1^{(\phi)} =& E\left(\phi(\mathbf{Z})\mathbf{Z}\mathbf{Z}^\top f_0(0 \mid \mathbf{Z})\{1 - G_0(0 \mid \mathbf{Z})\}\right), \\
V_1^{(\phi)} =& E\left(\phi(\mathbf{Z})^2 \mathbf{Z}\mathbf{Z}^\top (1-\tau)^2 \{1 - G_0(0 \mid \mathbf{Z})\}^2 \int_{-\infty}^{0} \frac{\lambda_0(u \mid \mathbf{Z}) du}{\{1 - F_0(u \mid \mathbf{Z})\}\{1 - G_0(u \mid \mathbf{Z})\}}\right).
\end{aligned}$$

Remark 2. *Define $\hat{\boldsymbol{\beta}}_{LT}$ and $\hat{\boldsymbol{\beta}}_{BGK}$ as the estimators proposed in Leng and Tong (2014) and De Becker et al. (2019), respectively. With correspondingly suitable regularity conditions hold, one can indeed show that*

$$n^{1/2}(\hat{\boldsymbol{\beta}}_{LT} - \boldsymbol{\beta}_0) \xrightarrow{d} N\left(0, {\Gamma_1^{(1)}}^{-1} V_1^{(1)} {\Gamma_1^{(1)}}^{-1}\right) \quad \text{and} \quad n^{1/2}(\hat{\boldsymbol{\beta}}_{BGK} - \boldsymbol{\beta}_0) \xrightarrow{d} N\left(0, {\Gamma_1^{(1)}}^{-1} V_1^{(1)} {\Gamma_1^{(1)}}^{-1}\right),$$

where $\Gamma_1^{(1)}$ and $V_1^{(1)}$ are defined in Theorem 2 above. The above variance form coincides with that of $n^{1/2}(\tilde{\boldsymbol{\beta}}^{(1)} - \boldsymbol{\beta}_0)$ as a special case justified by Theorem 2. For more details of the

derivations of the asymptotic variances of various estimators, readers are referred to Section E in the Supplementary Materials.

After we have established the consistency and asymptotic normality of our proposed estimator, we then present the following the following theorem which justifies that the optimal weight defined in (2.11) can guarantee the minimum variance attained by our estimator.

Theorem 3. *(Most efficient* $\mathbf{Z}$*-weight) Under conditions (C1)-(C8), we have*

$$\Gamma_1^{(\phi^{opt})^{-1}} V_1^{(\phi^{opt})} \Gamma_1^{(\phi^{opt})^{-1}} \leq \Gamma_1^{(\phi)^{-1}} V_1^{(\phi)} \Gamma_1^{(\phi)^{-1}},$$

for any $\phi(\mathbf{Z})$*, where* $A \leq B$ *means* $B - A$ *is non-negative definite for* A *and* B *two arbitrary square matrices of same dimension.*

While Theorem 3 ensures the optimality of the weight (2.11) when its true value is known, one can also show that our proposed kernel-based estimated optimal weight given by (2.16) is practically feasible in the sense that the resulting estimator can also achieve the same asymptotic minimum variance.

Theorem 4. *(Feasibility to attain minimum variance) Under conditions (C1)-(C8), we have* $\hat{\boldsymbol{\beta}}^{(\hat{\phi}^{opt})} \xrightarrow{p} \boldsymbol{\beta}_0$*, and*

$$n^{1/2}(\hat{\boldsymbol{\beta}}^{(\hat{\phi}^{opt})} - \boldsymbol{\beta}_0) \xrightarrow{d} N(0, \Gamma_1^{(\phi^{opt})^{-1}} V_1^{(\phi^{opt})} \Gamma_1^{(\phi^{opt})^{-1}}).$$

The matrices $\Gamma_1^{(\phi)}$ and $V_1^{(\phi)}$ involve unknown conditional density functions $f_0(\cdot \mid \mathbf{Z})$ and $g(\cdot \mid \mathbf{Z})$ that are difficult to estimate in finite samples. Therefore, we adopt a bootstrap resampling approach by resampling the triples $(\tilde{T}, \Delta, \mathbf{Z})$ with replacement. The performance of the bootstrap approach is shown to be satisfactory in the Monte Carlo studies conducted in Section 4. The same approach is adopted for inference of $\hat{\boldsymbol{\beta}}^{(\hat{\phi}^{opt})}$.

In real practice, we can choose the bandwidths h_n, d_n and b_n using the K-fold cross validation. The set of h_n, d_n and b_n that yields the smallest averaged prediction error is

selected. Moreover, from condition (C5), if we choose $v_b = n^{1/6}$, then we have $(4q_2)^{-1} < v_d < (3p)^{-1}$, which suggests that when there are two or more continuous covariates, we may take $h_n = d_n$ to simplify the procedure. Regarding the choice of loss function for the cross validation, Wang and Wang (2009) proposed to use the check function for uncensored data. Alternatively we could use the loss function

$$l(\mathbf{b}) = \left| \frac{1}{n_{CV}} \sum_{i=1}^{n_{CV}} \left(\Delta_i \left\{ I\left(e_i(\mathbf{b}) \leq 0\right) - \tau \right\} + (1-\Delta_i)(1-\tau) \left[1 - \exp\left\{ H_\tau \left(\hat{\Lambda}\left(e_i(b)\right) \right) \right\} \right] \right) \right|, \tag{3.1}$$

where n_{CV} is the sample size in the jth partition of the data by a little abuse of notation and $\hat{\Lambda}\left(e_i(b)\right)$ is the Nelson-Aalen estimate for the cumulative hazard function of $\epsilon_i(b)$. These two loss functions give similar results when censoring rate is not exceedingly high. Since the loss function $l(\mathbf{b})$ takes censored data into account, it can produce better result when the censoring rate increases. It should be noted that no matter which loss function is adopted, while checking for more bandwidths could help improve the numerical performance, there could be multiple bandwidths that achieve the smallest prediction error occasionally. For such situations, we suggest choosing a larger bandwidth as it usually results in a more stable set of estimates based on our practical experiences.

Remark 3. *Concerning $B_{di}(\mathbf{Z})$, the choice of kernel $K_2(\cdot)$ is different from the kernel $K_1(\cdot)$ for $B_{hi}(\mathbf{Z})$ when dimension increases. As shown in Theorem 4, let $d_n = c_d n^{-v_d}, b_n = c_b n^{-v_b}$, it suffices to have $(4q_2)^{-1} < v_d < (3p)^{-1}$ if we choose $v_b = 1/6$, which is adopted in simulation studies and data analysis. Compared to the constraints for the order of $K_1(\cdot)$, which requires $(2q_1)^{-1} < v < (3p)^{-1}$, we can increase the order of the kernel $K_2(\cdot)$ at a slower rate. In particular, when there are two continuous covariates, we still choose $K_2(\cdot)$ to be the biquadratic kernel by taking $v_d = 1/7$. However, for $K_1(\cdot)$, we are forced to use higher order kernels as $1/(2q_1) = 1/4$ when $q_1 = 2$, which is greater than $(3p)^{-1}$. Similar to the estimation of $\Lambda(\cdot \mid \mathbf{Z})$, we need to adjust the value for $\hat{\lambda}_*(1, \tilde{\boldsymbol{\beta}}^{(1)} \mid \mathbf{Z})$ when we turn to higher order kernels as it is possible to obtain negative values. We simply set $\hat{\lambda}_*(1, \tilde{\boldsymbol{\beta}}^{(1)} \mid \mathbf{Z}) = 0$ as*

needed.

4 Simulations

In this section, we assess the finite sample performance of the proposed methods via Monte Carlo simulations. We compare our estimator $\tilde{\boldsymbol{\beta}}^{(1)}$ presented in Section 2, which adopts a weight that involves cumulative hazard function only (Vanilla), with Leng and Tong (2014)'s (LT) procedure so as to examine the finite sample performances of these two methods. We further compare these results with the estimator $\hat{\boldsymbol{\beta}}^{(\hat{\phi}^{opt})}$ proposed in Section 2.3 to demonstrate the improvement by introducing the optimal $\mathbf{Z}$-measurable weight (Proposed opt) as well as Wang and Wang (2009)'s (WW) procedure. For each demonstration, we report the biases and root mean square errors (RMSE) of individual procedures based on 500 simulations. Standard errors of the biases and RMSE are also computed by repeating the simulations for 300 times. We also report the average coverage probabilities and average interval lengths of related resampling-based 95% confidence intervals. 300 bootstrap samples are simulated to obtain the confidence intervals in each simulation run. Throughout all examples, the computation time of Proposed opt is on average three times of that required for the vanilla estimator. In each example, we consider different parameters that give 20%, 40% and 60% censoring at median, respectively. The censoring rates at different quantile levels are quite close to the rate at $\tau = 0.5$ with difference not greater that 5%. All of our settings can satisfy the conditions in Section 3, in particular the identifiability condition (C3). As ϵ follows normal distribution in our settings, which is unbounded, we only need to compare the distribution of the censoring time C and the maximum value of $\exp(\mathbf{Z}^\top \boldsymbol{\beta}_0)$ with respect to $\mathbf{Z}$ in order to check the condition.

Example 1 We generate data from the model

$$\log(T) = \beta_0 + \beta_1 Z + \frac{\epsilon}{Z^2},$$

where $\boldsymbol{\beta} = (\beta_0, \beta_1) = (3, 5)$, $Z \sim \text{Uniform}(1, 2)$ and $\epsilon = \eta - Q_\tau(\eta)$ with η follows the standard normal distribution. We report the parameter estimates at $\tau = 0.25, 0.50$ and 0.75. The censoring time $\log(C)$ follows the uniform distribution $\text{Uniform}(0, \theta)$. We take $\theta = 52, 26$ and 17 to produce 20%, 40% and 60% censoring rates at $\tau = 0.50$, respectively.

Table 1 summarize the simulation results for three different τ values, namely 0.25, 0.50 and 0.75 with two different sample sizes including $n = 300$ and 500 at 60% censoring rate at median. The corresponding results for cases with censoring rates 20% and 40% are relegated to Tables 1 – 2 in Section G of the Supplementary Materials. For simplicity, we take the bandwidths as $h_n = n^{-1/3+0.01}$, $b_n = n^{-1/6}$ and $d_n = n^{-1/6}$. The bandwidth $h_n = n^{-1/3+0.01}$ was adopted in Leng and Tong (2014) as the optimal rate based of the asymptotic result. Our choice of b_n and d_n are motivated from the fact that the optimal choice for the estimation of $\lambda(\cdot \mid \mathbf{Z})$ is $b_n = d_n = O(n^{-1/6})$; see Remark 3.1 in Van Keilegom and Veraverbeke (2001). In this example, our vanilla procedure gives smaller RMSE in nearly all settings as compared to Leng and Tong (2014), especially when $\tau = 0.75$, showing that the proposed methods bring in improvement in the finite sample performance as compared with Leng and Tong (2014), which estimated the distribution function in the martingale-based formulation instead. As mentioned, since Wang and Wang (2009)'s procedure can be regarded as a special case in our procedure with the weight $\exp\{\Lambda_0(t \mid \mathbf{Z})\}$, the two methods, namely Vanilla and WW, give very similar numerical results. It is also noteworthy that the procedure with the optimal weight estimated gives the smallest RMSE and empirical mean lengths (EML) of the bootstrap confidence intervals in all settings, showing that it is more efficient compared to other existing methods. The empirical coverage probabilities (ECP) of the bootstrap confidence intervals for all methods are close to the nominal level of 95%.

Table 1: Simulation results for Example 1 when there is 60% censoring at median. The ECP and EML are the empirical coverage probabilities and empirical mean lengths for different confidence interval procedures with a nominal level of 0.95. Standard errors are in parenthesis.

			Bias		RMSE		ECP		EML	
τ	n	Method	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_0$	$\hat{\beta}_1$
0.25	300	LT	0.009	0.012	0.382	0.233	0.936	0.938	1.521	0.929
			(0.017)	(0.010)	(0.011)	(0.007)				
		WW	0.022	-0.018	0.396	0.241	0.932	0.936	1.557	0.950
			(0.017)	(0.010)	(0.012)	(0.007)				
		Proposed	0.010	-0.002	0.388	0.235	0.944	0.946	1.560	0.952
			(0.017)	(0.010)	(0.011)	(0.007)				
		Proposed opt	0.028	-0.014	0.381	0.230	0.938	0.942	1.541	0.937
			(0.017)	(0.010)	(0.011)	(0.007)				
	500	LT	0.003	0.014	0.283	0.172	0.948	0.950	1.154	0.703
			(0.014)	(0.008)	(0.009)	(0.006)				
		WW	0.011	-0.009	0.292	0.178	0.938	0.942	1.190	0.726
			(0.014)	(0.008)	(0.009)	(0.006)				
		Proposed	0.006	0.000	0.289	0.176	0.950	0.952	1.182	0.721
			(0.014)	(0.008)	(0.009)	(0.006)				
		Proposed opt	0.023	-0.012	0.278	0.168	0.952	0.952	1.152	0.700
			(0.013)	(0.008)	(0.009)	(0.005)				
0.5	300	LT	-0.065	0.073	0.342	0.219	0.942	0.934	1.427	0.873
			(0.016)	(0.010)	(0.012)	(0.008)				
		WW	-0.001	0.003	0.330	0.202	0.950	0.956	1.413	0.862
			(0.016)	(0.010)	(0.012)	(0.007)				
		Proposed	-0.007	0.008	0.329	0.202	0.954	0.958	1.422	0.868
			(0.016)	(0.010)	(0.012)	(0.007)				
		Proposed opt	0.009	-0.004	0.322	0.197	0.950	0.948	1.386	0.843
			(0.015)	(0.009)	(0.012)	(0.007)				
	500	LT	-0.074	0.070	0.294	0.187	0.916	0.904	1.096	0.670
			(0.013)	(0.008)	(0.009)	(0.006)				
		WW	-0.010	0.007	0.277	0.168	0.934	0.942	1.082	0.659
			(0.012)	(0.007)	(0.008)	(0.005)				
		Proposed	-0.017	0.013	0.280	0.170	0.926	0.932	1.088	0.663
			(0.012)	(0.007)	(0.009)	(0.005)				
		Proposed opt	-0.004	0.003	0.264	0.159	0.946	0.948	1.056	0.640
			(0.012)	(0.007)	(0.008)	(0.005)				
0.75	300	LT	-0.362	0.284	0.576	0.404	0.928	0.918	2.054	1.339
			(0.021)	(0.013)	(0.020)	(0.014)				
		WW	-0.012	0.010	0.355	0.218	0.958	0.958	1.559	0.951
			(0.017)	(0.010)	(0.013)	(0.008)				
		Proposed	-0.021	0.015	0.361	0.222	0.956	0.956	1.565	0.955
			(0.017)	(0.010)	(0.013)	(0.008)				
		Proposed opt	-0.002	0.000	0.345	0.210	0.954	0.952	1.524	0.926
			(0.016)	(0.010)	(0.012)	(0.007)				
	500	LT	-0.267	0.204	0.437	0.297	0.870	0.860	1.372	0.864
			(0.014)	(0.009)	(0.013)	(0.008)				
		WW	-0.012	0.009	0.297	0.180	0.924	0.934	1.177	0.719
			(0.013)	(0.008)	(0.010)	(0.006)				
		Proposed	-0.019	0.013	0.298	0.181	0.928	0.934	1.181	0.721
			(0.013)	(0.008)	(0.010)	(0.006)				
		Proposed opt	0.000	0.000	0.281	0.170	0.940	0.942	1.144	0.695
			(0.012)	(0.007)	(0.009)	(0.006)				

Example 2 In this example, the data are generated from similar setting in the previous example, but the number of covariates is increased. We generate data from the model

$$\log(T) = \beta_0 + \beta_1 Z_1 + \beta_2 Z_2 + \frac{\epsilon}{Z_1^2},$$

where $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2) = (3, 5, 1)$, $Z_j \sim \text{Uniform}(1, 2)$ for $j = 1, 2$ and $\epsilon = \eta - Q_\tau(\eta)$ with η follows the standard normal distribution. The censoring time was generated as $\log(C) \sim (5 - Z_1)\text{Uniform}(0, \theta)$. We take $\theta = 18, 9$ and 6 to produce 20%, 40% and 60% censoring rates at $\tau = 0.5$, respectively. Similar to Example 1, the results for the high censoring case 60% are summarized in Table 2; corresponding statistics for settings with 20% and 40% censoring rates can be found in Tables 3 – 4 in Section G of the Supplementary Materials.

In view of condition $(C5)$, to adjust for the increased number of covariates, we take the bandwidths as $h_n = n^{-1/7}$, $b_n = n^{-1/6}$ and $d_n = n^{-1/7}$. Similar to the previous examples, in this numerical demonstration, our procedure with the optimal weight gives the smallest RMSE as well as the smallest EML of the bootstrap confidence intervals in nearly all settings. It verifies that Proposed opt is more efficient compared to other methods when there are more than one covariates. The ECP of the bootstrap confidence intervals for all methods are close to the nominal level 95% whereas the corresponding values obtained from LT method demonstrate rather unsatisfactory ECP due to substantial biases incurred for high censoring cases.

Our numerical experience also finds that specific choices of the bandwidth parameters do not lead to substantially different parameter estimates. To this end, we alter one bandwidth from 0.3 to 0.7 while the remaining two other are held fixed. Figure 1 shows the corresponding RMSE given by our optimal procedure Proposed opt when $\tau = 0.5$, $n = 300$ and 40% censoring. The left, middle and the right panels correspond to the settings where b, d and h changes while other bandwidths are fixed at the values used in simulation. The black,

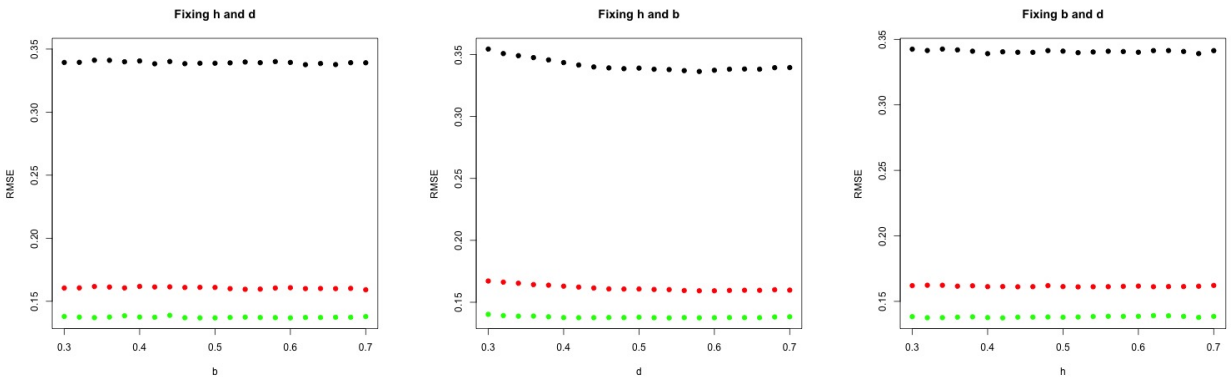

Figure 1: RMSE versus different bandwidths. The black, red and green dots represent the MSE for β_0, β_1 and β_2, respectively.

red and green dots represent the RMSE for β_0, β_1 and β_2, respectively. It shows that the performance of Proposed opt is stable with respect to changes in bandwidths.

Example 3 In this example, we generate data from the model

$$\log(T_i) = \beta_0 + \beta_1 Z_i + \beta_2 Z_i^* + 0.5 Z_i^{-2} \epsilon_i,$$

where $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2) = (3, 2, -1)$, $Z_i \sim$ Uniform$(1, 2)$, $Z_i^* = \log(Z_i) - \sum_{j=1}^{500} \log(Z_j)/500$ and $\epsilon = \eta - Q_\tau(\eta)$ with η follows the standard normal distribution. We report the parameter estimates at $\tau = 0.25, 0.50$ and 0.75. The censoring time was generated as $\log(C) \sim \theta \log((5 - Z_i)$Uniform$(0, 10))$. We take $\theta = 3$, $\theta = 2.3$ and $\theta = 2$ to produce 20%, 40% and 60% censoring rates at $\tau = 0.50$, respectively. The choice of bandwidths is the same as that of Example 2. This setting is more complicated than Example 2 in the sense that the covariates Z and Z^* are dependent. One can see from Table 3 and Tables 5 – 6 in Section G of the Supplementary Materials that Proposed opt is more efficient than other methods in general, and the difference is more significant when there is 60% censoring for $\tau = 0.75$. Similar improvements can also be found for estimates for high quantile levels under more complicated settings with high censoring rates. Figure 2 shows the corresponding RMSE using the same setup of Figure 1 in this example when $\tau = 0.5$, $n = 300$ and 60% censoring.

Table 2: Simulation results for Example 2 when there is 60% censoring at median. The ECP and EML are the empirical coverage probabilities and empirical mean lengths for different confidence interval procedures with a nominal level of 0.95. Standard errors are in parenthesis.

			Bias			RMSE			ECP			EML		
τ	n	Method	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$
0.25	300	LT	0.011	0.012	0.013	0.482	0.224	0.203	0.940	0.946	0.946	1.965	0.930	0.852
			(0.021)	(0.010)	(0.008)	(0.016)	(0.007)	(0.007)						
		WW	0.065	-0.046	-0.004	0.489	0.236	0.208	0.948	0.952	0.950	2.054	0.976	0.897
			(0.023)	(0.011)	(0.009)	(0.016)	(0.008)	(0.007)						
		Proposed	-0.007	0.005	0.009	0.490	0.229	0.208	0.944	0.946	0.950	2.036	0.964	0.882
			(0.022)	(0.010)	(0.009)	(0.016)	(0.008)	(0.007)						
		Proposed opt	0.024	-0.012	0.006	0.487	0.221	0.208	0.958	0.958	0.966	2.105	0.993	0.902
			(0.022)	(0.010)	(0.009)	(0.016)	(0.008)	(0.007)						
	500	LT	-0.014	0.030	0.004	0.371	0.176	0.160	0.952	0.954	0.946	1.484	0.698	0.640
			(0.016)	(0.008)	(0.007)	(0.011)	(0.005)	(0.005)						
		WW	0.035	-0.024	-0.007	0.373	0.172	0.162	0.958	0.956	0.964	1.539	0.724	0.669
			(0.016)	(0.008)	(0.007)	(0.011)	(0.005)	(0.005)						
		Proposed	-0.018	0.017	0.001	0.365	0.171	0.158	0.956	0.952	0.964	1.530	0.719	0.662
			(0.016)	(0.008)	(0.007)	(0.011)	(0.005)	(0.005)						
		Proposed opt	0.013	-0.001	-0.001	0.361	0.170	0.149	0.962	0.952	0.972	1.554	0.726	0.669
			(0.015)	(0.008)	(0.007)	(0.011)	(0.005)	(0.005)						
0.5	300	LT	-0.190	0.159	0.024	0.506	0.289	0.199	0.930	0.896	0.954	1.913	0.924	0.826
			(0.021)	(0.011)	(0.009)	(0.017)	(0.009)	(0.006)						
		WW	-0.002	0.004	-0.001	0.432	0.210	0.190	0.960	0.948	0.954	1.829	0.859	0.790
			(0.019)	(0.009)	(0.008)	(0.014)	(0.007)	(0.006)						
		Proposed	-0.038	0.025	0.004	0.431	0.211	0.189	0.954	0.952	0.942	1.846	0.867	0.798
			(0.019)	(0.009)	(0.008)	(0.015)	(0.007)	(0.006)						
		Proposed opt	0.000	0.005	-0.003	0.423	0.206	0.180	0.964	0.956	0.954	1.865	0.874	0.800
			(0.019)	(0.009)	(0.008)	(0.014)	(0.006)	(0.005)						
	500	LT	-0.205	0.159	0.024	0.424	0.231	0.154	0.902	0.862	0.950	1.454	0.686	0.621
			(0.015)	(0.008)	(0.007)	(0.012)	(0.006)	(0.004)						
		WW	-0.008	0.001	0.005	0.345	0.158	0.141	0.950	0.952	0.954	1.405	0.652	0.598
			(0.015)	(0.007)	(0.007)	(0.011)	(0.005)	(0.004)						
		Proposed	-0.040	0.019	0.012	0.350	0.161	0.143	0.950	0.958	0.952	1.417	0.658	0.605
			(0.015)	(0.007)	(0.007)	(0.011)	(0.005)	(0.004)						
		Proposed opt	-0.003	-0.003	0.007	0.332	0.156	0.132	0.954	0.968	0.964	1.420	0.660	0.597
			(0.014)	(0.007)	(0.006)	(0.010)	(0.005)	(0.004)						
0.75	300	LT	-1.959	1.398	0.146	2.753	2.145	0.445	0.974	0.978	0.990	37.140	31.840	4.642
			(0.113)	(0.093)	(0.020)	(1.164)	(1.051)	(0.074)						
		WW	-0.053	0.026	0.011	0.488	0.240	0.210	0.952	0.936	0.944	2.013	0.960	0.862
			(0.021)	(0.010)	(0.009)	(0.016)	(0.008)	(0.007)						
		Proposed	-0.062	0.030	0.011	0.496	0.244	0.212	0.954	0.934	0.944	2.027	0.967	0.870
			(0.022)	(0.010)	(0.009)	(0.016)	(0.008)	(0.007)						
		Proposed opt	-0.043	0.015	0.008	0.484	0.233	0.200	0.958	0.960	0.960	2.025	0.971	0.870
			(0.020)	(0.010)	(0.009)	(0.016)	(0.008)	(0.006)						
	500	LT	-1.195	0.796	0.100	1.378	0.886	0.265	0.708	0.656	0.952	3.085	1.858	1.103
			(0.028)	(0.017)	(0.011)	(0.029)	(0.019)	(0.009)						
		WW	-0.063	0.032	0.012	0.377	0.177	0.145	0.942	0.952	0.964	1.532	0.721	0.651
			(0.017)	(0.008)	(0.007)	(0.012)	(0.006)	(0.005)						
		Proposed	-0.075	0.037	0.013	0.378	0.178	0.146	0.948	0.958	0.970	1.541	0.727	0.655
			(0.017)	(0.008)	(0.007)	(0.012)	(0.006)	(0.005)						
		Proposed opt	-0.050	0.020	0.010	0.363	0.166	0.141	0.956	0.972	0.974	1.531	0.719	0.649
			(0.016)	(0.007)	(0.007)	(0.011)	(0.005)	(0.005)						

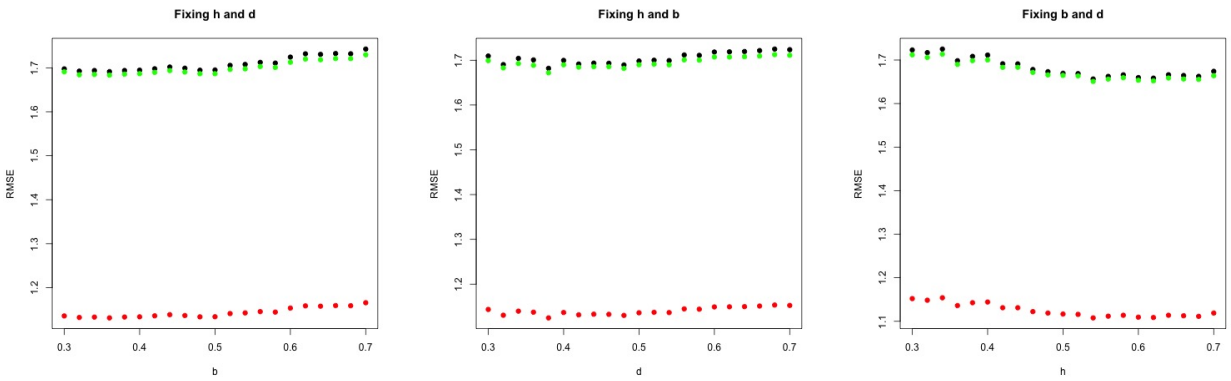

Figure 2: RMSE versus different bandwidths in Example 3. The black, red and green dots represent the MSE for β_0, β_1 and β_2, respectively.

And it is consistent with our experience the choices of the bandwidth parameters do not lead to substantially different parameter estimates.

Remark 4. *In all Examples 1 to 3, the numerical results suggest that our method performs best for higher quantile level estimates especially when the censoring rate is high. This empirical observation can possibly be relevant to the properties of the estimation methods. The method Proposed opt involves kernel estimation in each datum, so some efficiency gain brought by the optimal weight may be offset by the instability of kernel estimation, making our efficiency gain less significant in low quantile and low censoring case compared to other methods, as the involvement of kernel estimation in these methods are relatively small in such a setting. That is because both methods would not use the kernel estimate for data whose estimated conditional cumulative distribution function is less than τ, and kernel estimation is not even involved for uncensored data in Wang and Wang (2009).*

5 Real Data Analysis

We illustrate the proposed method by analyzing the data on the survival time of gastric adenocarcinoma patients who underwent surgery at the Helsinki University Hospital, Finland. The dataset is available at https://datadryad.org/stash/dataset/doi:10.5061/dryad.hb62394.

Table 3: Simulation results for Example 3 when there is 60% censoring at median. The ECP and EML are the empirical coverage probabilities and empirical mean lengths for different confidence interval procedures with a nominal level of 0.95. Standard errors are in parenthesis.

			Bias			RMSE			ECP			EML		
τ	n	Method	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$
0.25	300	LT	-0.508	0.353	-0.407	2.742	1.839	2.647	0.956	0.958	0.958	10.906	7.315	10.553
			(0.107)	(0.072)	(0.105)	(0.090)	(0.061)	(0.084)						
		WW	0.486	-0.327	0.438	2.032	1.359	2.028	0.954	0.954	0.954	8.331	5.576	8.292
			(0.081)	(0.054)	(0.082)	(0.062)	(0.042)	(0.063)						
		Proposed	-0.433	0.296	-0.375	2.506	1.678	2.437	0.960	0.962	0.964	10.181	6.824	9.911
			(0.101)	(0.068)	(0.098)	(0.083)	(0.055)	(0.079)						
		Proposed opt	0.027	-0.016	0.014	2.029	1.357	2.006	0.954	0.952	0.948	12.732	8.544	12.132
			(0.084)	(0.056)	(0.084)	(0.060)	(0.040)	(0.060)						
	500	LT	-0.445	0.307	-0.360	1.847	1.239	1.791	0.958	0.962	0.960	7.743	5.190	7.584
			(0.084)	(0.056)	(0.082)	(0.068)	(0.046)	(0.064)						
		WW	0.375	-0.255	0.328	1.435	0.961	1.426	0.964	0.964	0.964	6.205	4.153	6.207
			(0.066)	(0.044)	(0.066)	(0.046)	(0.031)	(0.046)						
		Proposed	-0.474	0.322	-0.420	1.748	1.172	1.702	0.972	0.972	0.966	7.351	4.925	7.223
			(0.077)	(0.052)	(0.076)	(0.063)	(0.042)	(0.060)						
		Proposed opt	-0.046	0.032	-0.048	1.398	0.937	1.394	0.970	0.968	0.964	6.354	4.254	6.322
			(0.066)	(0.044)	(0.066)	(0.049)	(0.033)	(0.049)						
0.5	300	LT	-1.791	1.224	-1.446	4.510	3.043	4.078	0.984	0.984	0.980	28.153	19.142	25.058
			(0.167)	(0.112)	(0.156)	(0.299)	(0.203)	(0.249)						
		WW	0.114	-0.074	0.126	1.828	1.222	1.805	0.972	0.972	0.970	8.503	5.697	8.350
			(0.089)	(0.060)	(0.089)	(0.067)	(0.045)	(0.065)						
		Proposed	-0.574	0.390	-0.489	2.363	1.585	2.267	0.962	0.964	0.962	9.679	6.492	9.387
			(0.103)	(0.069)	(0.100)	(0.086)	(0.058)	(0.081)						
		Proposed opt	0.111	-0.073	0.107	1.694	1.133	1.685	0.968	0.966	0.962	11.435	7.699	10.898
			(0.082)	(0.055)	(0.082)	(0.076)	(0.051)	(0.070)						
	500	LT	-0.958	0.658	-0.718	2.426	1.633	2.262	0.966	0.964	0.964	10.000	6.724	9.495
			(0.099)	(0.067)	(0.095)	(0.117)	(0.079)	(0.104)						
		WW	0.094	-0.062	0.105	1.473	0.986	1.471	0.944	0.944	0.944	6.097	4.082	6.043
			(0.060)	(0.040)	(0.060)	(0.048)	(0.032)	(0.047)						
		Proposed	-0.418	0.284	-0.353	1.672	1.120	1.637	0.952	0.952	0.954	6.885	4.614	6.739
			(0.069)	(0.046)	(0.067)	(0.060)	(0.041)	(0.058)						
		Proposed opt	0.032	-0.021	0.037	1.359	0.909	1.362	0.956	0.958	0.954	5.921	3.965	5.868
			(0.059)	(0.039)	(0.059)	(0.042)	(0.028)	(0.042)						
0.75	300	LT	-217.940	150.184	-182.336	340.413	234.547	285.232	0.906	0.904	0.918	1095.008	754.463	923.124
			(12.289)	(8.458)	(10.300)	(12.821)	(8.801)	(10.809)						
		WW	-0.374	0.256	-0.311	2.651	1.777	2.555	0.960	0.962	0.962	10.521	7.060	10.196
			(0.103)	(0.069)	(0.101)	(0.099)	(0.067)	(0.089)						
		Proposed	-0.703	0.476	-0.605	2.803	1.878	2.689	0.972	0.974	0.970	10.860	7.286	10.520
			(0.113)	(0.076)	(0.109)	(0.107)	(0.072)	(0.098)						
		Proposed opt	0.198	-0.132	0.188	1.860	1.243	1.858	0.968	0.968	0.966	13.109	8.853	12.418
			(0.081)	(0.054)	(0.081)	(0.062)	(0.041)	(0.061)						
	500	LT	-53.418	36.718	-44.755	136.020	93.609	114.077	0.978	0.978	0.978	589.442	406.004	496.216
			(6.104)	(4.206)	(5.128)	(11.445)	(7.878)	(9.644)						
		WW	-0.106	0.074	-0.065	1.665	1.116	1.656	0.968	0.968	0.962	7.229	4.846	7.080
			(0.069)	(0.046)	(0.069)	(0.061)	(0.041)	(0.058)						
		Proposed	-0.409	0.277	-0.338	1.817	1.218	1.781	0.962	0.962	0.962	7.593	5.091	7.409
			(0.079)	(0.053)	(0.077)	(0.073)	(0.049)	(0.068)						
		Proposed opt	0.173	-0.115	0.173	1.442	0.965	1.451	0.964	0.964	0.958	7.802	5.248	7.564
			(0.061)	(0.041)	(0.061)	(0.050)	(0.033)	(0.049)						

The dataset contains 301 subjects. We are interested in estimating the conditional median of the log survival time (in years), denoted as y, given the age of patient on date of surgery (in years), denoted as Z_1^* as well as gender of the patient (Male=1, Female=0), denoted as Z_2. Approximately 60% of the observations are censored. Regarding the identifiability issue, following the idea proposed in Wang and Wang (2009), we examine if Condition (C3) can be satisfied by computing $1 - \hat{G}(\max(\mathbf{Z}_j^\top \hat{\boldsymbol{\beta}}^{(\hat{\phi}^{opt})} : j = 1, \ldots, n) \mid \mathbf{Z}_i)$ for i in 1 to n, where $\hat{G}(\cdot \mid \mathbf{Z})$ is the local Kaplan-Meier estimator of $G_0(\cdot \mid \mathbf{Z})$. There are 15 zero entries. We have also computed the number of zero entries for Examples 2 and 3 when n=300 and τ =0.5. The numbers are 35 and 10, respectively. Given these numbers, we believe that it is reasonable to assume that the data set satisfies condition (C3).

In the previous simulation examples, all the bandwidths adopted are selected for covariates with range 0 to 1. Hence, in this data analysis, we standardise the covariates by defining $Z_1 = Z_1^*/\{\max(Z_1^*) - \min(Z_1^*)\}$. Moreover, when choosing bandwidths, 5-fold cross validation is implemented for all four methods with the check function as the loss function for comparison. Using the Vanilla weight, our estimator gives the estimate

$$y = 3.56 - 1.25Z_1 - 0.09Z_2, \tag{5.1}$$

and adopting the Proposed opt weight, our estimate is

$$y = 3.43 - 1.12Z_1 - 0.11Z_2. \tag{5.2}$$

For the bootstrap confidence intervals, the bootstrap 95% confidence intervals for Z_1 in our Vanilla and Proposed opt approaches are $(-1.76, -0.74)$ and $(-1.59, -0.65)$, respectively, demonstrating that Z_1 is deemed significant by the model using both methods. In addition, Proposed opt offers a shorter confidence interval. On the contrary, for Z_2, the 95% confidence intervals for Z_2 in our Vanilla and Proposed opt approaches are $(-0.27, 0.09)$ and $(-0.28, 0.06)$, respectively, which concludes that Z_2 is not significant; the lengths of confi-

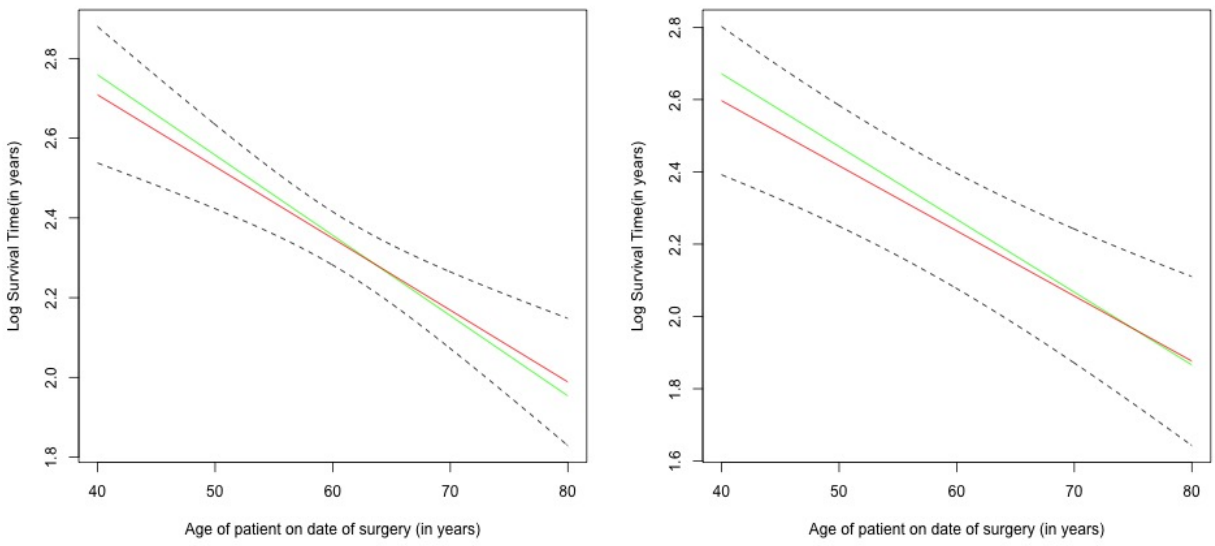

Figure 3: Estimated conditional median survival times from the vanilla and propose opt weight functions for (left) a female given her age and (right) a male given his age. The green and red lines denote the estimations from Vanilla and Proposed opt, respectively.

dence intervals for both estimators are close. In Section 3, we have proposed an alternate loss function 3.1. The estimates as well as the lengths of confidence intervals of Proposed opt using both loss functions are similar, so we only report the result using the check function as loss function here.

Figure 3 shows the estimated median log survival time (in years) versus age of patient on date of surgery (in years) for both genders. The green and red lines are the estimation from Vanilla and Proposed opt, respectively. The dashed lines represent the upper limit and lower limit of the 95% pointwise bootstrap confidence interval from Proposed opt respectively. As we can see in Table 4, which summarizes the lengths of the confidence intervals given by the four procedures considered, Proposed opt produces the smallest length in the confidence intervals of Z_1 as well as the intercept. The lengths of confidence intervals for Z_2 are close among different methods, so Proposed opt is the most efficient method in this case.

Table 4: Parameter estimates (Est), bootstrap standard errors (BSE) and lengths of confidence intervals (LCI) for various methods

	Est			BSE			LCI		
Method	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$
LT	3.56	-1.24	-0.06	0.25	0.26	0.08	0.98	1.02	0.30
WW	3.67	-1.37	-0.13	0.26	0.27	0.09	1.03	1.06	0.33
Vanilla	3.56	-1.25	-0.09	0.25	0.26	0.09	0.98	1.02	0.35
Proposed opt	3.43	-1.12	-0.11	0.24	0.24	0.09	0.93	0.94	0.34

6 Conclusion

With its growing popularity, censored quantile regression has been an active research area with a wide range of applications. In the existing literature, inference of the model parameters of interest is usually conducted via establishing martingale-based estimating equations or optimizing convex objective function that generalizes the Koenker and Bassett (1978) celebrated check function with suitable weights adjusted for partial information of censored subjects. Although quite a few methods have been suggested, there has been little discussion on the relationships amongst these approaches, including, for instance, Peng and Huang (2008), Wang and Wang (2009), Leng and Tong (2014) and De Becker et al. (2019); the interrelationships amongst them still remain unclear. Another related open question that has seldom been discussed is the estimation efficiency. Given the intricate forms of the asymptotic variance of estimators for CQR models, the challenges involved in deriving a statistically most efficient estimator are particularly prominent, especially when there is further complication due to censored observations.

In this article, we begin with constructing a mean-zero estimating equation from which our purposed vanilla estimator is asymptotically equivalent with many of the existing ones in terms of estimation efficiency. Furthermore, by studying the asymptotic variance of our estimator given an arbitrary weight function, we derive the most efficient estimator among a

general class of $\mathbf{Z}$-measurable weight functions. This optimal estimator coincide with the corresponding counterpart for non-censored data as discussed in Koenker (2005). Consistency and asymptotic normality of the proposed estimator are established, justified also numerically via simulations and our analysis of gastric adenocarcinoma patients data at Helsinki University Hospital. With the most efficient weight incorporated, our optimal estimator outperforms the existing contenders.

As a final note, we have focused our discussion on ordinary right-censored time-to-event data in this work. In various applications, however, lifetime data may be biased sampled in nature due to study design or data collecting mechanism. Special treatments are required to properly handle data with various bias sampling schemes. In particular, Xu et al. (2017) developed a martingale based estimating procedure under general bias sampling schemes. However, the corresponding efficiency issues have not yet been thoroughly investigated. The corresponding problem of seeking most efficient weight for inference merits further detailed investigations.

Supplementary Materials

The supplementary materials contain further technical details of the remark and the proposition presented in Section 2. Proofs for the main theorems summarized in Section 3 are also included.

Acknowledgement

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 Grants Council RGC-14301618, RGC-14301920 and RGC-14307221. Xu's work was partially supported by NSF SES-1659328, SES-1846747 and DMS-1712717.

References

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 Becker, M., Ghouch, A. E., and Van Keilegom, I. (2019), "An adapted loss function for censored quantile regression," *Journal of the American Statistical Association*, 114, 1126–37.

Efron, B. (1967), "The two-sample problem with censored data," in *Proceeding of the Fifth Berkeley Symposium in Mathematical Statistics IV, New York: Prentice-Hall*, eds. Le Cam, L. and Neyman, J., pp. 831–53.

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: New York.

Huang, Y. (2010), "Quantile calculus and censored regression," *Annals of Statistics*, 38, 1607–37.

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.

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.

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

Liang, H.-Y., de Uña-Álvarez, J., and del Carmen Iglesias-Pérez, M. (2012), "Asymptotic properties of conditional distribution estimator with truncated, censored and dependent data," *Test*, 21, 790–810.

Lin, Y. Y. and Chen, K. (2013), "Efficient estimation of the censored linear regression model," *Biometrika*, 100, 525–30.

Ma, Y. and Wei, Y. (2012), "Quantile analysis on residual life with longitudinal measurements," *Statistica Sinica*, 22, 47–68.

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

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

Tripathi, G. (1999), "A matrix extension of the Cauchy-Schwarz inequality," *Economics Letters*, 63, 1–3.

Tsiatis, A. A. (1990), "Estimating regression parameters using linear rank tests for censored data," *The Annals of Statistics*, 18, 354–372.

Van Keilegom, I. and Veraverbeke, N. (2001), "Hazard rate estimation in nonparametric regression with censored data," *Annals of the Institute of Statistical Mathematics*, 53, 730–745.

Wang, H. and Fygenson, M. (2009), "Inference for censored quantile regression models in longitudinal studies," *Annals of Statistics*, 37, 756–81.

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

Wei, L. J., Ying, Z., and Lin, D. Y. (1990), "Linear regression analysis of censored survival data based on rank tests," *Biometrika*, 77, 845–51.

Wei, Y. (2008), "An approach to multivariate covariate-dependent quantile contours with application to bivariate conditional growth charts," *Journal of American Statistical Association*, 193, 397–409.

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

Xu, G., Sit, T., Wang, L., and Huang, C.-Y. (2017), "Efficient estimation and inference of quantile regression under general biased sampling," *Journal of the American Statistical Association*, 112, 1571–86.

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.