

Locally Weighted Censored Quantile Regression

Huixia Judy WANG and Lan WANG

Censored quantile regression offers a valuable supplement to Cox proportional hazards model for survival analysis. Existing work in the literature often requires stringent assumptions, such as unconditional independence of the survival time and the censoring variable or global linearity at all quantile levels. Moreover, some of the work uses recursive algorithms, making it challenging to derive asymptotic normality. To overcome these drawbacks, we propose a new locally weighted censored quantile regression approach that adopts the redistribution-of-mass idea and employs a local reweighting scheme. Its validity only requires conditional independence of the survival time and the censoring variable given the covariates, and linearity at the particular quantile level of interest. Our method leads to a simple algorithm that can be conveniently implemented with R software. Applying recent theory of M-estimation with infinite dimensional parameters, we establish the consistency and asymptotic normality of the proposed estimator. The proposed method is studied via simulations and is illustrated with the analysis of an acute myocardial infarction dataset.

KEY WORDS: Kaplan–Meier estimator; Kernel; Quantile regression; Random censoring; Semiparametric; Survival analysis.

1. INTRODUCTION

Let T_i be an uncensored dependent variable of interest, such as the survival time or some transformation of the survival time, and let \mathbf{x}_i be an observable $p \times 1$ vector of covariates. In many applications, especially in biomedical studies, T_i cannot be completely observed due to possible censoring, for instance, withdrawal of patients from the study, or death from a cause unrelated to the specific disease of being studied, etc. In this paper, we focus on random right censoring, although the methodology developed can be directly extended to left censoring. Let C_i denote the censoring variable, whose distribution may depend on \mathbf{x}_i , $i = 1, \dots, n$. Due to right censoring, we only observe the triples $(\mathbf{x}_i, Y_i, \delta_i)$, where

$$Y_i = \min(T_i, C_i), \quad \delta_i = I(T_i \leq C_i) \quad (1)$$

are the observed (possibly censored) response variable and the censoring indicator, respectively. Our main objective is to estimate the p -dimensional vector quantile coefficient $\beta_0(\tau)$ for some $\tau \in (0, 1)$ in the following latent quantile regression model:

$$T_i = \mathbf{x}_i^T \beta_0(\tau) + e_i(\tau), \quad i = 1, \dots, n, \quad (2)$$

where $e_i(\tau)$ is the random error whose τ th quantile conditional on \mathbf{x}_i equals 0. Throughout the paper we use superscript T to denote the matrix transpose. Together (1) and (2) specify a censored quantile regression model. When log transformed survival time is used, (2) corresponds to the familiar accelerated failure time model.

Censored quantile regression provides a valuable complement to traditional Cox proportional hazards model for survival analysis. It relaxes the proportionality constraint on the hazard and allows for modeling heterogeneity of the data. Moreover, the quantiles of the survival time are directly interpretable. The readers are referred to Koenker and Geling (2001) and Portnoy (2003) for detailed discussions of censored quantile regression.

Censored quantile regression was first studied by Powell (1984, 1986) for fixed censoring, where the censoring times C_i are known for all observations, even for those uncensored. The fixed censoring often occurs, for instance, in biomedical studies where the measurements are censored by some quantification limits, or in social surveys where data are up to some “top-coding” or ceiling effects. However, in most survival analysis, censoring times are not always observable. In addition, the nonconvexity of Powell’s objective function imposes computational challenges; see Buchinsky (1994), Fitzenberger (1997), Koenker and Park (1996), and Fitzenberger and Winker (2007) for discussions of different algorithms. In this paper, we focus on random censoring, where censoring points are unknown for uncensored observations.

Existing literature on censored quantile regression often require stringent assumptions or complex computation. Assuming that T_i and C_i are unconditionally independent, Ying, Jung, and Wei (1995) proposed a semiparametric estimation procedure for a censored median regression model. Their procedure requires solving a discrete and nonmonotone estimating equation and thus can be difficult to implement in practice. Using the missing information principle (MIP), McKeague, Subramanian, and Sun (2001) proposed an estimating equation approach for median regression, in which the estimating functions for the censored cases are replaced by the conditional expectation given the data. In the case of discrete covariates, they studied the large sample properties of the MIP estimator, and showed that it is asymptotically equivalent to the estimator introduced by Ying, Jung, and Wei (1995). For medical cost data with informative censoring, Bang and Tsiatis (2002) developed two weighted estimating equation approaches to estimate the parameters in the median regression models. Honoré, Khan, and Powell (2002) extended Powell’s approach to random censoring assuming that C_i is independent of both \mathbf{x}_i and T_i . Lindgren (1997) transformed the problem of finding the τ th conditional quantile of T_i into a problem of finding the q th quantile of Y_i , where q is estimated using a local approach. Lindgren’s method involves an iterative minimization procedure that lacks theoretical justification. Recently, Gannoun, Saracco, and Yu (2007)

Huixia Judy Wang is Assistant Professor, Department of Statistics, North Carolina State University, Raleigh, NC 27695 (E-mail: wang@stat.ncsu.edu). Lan Wang is Associate Professor, Department of Statistics, University of Minnesota, Minneapolis, MN 55455 (E-mail: lan@stat.umn.edu). The research is partially supported by the NSF awards DMS-07-06963 and DMS-07-06842. The authors thank two referees, an Associate Editor, the Editor, and Professor Dennis D. Boos for helpful comments and suggestions, and Professors Xuming He, Wenbin Lu, and Tereza Neocleous for inspiring discussions.

proposed a local linear estimator of the conditional quantile, and compared its performance to two kernel estimators in a simulation study.

Under the more relaxed conditional independence assumption, that is, T_i and C_i are independent given \mathbf{x}_i , Portnoy (2003) developed a novel “recursive reweighting” scheme that generalizes the Kaplan–Meier estimator. Extending the redistribution-of-mass idea of Efron (1967) in the univariate case, Portnoy’s method redistributes the conditional probability $P(T_i > C_i | C_i, \mathbf{x}_i)$ of the censored cases to the right. For each censored observation, the mass is obtained through sequential pivots by fitting an entire linear quantile process. The recursive approach, however, makes it challenging to establish the asymptotic normality theory. Recently, Peng and Huang (2008) proposed a martingale-based estimating procedure that showed similar performance to Portnoy’s estimator. Both Portnoy’s and Peng and Huang’s approaches rely on a very strong global assumption, that is, in order to estimate the τ th conditional quantile of T_i given \mathbf{x}_i , it is necessary to assume that the conditional functionals at lower quantiles are all linear. This global assumption is often found to be too restrictive in practice.

We propose a new estimating procedure that relaxes the global linearity assumption. We adopt a similar redistribution-of-mass idea, but instead of estimating the censoring probability $P(T_i > C_i | C_i, \mathbf{x}_i)$ by fitting an entire linear quantile regression process, we estimate this probability nonparametrically using the local Kaplan–Meier method. Our work is partly motivated by an acute myocardial infarction study, where we are interested in the effects of gender and age on patients’ survival time. Exploratory analysis reveals that age exhibits a nonlinear effect on survival time at lower quantiles $\tau \leq 0.4$; see Figure 2 in Section 5. Therefore, Portnoy’s approach may yield bias for estimating the median survival time. Furthermore, our approach detects a significant gender effect, which is missed by Portnoy’s method.

Compared to existing procedures, our approach enjoys several distinctive advantages. First, our method only assumes linearity at one prespecified quantile level τ of interest, and thus relaxes the global assumption of Portnoy (2003) and provides more flexibility. In fact, when the global linearity assumption is seriously violated, Portnoy’s approach yields a biased estimator, and the bias does not vanish as $n \rightarrow \infty$; see Section 4 for detailed discussion. Second, our algorithm is computationally simple and easy to implement. In our method, the weights at all censored observations are estimated in one single step, and consequently the quantile estimate $\hat{\beta}(\tau)$ is obtained by minimizing only one convex objective function at the quantile of interest. Finally, our estimated weights enjoy the nice property of uniform consistency, making it feasible to establish both the consistency and asymptotic normality of the resulting estimator.

The rest of this article is organized as follows. In Section 2, we present the proposed estimating procedure. In Section 3, we establish the consistency and asymptotic normality of the resulting quantile coefficient estimator. The finite sample performance of the proposed method is investigated and compared to Portnoy’s algorithm in Section 4. The proposed method is illustrated with the analysis of the acute myocardial infarction dataset in Section 5. Section 6 concludes the paper with some discussions. The technical proofs are in the Appendix.

2. THE PROPOSED METHOD

2.1 Background

We note that model (2) amounts to assuming

$$Q_{T_i}(\tau | \mathbf{x}_i) = \mathbf{x}_i^T \boldsymbol{\beta}_0(\tau), \quad (3)$$

where $Q_{T_i}(\tau | \mathbf{x}_i) = \inf\{t: F_0(t | \mathbf{x}_i) \geq \tau\}$ is the τ th conditional quantile of T_i given the covariate \mathbf{x}_i . Before introducing the locally weighted censored quantile regression, we briefly review quantile regression for noncensored data.

When there is no censoring, $Y_i = T_i$ are observed everywhere, and $\boldsymbol{\beta}_0(\tau)$ in model (3) can be estimated by minimizing the quantile objective function

$$S_n(\boldsymbol{\beta}) = n^{-1} \sum_{i=1}^n \rho_\tau(Y_i - \mathbf{x}_i^T \boldsymbol{\beta}), \quad (4)$$

where $\rho_\tau(u) = u\{\tau - I(u < 0)\}$ is the quantile loss function. Note that a minimizer of (4) is also a root of the following estimating equation:

$$\mathbf{D}_n(\boldsymbol{\beta}) = n^{-1} \sum_{i=1}^n \mathbf{x}_i \{\tau - I(Y_i - \mathbf{x}_i^T \boldsymbol{\beta} \leq 0)\} = o_p(a_n),$$

where $a_n \rightarrow 0$ as $n \rightarrow \infty$. Under model (3), $E[\mathbf{D}_n(\boldsymbol{\beta}_0(\tau))] = 0$, and thus $\mathbf{D}_n(\boldsymbol{\beta})$ is an unbiased estimating function for $\boldsymbol{\beta}_0(\tau)$. We refer to Koenker (2005) for a comprehensive presentation of the methodology and theory of quantile regression for noncensored data.

2.2 Locally Weighted Censored Quantile Regression

Similarly to Portnoy’s approach, our method also adopts Efron’s redistribution-of-mass idea that redistributes the mass of each censored observation to the uncensored ones to the right. In the censored quantile regression setting, the idea is to redistribute the probability mass $P(T_i > C_i | C_i, \mathbf{x}_i)$ of the censored cases to the right; see Remark 1 for more explanation of the intuition. Since the quantile regression estimator is determined by the signs of residuals, the mass that is redistributed can be sent to $+\infty$.

We first consider an ideal situation where $F_0(t | \mathbf{x}_i)$, the conditional cumulative distribution function of the (sometimes unobservable) survival time T_i given \mathbf{x}_i , is assumed to be known. In this case, we define the following weight function:

$$w_i(F_0) = \begin{cases} 1, & \delta_i = 1 \text{ or } F_0(C_i | \mathbf{x}_i) > \tau \\ \frac{\tau - F_0(C_i | \mathbf{x}_i)}{1 - F_0(C_i | \mathbf{x}_i)}, & \delta_i = 0 \text{ and } F_0(C_i | \mathbf{x}_i) < \tau, \end{cases} \quad (5)$$

$i = 1, \dots, n$. If $F_0(t | \mathbf{x})$ is known, the quantile coefficient $\boldsymbol{\beta}_0(\tau)$ can be estimated by minimizing the following weighted objective function:

$$Q_n(\boldsymbol{\beta}, F_0) = n^{-1} \sum_{i=1}^n [w_i(F_0) \rho_\tau(Y_i - \mathbf{x}_i^T \boldsymbol{\beta}) + \{1 - w_i(F_0)\} \rho_\tau(Y^{+\infty} - \mathbf{x}_i^T \boldsymbol{\beta})], \quad (6)$$

where $Y^{+\infty}$ is any value sufficiently large to exceed all $\mathbf{x}_i^T \times \boldsymbol{\beta}_0(\tau)$. It is easy to derive the negative subgradient of $Q_n(\boldsymbol{\beta}, F_0)$ with respect to $\boldsymbol{\beta}$:

$$\mathbf{M}_n(\boldsymbol{\beta}, F_0) = n^{-1} \sum_{i=1}^n \mathbf{x}_i \{\tau - w_i(F_0) I(Y_i - \mathbf{x}_i^T \boldsymbol{\beta} < 0)\}. \quad (7)$$

Remark 1. To understand the weighting scheme, we note that in quantile regression, the contribution of each point to the subgradient condition only depends on the sign of the residual $T_i - \mathbf{x}_i^T \boldsymbol{\beta}_0(\tau)$. For uncensored observations, $Y_i = T_i < C_i$ and $I(T_i - \mathbf{x}_i^T \boldsymbol{\beta}_0(\tau) < 0)$ are observed. For censored observations, if $Y_i = C_i > \mathbf{x}_i^T \boldsymbol{\beta}_0(\tau)$, we immediately know that $I(T_i - \mathbf{x}_i^T \boldsymbol{\beta}_0(\tau) < 0) = 0$ since $T_i > C_i$. The ambiguous situation is $\delta_i = 0$ and $C_i < \mathbf{x}_i^T \boldsymbol{\beta}_0(\tau)$ [or equivalently $F_0(C_i | \mathbf{x}_i) < \tau$], in which case, when given (\mathbf{x}_i, C_i) ,

$$\begin{aligned} E\{I(T_i - \mathbf{x}_i^T \boldsymbol{\beta}_0(\tau) < 0) | T_i > C_i\} &= \frac{P(C_i < T_i < \mathbf{x}_i^T \boldsymbol{\beta}_0(\tau))}{P(T_i > C_i)} \\ &= \frac{\tau - F_0(C_i | \mathbf{x}_i)}{1 - F_0(C_i | \mathbf{x}_i)}. \end{aligned}$$

Therefore, in the first two situations, we assign weight $w_i(F_0) = 1$ to the observed data (\mathbf{x}_i, Y_i) . In the ambiguous situation, we assign weight $w_i(F_0) = \frac{\tau - F_0(C_i | \mathbf{x}_i)}{1 - F_0(C_i | \mathbf{x}_i)}$ to the ‘‘pseudo observation’’ at (\mathbf{x}_i, C_i) and redistribute the complimentary weight $1 - w_i(F_0)$ to any point above (\mathbf{x}_i, C_i) , for example, $(\mathbf{x}_i, Y^{+\infty})$, or $(\mathbf{x}_i, +\infty)$ without altering the quantile fit.

In reality, $F_0(C_i | \mathbf{x}_i)$ and the corresponding weights $w_i(F_0)$ are unknown and have to be estimated. We propose to estimate $F_0(\cdot | \mathbf{x})$ nonparametrically using the local Kaplan–Meier estimator $\hat{F}(\mathbf{x})$. Specifically,

$$\hat{F}(t | \mathbf{x}) = 1 - \prod_{j=1}^n \left\{ 1 - \frac{B_{nj}(\mathbf{x})}{\sum_{k=1}^n I(Y_k \geq Y_j) B_{nk}(\mathbf{x})} \right\}^{\eta_j(t)}, \quad (8)$$

where $\eta_j(t) = I(Y_j \leq t, \delta_j = 1)$ and $B_{nk}(\mathbf{x})$ is a sequence of non-negative weights adding up to 1. When $B_{nj}(\mathbf{x}) = 1/n$ for all j , $1 - \hat{F}(t | \mathbf{x})$ is just the classical Kaplan–Meier estimator of the survival function of T . Here we employ the commonly used Nadaraya–Watson’s type weights, that is,

$$B_{nk}(\mathbf{x}) = \frac{K\left(\frac{\mathbf{x} - \mathbf{x}_k}{h_n}\right)}{\sum_{i=1}^n K\left(\frac{\mathbf{x} - \mathbf{x}_i}{h_n}\right)}, \quad (9)$$

where K is a density kernel function, and $h_n \in \mathbb{R}^+$ is the bandwidth converging to zero as $n \rightarrow \infty$.

Remark 2. Nonparametric estimation of the conditional distribution function of the survival time under random censoring was first introduced by Beran (1981), and further studied by Dabrowska (1989), Gonzalez-Manteiga and Cadarso-Suarez (1994), Leconte, Poiraud-Casanova, and Thomas-Agnan (2002), among others. In practice, a variety of $B_{nk}(\mathbf{x})$ sequences are possible, and the readers are referred to Leconte, Poiraud-Casanova, and Thomas-Agnan (2002) for more detailed discussions.

By plugging the local Kaplan–Meier estimator of $F_0(C_i | \mathbf{x}_i)$ into (5), we obtain the estimated local weights $w_i(\hat{F})$. Our proposed locally weighted censored quantile regression estimator,

denoted by $\hat{\boldsymbol{\beta}}(\tau)$, is defined as the value of $\boldsymbol{\beta}$ that minimizes the locally weighted objective function

$$\begin{aligned} Q_n(\boldsymbol{\beta}, \hat{F}) &= n^{-1} \sum_{i=1}^n [w_i(\hat{F}) \rho_\tau(Y_i - \mathbf{x}_i^T \boldsymbol{\beta}) \\ &\quad + \{1 - w_i(\hat{F})\} \rho_\tau(Y^{+\infty} - \mathbf{x}_i^T \boldsymbol{\beta})]. \quad (10) \end{aligned}$$

Our method only requires the linear dependence of T_i on \mathbf{x}_i at the particular quantile of interest and thus works for a wide class of data. In addition, with our method, the weights for all the censored observations can be obtained in one single step, and we only need solve one minimization problem at the quantile level τ .

2.3 Computation

The locally weighted censored quantile regression is simple to implement with currently available R software. With estimated weights, it reduces to a weighted quantile regression on a set of appropriately defined pseudo observations. To illustrate the algorithm, we assume without loss of generality that the first n_0 observations are censored. Appending n_0 pseudo paired observations $\{(\mathbf{x}_1, Y^{+\infty}), \dots, (\mathbf{x}_{n_0}, Y^{+\infty})\}$ to the original observed data $\{(\mathbf{x}_i, Y_i), i = 1, \dots, n\}$, we obtain an augmented data set $\{(\mathbf{x}_i, Y_i), i = 1, \dots, n + n_0\}$. We first compute the local Kaplan–Meier estimator $\hat{F}(\cdot | \mathbf{x}_i)$ and weights $w_i(\hat{F}), i = 1, \dots, n$, following (8) and (5). Then the quantile estimator $\hat{\boldsymbol{\beta}}(\tau)$ can be readily computed by regressing $(Y_1, \dots, Y_{n+n_0})^T$ onto $(\mathbf{x}_1, \dots, \mathbf{x}_{n+n_0})^T$ using the function ‘‘rq’’ in R package *quantreg*, with corresponding local weights $(w_1(\hat{F}), \dots, w_n(\hat{F}), 1 - w_1(\hat{F}), \dots, 1 - w_{n_0}(\hat{F}))^T$. Since the subgradient (7) only depends on the signs of the residuals, we may take $Y^{+\infty} = +\infty$. In our implementation, we take $Y^{+\infty} = 100 \max\{Y_1, \dots, Y_n\}$.

When computing the local Kaplan–Meier estimator, we use the biquadratic kernel, that is, $K(x) = \frac{15}{16} (1 - x^2) I(|x| \leq 1)$. Different types of kernel functions may also be used and they result in little difference in practice. For multivariate covariates, it is customary to adopt a product kernel. For example, in the bivariate case we use $K(x_1, x_2) = K_1(x_1)K_2(x_2)$, where $K_1(\cdot)$ and $K_2(\cdot)$ are two univariate biquadratic kernel functions.

Although we observe in the simulations that the proposed estimator is not sensitive to the choice of smoothing parameter (see Section 4 and Remark 3 therein), for practical data analysis we recommend the computationally more intensive m -fold cross-validation method for choosing h_n ; see, for example, section 7.10 of Hastie, Tibshirani, and Friedman (2001). More specifically, we randomly divide the data into m nonoverlapped and roughly equal-sized parts. For the j th part, we fit the model using the data from the other $m - 1$ parts, and calculate the quantile loss from predicting the τ th conditional quantile of T for the uncensored cases in the j th part of the data. We repeat the above procedure for $j = 1, \dots, m$ and calculate the averaged prediction error. The h_n that yields the smallest averaged prediction error is selected. Based on our simulation experience, we recommend $m = 10$ for practical use.

3. LARGE SAMPLE PROPERTIES

3.1 Notations and Assumptions

Before stating the main theoretical results, we first define the following functions:

- (i) $1 - F_0(t|\mathbf{x}) = P(T_i > t|\mathbf{x})$, the survival function of T_i conditional on \mathbf{x} ;
- (ii) $1 - G(t|\mathbf{x}) = P(C_i > t|\mathbf{x})$, the survival function of C_i conditional on \mathbf{x} ;
- (iii) $\bar{h}(\mathbf{x})$, the marginal density function of \mathbf{x} .

To establish the asymptotic results in this paper, we require the following assumptions:

- A1 There exists a constant $K_{\mathbf{x}}$ such that $E\|\mathbf{x}\|^3 \leq K_{\mathbf{x}}$. In addition, $\max_{1 \leq i \leq n} \|\mathbf{x}_i\| = O_p(n^{1/2}(\log n)^{-1})$, and $E(\mathbf{xx}^T)$ is a positive definite $p \times p$ matrix.
- A2 The functions $F_0(t|\mathbf{x})$ and $G(t|\mathbf{x})$ have first derivatives with respect to t , denoted as $f_0(t|\mathbf{x})$ and $g(t|\mathbf{x})$, which are uniformly bounded away from infinity. In addition, $F(t|\mathbf{x})$ and $G(t|\mathbf{x})$ have bounded (uniformly in t) second-order partial derivatives with respect to \mathbf{x} .
- A3 For $\boldsymbol{\beta}$ in the neighborhood of $\boldsymbol{\beta}_0(\tau)$, $E[\mathbf{xx}^T f_0(\mathbf{x}^T \boldsymbol{\beta}|\mathbf{x}) \times \{1 - G(\mathbf{x}^T \boldsymbol{\beta}|\mathbf{x})\}]$ is positive definite.
- A4 The bandwidth is such that $h_n = O(n^{-1/2+\gamma_0})$, where $0 < \gamma_0 < 1/4$.
- A5 The kernel function $K(\cdot) \geq 0$ has a compact support. It is Lipschitz continuous of order 1 and satisfies $\int K(u) du = 1$, $\int uK(u) du = 0$, $\int K^2(u) du < \infty$, and $\int |u|^2 K(u) du < \infty$.

Assumption A1 states some tail and moment conditions on the covariate \mathbf{x} . These conditions are standard in the quantile regression literature. Assumption A2 is needed for the local Kaplan–Meier estimator. It allows us to obtain the local expansions of $F_0(t|\mathbf{x})$ and $G(t|\mathbf{x})$ in the neighborhood of $\mathbf{x}^T \boldsymbol{\beta}_0(\tau)$, and to obtain the uniform consistency and the linear representation of $\hat{F}(t|\mathbf{x})$, which are needed for deriving the consistency and asymptotic normality results in Sections 3.2 and 3.3. Assumption A3 ensures that the expectation of the estimating function $E\{M_n(\boldsymbol{\beta}, F_0)\}$ has a unique zero at $\boldsymbol{\beta}_0(\tau)$, and it is needed to establish the asymptotic distribution of $\hat{\boldsymbol{\beta}}(\tau)$. Assumptions A4 and A5 specify the conditions on the bandwidth and the kernel function. These are regular assumptions for kernel-based nonparametric smoothing.

3.2 Consistency

The following Theorem 1 states the consistency of the locally weighted censored quantile regression estimator $\hat{\boldsymbol{\beta}}(\tau)$. Its proof is given in the Appendix.

Theorem 1. Assume that the triples $\{\mathbf{x}_i, Y_i, \delta_i\}$, $i = 1, \dots, n$, constitute an iid multivariate random sample, and that the censoring variable C_i is independent of T_i conditional on the covariate \mathbf{x}_i . Under model (3) and the assumptions A1–A5,

$$\hat{\boldsymbol{\beta}}(\tau) \rightarrow \boldsymbol{\beta}_0(\tau) \quad (11)$$

in probability, as $n \rightarrow \infty$.

We show in the Appendix that $E\{\mathbf{M}_n(\boldsymbol{\beta}_0(\tau), F_0)\} = 0$ under model (2), therefore $\mathbf{M}_n(\boldsymbol{\beta}, F)$ is an unbiased estimating

function for $\boldsymbol{\beta}_0(\tau)$. To prove consistency, the main complexity comes from the fact that the objective function (10) is a non-smooth function involving a preliminary nonparametric estimator of the conditional distribution function $F_0(\cdot|\mathbf{x})$. As a result, standard M-estimation theory that involves only finite dimensional parameters such as in Pakes and Pollard (1989), He and Shao (1996), among others, is not applicable in our situation.

In this paper, we employ the recently developed empirical-process-based general theory of Chen, Linton, and Van Keilegom (2003) on nonsmooth objective functions with infinite dimensional parameters. The application of their theory is possible by using the uniform consistency of $\hat{F}(\cdot|\mathbf{x})$ for $F_0(\cdot|\mathbf{x})$; see Lemma A.1 in the Appendix.

3.3 Asymptotic Normality

In censored quantile regression, establishing asymptotic normality turns out to be much more challenging than proving consistency. Although several authors derived asymptotic normality of quantile estimators in censored regression, stringent model assumptions are needed. For instance, Powell (1986) assumed fixed censoring, and Honoré, Khan, and Powell (2002) assumed unconditional independence of T_i and C_i .

Portnoy (2003) and Neocleous, Branden, and Portnoy (2006) relaxed the unconditional independence condition, but they only verified consistency or \sqrt{n} -consistency of the quantile coefficient estimator, and they assumed global linearity. We next briefly discuss the difficulties faced by these two papers as they are more related to our work. Portnoy (2003) used a simplex pivoting algorithm in which a linear quantile regression model is fitted at a sequence of breakpoints. The censoring probability $F_0(C_i|\mathbf{x}_i)$ is then estimated by the quantile level $\hat{\tau}_i$ at which the fitted quantile regression line crosses C_i . On the other hand, Neocleous, Branden, and Portnoy (2006) proposed an improved “grid” algorithm. For both algorithms, the estimates at upper quantiles are determined by the fits at lower quantiles. This recursive nature makes it difficult to obtain the stochastic expansion of $\hat{F}(C_i|\mathbf{x}_i) - F_0(C_i|\mathbf{x}_i)$, thus complicating the derivation of asymptotic normality.

For the proposed locally weighted censored quantile regression, the local Kaplan–Meier estimator $\hat{F}(\cdot|\mathbf{x})$ is uniformly consistent for $F_0(\cdot|\mathbf{x})$, and it possesses a nice linear representation. This allows us to apply the general theory of Chen, Linton, and Van Keilegom (2003). More specifically, Chen, Linton, and Van Keilegom (2003) suggests that to establish the asymptotic normality of $\hat{\boldsymbol{\beta}}(\tau)$, the key is to show that the pathwise derivative of $E(\mathbf{M}_n(\boldsymbol{\beta}, F))$, denoted as $\boldsymbol{\Gamma}_2(\boldsymbol{\beta}_0(\tau), F_0)[\hat{F} - F_0]$, is a smooth linear function of $\hat{F} - F_0$. Using the Bahadur representation of $\hat{F}(\cdot|\mathbf{x})$, we verify in the Appendix that $\boldsymbol{\Gamma}_2(\boldsymbol{\beta}_0(\tau), F_0)[\hat{F} - F_0]$ is approximately a sum of independent mean zero random vectors, and this permits an application of the central limit theorem. The limiting distribution of $\hat{\boldsymbol{\beta}}(\tau)$ is given in the following theorem.

Theorem 2. Under the assumptions of Theorem 1 and $1/6 < \gamma_0 < 1/4$, we have

$$n^{1/2}(\hat{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}_0(\tau)) \xrightarrow{D} N(0, \boldsymbol{\Gamma}_1^{-1} \mathbf{V} \boldsymbol{\Gamma}_1^{-1}), \quad (12)$$

where $\boldsymbol{\Gamma}_1 = E[\mathbf{xx}^T \{1 - G(\mathbf{x}^T \boldsymbol{\beta}_0(\tau)|\mathbf{x})\} f_0(\mathbf{x}^T \boldsymbol{\beta}_0(\tau)|\mathbf{x})]$ and

$$\mathbf{V} = \text{Cov}(\mathbf{m}_i(\boldsymbol{\beta}_0, F_0) + (1 - \tau)\boldsymbol{\phi}(Y_i, \delta_i, \mathbf{x}_i^T \boldsymbol{\beta}_0(\tau), \mathbf{x}_i)), \quad (13)$$

with $\mathbf{m}_i(\boldsymbol{\beta}_0, F_0) = \mathbf{x}_i\{\tau - w_i(F_0)I(Y_i < \mathbf{x}_i^T \boldsymbol{\beta}_0(\tau))\}$ and $\boldsymbol{\phi}(Y_i, \delta_i, \mathbf{x}_i^T \boldsymbol{\beta}_0(\tau), \mathbf{x}_i)$ defined in equation (A.10) in the Appendix.

Inference on $\boldsymbol{\beta}_0(\tau)$ is of practical importance. However, it is seen from Theorem 2 and its proof that the asymptotic covariance matrix of $\hat{\boldsymbol{\beta}}(\tau)$ takes a complex form, and both \mathbf{V} and $\boldsymbol{\Gamma}_1$ involve unknown density functions $f_0(\cdot|\mathbf{x})$ and $g(\cdot|\mathbf{x})$ that are difficult to estimate in finite samples. For practical implementation, we adopt a simple percentile bootstrap approach through resampling the triples $(\mathbf{x}_i, Y_i, \delta_i)$ with replacement. The 95% bootstrap confidence interval for $\boldsymbol{\beta}_0(\tau)$ is constructed by taking the 2.5th and 97.5th percentiles of the bootstrap coefficients as the end points. The performance of the percentile bootstrap approach is shown to be satisfactory in the Monte Carlo studies carried out in Section 4.

4. SIMULATION RESULTS

In this section, we assess the finite sample performance of the proposed method via Monte Carlo simulations. We are mainly interested in comparing the performance of the locally weighted censored regression of quantiles (*Lcrq*) with that of Portnoy's censored regression of quantiles (*crq*). We also include another two procedures: the omniscient procedure (*omni*) that assumes knowledge of the sometimes unobservable T_i in (2) and applies the regular quantile regression to (\mathbf{x}_i, T_i) , and the naive procedure (*naive*) that completely ignores censoring and applies the regular quantile regression to (\mathbf{x}_i, Y_i) .

The four procedures are evaluated in two different settings. In the first example, we consider a conditional quantile function that is linear in the covariate at every quantile, while in the second example, the conditional quantile function is linear in the covariate only at a particular quantile.

For each example, we report the bias and mean squared error of each procedure based on 500 simulation runs. For *Lcrq* and *crq*, we also report the average coverage probabilities and average interval lengths of related resampling-based 95% confidence intervals. The results of *crq* are obtained by using the default options in the R package *quantreg* version 4.24. In this package, the *crq* confidence intervals are computed by adjusting the interquartiles of the bootstrap coefficients based on normality as described in Portnoy (2003). For both *crq* and *lcrq*, 300 bootstrap samples are used to obtain the confidence intervals in each simulation run.

Example 1. We generate random data (x_i, T_i, C_i) from the model

$$T_i = b_0 + b_1 x_i + \epsilon_i, \quad i = 1, \dots, n, \quad (14)$$

where $b_0 = 3$, $b_1 = 5$, $x_i \sim U(0, 1)$, $\epsilon_i = \eta_i - \Phi^{-1}(\tau)$, with η_1, \dots, η_n being iid standard normal random variables and Φ^{-1} being the quantile function of the standard normal distribution. Therefore, the random errors ϵ_i are centered to have zero τ th quantile. The censoring variable C_i has a uniform distribution on the interval $(0, 14)$, resulting in 40% censoring at the median. The observed response variable is $Y_i = \min(T_i, C_i)$. The global linearity assumption required for *crq* is satisfied in this example.

Table 1 summarizes the simulation results for two different values of τ : 0.5 and 0.7, and two different sample sizes $n = 200$ and 500. For *Lcrq*, the results are reported for $h_n = 0.1$ and 0.05

at $n = 200$ and 500, respectively. We also report the results of *Lcrq_{CV}*, that is, the *Lcrq* with h_n chosen by the 10-fold cross-validation method.

For this ideal example, the performance of *Lcrq* and *Lcrq_{CV}* is very close to that of *crq* and to that of the omniscient procedure in terms of bias and mean squared error. Not surprisingly, the naive procedure performs poorly. The bootstrap confidence intervals of *Lcrq* or *Lcrq_{CV}* have empirical coverage probabilities (*ECP*) close to the nominal level 95%. In contrast, the *crq* bootstrap confidence intervals have *ECP* generally below 95%, especially those for the intercept. In addition, *crq* leads to universally wider confidence intervals than *Lcrq* and *Lcrq_{CV}*. Similar patterns are observed for $n = 1,000$ and censoring probabilities 0.2, 0.3, and 0.5, and the results are not reported due to space limitation.

Example 2. We generate random data (x_i, T_i, C_i) , $i = 1, \dots, n$, such that

$$T_i = b_0 + b_1 x_i + (0.2 + 2(x_i - 0.5)^2)\epsilon_i, \quad (15)$$

where $b_0 = 2$, $b_1 = 1$, $x_i \sim N(0, 1)$, and ϵ_i are generated following the same way as in Example 1. The censoring variable C_i has a uniform distribution on the interval $(0, 7)$ yielding 40% censoring at the median.

Table 2 summarizes the simulation results for $\tau = 0.5$ and 0.7 with sample sizes $n = 200$ and 500, respectively. In this example, the conditional quantile of T_i is linear in x_i only at the τ th quantile and is quadratic in x_i everywhere else. The *crq* procedure is quite robust to the linearity assumption in the sense that its bias is not unreasonably large; however, we observe that the bias does not diminish with increased sample size. The bias of *crq* for estimating b_1 increases to around 0.10 when the sample size n increases from 200 to 500. Even for sample size 1,000 (not reported), we observe that the bias of *crq* is still around 0.1, while that of *Lcrq* or *Lcrq_{CV}* is below 0.01 at both $\tau = 0.5$ and 0.7. Similarly as in Example 1, the *crq* bootstrap confidence intervals have coverage probabilities considerably below 95%. The *Lcrq* and *Lcrq_{CV}* intervals not only maintain more accurate coverage probabilities, but also are shorter than those of *crq*. More specifically, the average lengths of the *Lcrq* confidence intervals are about 89% of those of *crq* for estimating b_0 , and about 90% of those of *crq* for estimating b_1 .

Remark 3. In the extensive simulation studies carried out, we note that the performance of *Lcrq* is not sensitive to the choice of h_n . To give an idea of the sensitivity, we plot in Figure 1 the MSE versus h_n for estimating b_0 and b_1 in Example 1 with $n = 200$. The figure suggests that the MSE varies little for $h_n \in [0.01, 0.15]$.

5. ANALYSIS OF THE ACUTE MYOCARDIAL INFARCTION DATASET

To illustrate the proposed method, we analyze a subset of data from a study conducted at the University Clinical Center in Ljubljana. The purpose of the study was to investigate the survival of patients after acute myocardial infarction. The dataset contains information on 972 patients ranging in age from 40 to 80 years, of which 48% have censored outcomes. More details about the study can be found in Pohar and Stare (2006), and the dataset is included in the R package *relsurv*. We are interested

Table 1. Simulation results for Example 1, based on 500 simulation runs. 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. The last two rows are the averaged Monte Carlo standard errors of different measurements for *crq* and *Lcrq_{CV}*, respectively

τ	n	Method	Bias		MSE		ECP		EML	
			b_0	b_1	b_0	b_1	b_0	b_1	b_0	b_1
0.5	200	<i>omni</i>	0.004	-0.002	0.029	0.092				
		<i>crq</i>	-0.010	-0.011	0.042	0.157	0.916	0.942	0.835	1.567
		<i>Lcrq</i>	-0.005	-0.019	0.041	0.157	0.960	0.968	0.816	1.564
		<i>Lcrq_{CV}</i>	-0.010	-0.024	0.041	0.157	0.966	0.976	0.816	1.562
		<i>naive</i>	-0.208	-1.120	0.102	1.540				
	500	<i>omni</i>	-0.003	0.003	0.012	0.036				
		<i>crq</i>	-0.008	-0.005	0.016	0.058	0.928	0.936	0.512	0.987
		<i>Lcrq</i>	-0.008	-0.014	0.016	0.059	0.966	0.956	0.512	0.979
		<i>Lcrq_{CV}</i>	-0.010	-0.016	0.016	0.060	0.968	0.954	0.512	0.977
		<i>naive</i>	-0.212	-1.099	0.068	1.324				
0.7	200	<i>omni</i>	0.003	-0.005	0.034	0.098				
		<i>crq</i>	-0.014	0.001	0.045	0.158	0.900	0.932	0.882	1.691
		<i>Lcrq</i>	-0.005	-0.010	0.047	0.163	0.958	0.964	0.862	1.656
		<i>Lcrq_{CV}</i>	-0.009	-0.018	0.048	0.164	0.946	0.964	0.858	1.644
		<i>naive</i>	-0.172	-0.609	0.073	0.526				
	500	<i>omni</i>	-0.006	0.003	0.013	0.041				
		<i>crq</i>	-0.010	-0.000	0.018	0.063	0.908	0.930	0.545	1.035
		<i>Lcrq</i>	-0.008	-0.007	0.018	0.063	0.944	0.968	0.531	1.017
		<i>Lcrq_{CV}</i>	-0.010	-0.007	0.018	0.063	0.936	0.964	0.531	1.019
		<i>naive</i>	-0.165	-0.605	0.045	0.428				
		<i>crq</i>	0.008	0.014	0.002	0.007	0.013	0.011	0.011	0.016
		<i>Lcrq_{CV}</i>	0.008	0.014	0.002	0.007	0.009	0.008	0.007	0.011

Table 2. Simulation results for Example 2, based on 500 simulation runs. 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. The last two rows are the averaged Monte Carlo standard errors of different measurements for *crq* and *Lcrq_{CV}*, respectively

τ	n	Method	Bias		MSE		ECP		EML	
			b_0	b_1	b_0	b_1	b_0	b_1	b_0	b_1
0.5	200	<i>omni</i>	-0.004	0.013	0.020	0.079				
		<i>crq</i>	0.034	-0.084	0.025	0.102	0.928	0.900	0.641	1.247
		<i>Lcrq</i>	-0.053	0.007	0.022	0.074	0.952	0.948	0.577	1.131
		<i>Lcrq_{CV}</i>	-0.065	0.020	0.022	0.073	0.948	0.948	0.571	1.113
		<i>naive</i>	-0.417	-0.264	0.194	0.123				
	500	<i>omni</i>	-0.002	0.001	0.009	0.034				
		<i>crq</i>	0.035	-0.099	0.011	0.051	0.904	0.878	0.387	0.762
		<i>Lcrq</i>	-0.052	-0.001	0.011	0.035	0.898	0.960	0.353	0.697
		<i>Lcrq_{CV}</i>	-0.061	0.013	0.011	0.033	0.900	0.952	0.353	0.702
		<i>naive</i>	-0.414	-0.256	0.179	0.086				
0.7	200	<i>omni</i>	-0.000	0.015	0.023	0.089				
		<i>crq</i>	0.042	-0.091	0.028	0.108	0.912	0.892	0.658	1.268
		<i>Lcrq</i>	-0.034	-0.007	0.023	0.086	0.942	0.956	0.583	1.150
		<i>Lcrq_{CV}</i>	-0.049	0.000	0.023	0.081	0.930	0.950	0.567	1.125
		<i>naive</i>	-0.216	-0.134	0.062	0.077				
	500	<i>omni</i>	-0.001	0.008	0.010	0.037				
		<i>crq</i>	0.043	-0.106	0.014	0.055	0.890	0.852	0.407	0.781
		<i>Lcrq</i>	-0.037	-0.010	0.011	0.039	0.902	0.932	0.368	0.723
		<i>Lcrq_{CV}</i>	-0.045	0.002	0.011	0.040	0.918	0.932	0.364	0.717
		<i>naive</i>	-0.220	-0.128	0.055	0.040				
		<i>crq</i>	0.006	0.012	0.001	0.005	0.013	0.014	0.007	0.014
		<i>Lcrq_{CV}</i>	0.005	0.010	0.001	0.004	0.012	0.010	0.004	0.008

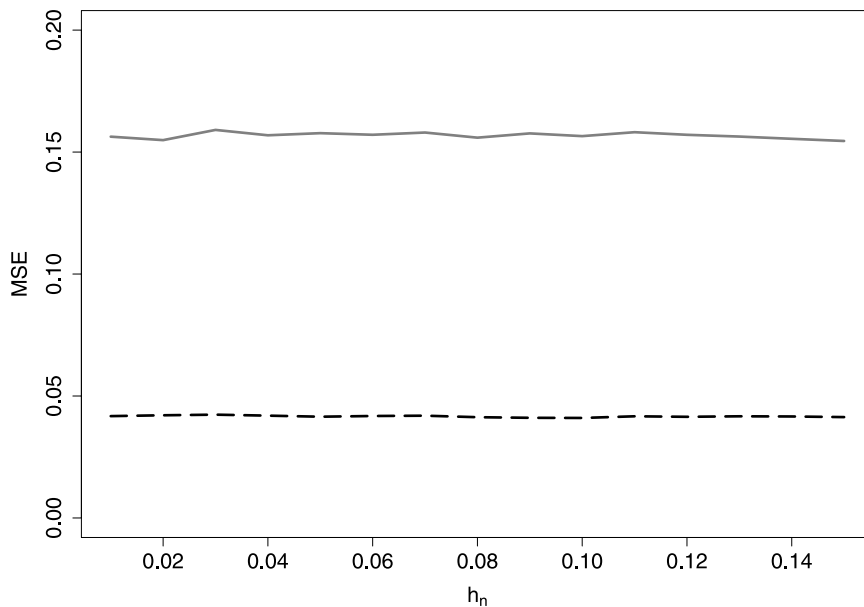


Figure 1. Plot of MSE versus h_n in Example 2 with $n = 200$. The dashed and solid lines represent the MSE for estimating b_0 and b_1 , respectively.

in estimating the conditional median of the log survival time (in days), a quantity of interest to medical doctors, given age (in years) and gender (1 = male, 0 = female).

To explore the dependence of log survival time and age at different quantiles, in a preliminary analysis we fit a semiparametric regression model by using polynomial B-splines to approximate the effect of age. Figure 2 plots the estimated log survival times against age at quantiles $\tau = 0.01, 0.03, 0.05, 0.1, 0.2, 0.3, 0.4,$ and 0.5 . This plot reveals that at lower quantiles the assumption of linearity of age on survival time might be violated. This suggests that *Lcrq* may be more appropriate than *crq* for analyzing this dataset.

Next, we report and compare results from both *Lcrq* and *crq* for estimating the median survival time. As the interaction effect of age and gender is shown to be insignificant in the preliminary analysis, we consider the model with main effects of age and gender. Let y denote the conditional median log survival time. The *Lcrq* yields that at $\tau = 0.5$,

$$y = 10.506 - 0.042 \text{ age} + 0.222 \text{ gender}. \quad (16)$$

The 95% bootstrap confidence interval for age is $(-0.052, -0.031)$ and for gender is $(0.012, 0.355)$. This suggests that increased age and being a female are associated with significantly shorter median survival time. For comparison, Portnoy's

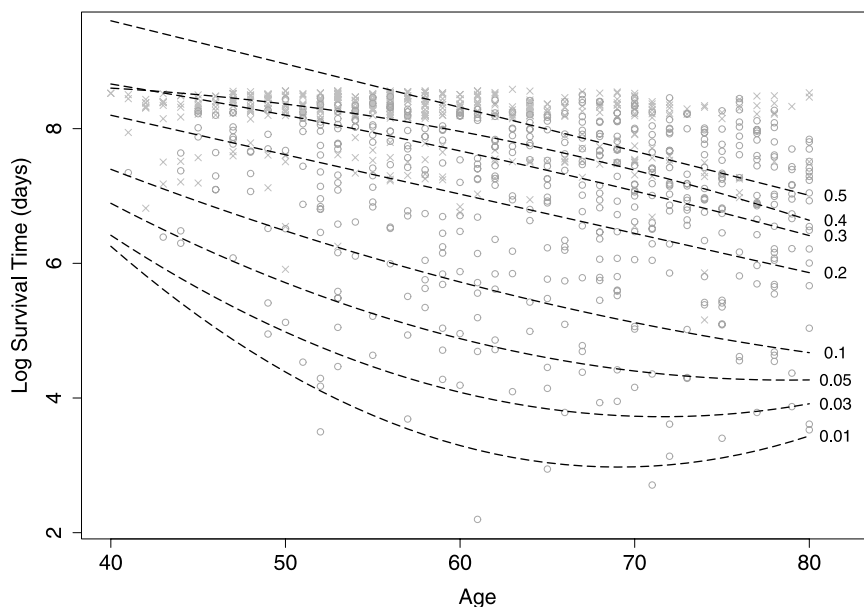


Figure 2. Acute myocardial infarction data: the fitted log survival times (from a partial linear model) against age at quantiles 0.01, 0.03, 0.05, 0.1, 0.2, 0.3, 0.4, and 0.5. The open circles denote uncensored observations, and the crosses represent censored points.

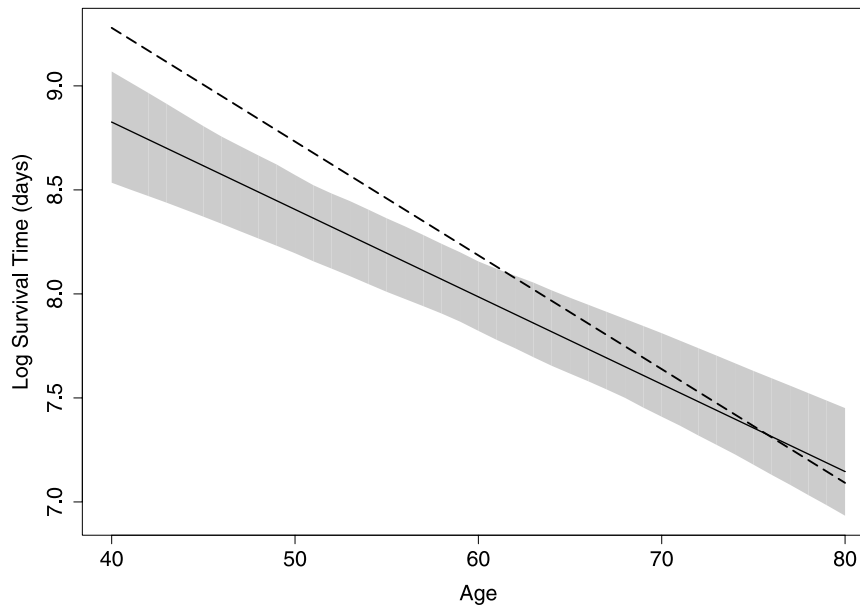


Figure 3. The estimated conditional median survival time from *crq* and *Lcrq* for female. The dashed and solid lines are the estimation from *crq* and *Lcrq*, respectively. The shaded area represents the 95% pointwise bootstrap confidence band from *Lcrq*.

crq gives

$$y = 11.466 - 0.055 \text{ age} + 0.157 \text{ gender.} \quad (17)$$

The associated 95% bootstrap confidence interval for age is $(-0.072, -0.037)$ and for gender is $(-0.125, 0.439)$. As observed in the simulation study, *crq* gives wider confidence intervals than *Lcrq*. Moreover, *crq* fails to identify the significant gender effect.

Figure 3 shows the estimated conditional median survival time (on log scale) from *Lcrq* and *crq* for female patients. The dashed line is the conditional median estimation from *crq*, and the solid line is from *Lcrq*. The shaded area represents a 95% pointwise bootstrap confidence band following our method. For age < 60 , the median estimates from *crq* fall outside the confidence band, suggesting a noticeable difference from the *Lcrq* estimates. Such difference is also observed in the estimation of male median survival time.

6. CONCLUDING REMARKS

We propose a new locally weighted censored quantile regression approach that relaxes the stringent model assumptions, such as unconditional independence or global linearity, required by many existing procedures. Our approach employs the redistribution-of-mass idea and is computationally simpler, as it directly estimates the quantile of interest and avoids fitting an entire quantile process. The new procedure adopts a preliminary local Kaplan–Meier estimator and leads to a weighted quantile regression. Utilizing results in modern empirical process theory, we establish the consistency and asymptotic normality of the resulted estimator. Both simulation studies and the analysis of acute myocardial infarction data show that the proposed method leads to shorter interval estimates than Portnoy’s *crq* procedure.

For right censored data, quantile functions with τ close to 1 may not be identifiable due to censorship. To ensure $\beta_0(\tau)$ is

estimable, we need some area of \mathbf{x} where not all censoring happens below the τ th conditional quantile line, otherwise the matrix $E[\mathbf{xx}^T f_0(\mathbf{x}^T \boldsymbol{\beta} | \mathbf{x}) \{1 - G(\mathbf{x}^T \boldsymbol{\beta} | \mathbf{x})\}]$ in condition A3 may be singular. A sufficient but not necessary condition to ensure the identifiability is that on some subset of \mathbf{x} with positive measure, we have $1 - G(\mathbf{x}^T \boldsymbol{\beta}_0(\tau) | \mathbf{x}) = P(C > \mathbf{x}^T \boldsymbol{\beta}_0(\tau) | \mathbf{x}) > 0$. In principle, the quantile level τ is chosen according to practical interest. To check whether $\boldsymbol{\beta}_0(\tau)$ is estimable, empirically we can estimate $G(\cdot | \mathbf{x}_i)$ by the local Kaplan–Meier estimator $\hat{G}(\cdot | \mathbf{x}_i)$ following the procedure described in Section 2.2. If a large fraction of $\{1 - \hat{G}(\mathbf{x}_i^T \hat{\boldsymbol{\beta}}(\tau) | \mathbf{x}_i), i = 1, \dots, n\}$ are zero, we should be cautious about estimating and making inference of the τ th quantile.

The interquartile range method was introduced in Portnoy (2003) to construct confidence intervals. The purpose was to account for the randomness of the τ -cutoff point (the largest quantile level up to the identifiability constraint) in different bootstrap samples. We find that in finite samples the empirical distribution of $\hat{\boldsymbol{\beta}}(\tau)$ from the bootstrap samples often deviates from normality. Our simulations also confirm that the interquartile range method often gives confidence intervals that tend to under cover the true coefficients. When the τ th quantile is identifiable, we recommend use of the simpler percentile method described in Section 3.3.

Finally, we observe in our empirical studies that Portnoy’s *crq* algorithm exhibits good robustness against modest violation of the global linearity assumption. However, when global linearity is seriously violated, *crq* leads to estimators with bias that does not diminish with increased sample size. In such cases, the results from the proposed *Lcrq* method are more trustable.

APPENDIX

Throughout the appendix, the τ will be omitted in various expressions such as $\boldsymbol{\beta}_0(\tau)$ and $e_i(\tau)$, but we should bear in mind that these quantities are all τ -specific. Let \mathcal{B} denote a compact finite dimensional parameter set, and \mathcal{H} denote an infinite dimensional parameter space.

Let $F = F(\cdot|\mathbf{x}) \in \mathcal{H}$ be a nonparametric conditional distribution function. For notational simplicity, from now on we shall write $F_0(t|\mathbf{x}_i)$, $F(t|\mathbf{x}_i)$, $G(t|\mathbf{x}_i)$, $g(t|\mathbf{x}_i)$ as $F_{0i}(t)$, $F_i(t)$, $G_i(t)$, and $g_i(t)$, respectively, whenever clear from the context. Recall that the negative subgradient of the objective function $Q_n(\boldsymbol{\beta}, F)$ is defined in (7) as

$$\mathbf{M}_n(\boldsymbol{\beta}, F) = n^{-1} \sum_{i=1}^n \mathbf{m}_i(\boldsymbol{\beta}, F), \quad (\text{A.1})$$

where

$$\begin{aligned} \mathbf{m}_i(\boldsymbol{\beta}, F) &= \mathbf{x}_i \{\tau - w_i(F)I(Y_i \leq \mathbf{x}_i^T \boldsymbol{\beta})\} \\ &= \mathbf{x}_i \left[I(T_i \leq C_i) \{\tau - I(T_i \leq \mathbf{x}_i^T \boldsymbol{\beta})\} + I(T_i > C_i) \right. \\ &\quad \times \left. \left\{ \tau - \left(1 - \frac{1-\tau}{1-F_i(C_i)} I(F_i(C_i) < \tau) \right) I(C_i \leq \mathbf{x}_i^T \boldsymbol{\beta}) \right\} \right] \\ &= \mathbf{x}_i \left[I(C_i \leq \mathbf{x}_i^T \boldsymbol{\beta}) \left\{ I(T_i \leq C_i)(\tau - 1) \right. \right. \\ &\quad \left. \left. + I(T_i > C_i) \left(\tau - 1 + \frac{1-\tau}{1-F_i(C_i)} I(F_i(C_i) < \tau) \right) \right\} \right. \\ &\quad \left. + I(C_i > \mathbf{x}_i^T \boldsymbol{\beta}) \right. \\ &\quad \left. \times \left\{ \tau I(T_i \leq C_i) - I(T_i \leq \mathbf{x}_i^T \boldsymbol{\beta}) + \tau I(T_i > C_i) \right\} \right] \\ &= \mathbf{x}_i \left[\left\{ \tau - I(C_i > \mathbf{x}_i^T \boldsymbol{\beta}, T_i \leq \mathbf{x}_i^T \boldsymbol{\beta}) - I(C_i \leq \mathbf{x}_i^T \boldsymbol{\beta}, T_i \leq C_i) \right\} \right. \\ &\quad \left. - I(C_i \leq \mathbf{x}_i^T \boldsymbol{\beta}, T_i > C_i) \right. \\ &\quad \left. \times \left(1 - \frac{1-\tau}{1-F_i(C_i)} I(F_i(C_i) < \tau) \right) \right]. \quad (\text{A.2}) \end{aligned}$$

By the conditional independence of T and C given \mathbf{x} , we have

$$\begin{aligned} E\{I(C > t, T < t)|\mathbf{x}\} &= P(C > t|\mathbf{x})P(T < t|\mathbf{x}) = \{1 - G(t|\mathbf{x})\}F_0(t|\mathbf{x}), \\ E\{I(C < t, T < C)|\mathbf{x}\} &= E_{C|\mathbf{x}}\{I(C < t)P(T < C|C, \mathbf{x})\} \\ &= \int_{-\infty}^t F_0(u|\mathbf{x})g(u|\mathbf{x}) du. \end{aligned}$$

Therefore,

$$\mathbf{M}(\boldsymbol{\beta}, F) \doteq E[\mathbf{m}_i(\boldsymbol{\beta}, F)] = E[\mathbf{x}\{\tau - H(\mathbf{x}^T \boldsymbol{\beta}|\mathbf{x}) - R(\boldsymbol{\beta}, F|\mathbf{x})\}],$$

where

$$\begin{aligned} H(t|\mathbf{x}) &= \{1 - G(t|\mathbf{x})\}F_0(t|\mathbf{x}) + \int_{-\infty}^t F_0(u|\mathbf{x})g(u|\mathbf{x}) du, \\ R(\boldsymbol{\beta}, F|\mathbf{x}) &= \int_{-\infty}^{\mathbf{x}^T \boldsymbol{\beta}} \{1 - F_0(u|\mathbf{x})\} \\ &\quad \times \left[1 - \frac{1-\tau}{1-F(u|\mathbf{x})} I\{F(u|\mathbf{x}) < \tau\} \right] g(u|\mathbf{x}) du. \end{aligned}$$

In $\mathbf{M}(\boldsymbol{\beta}, F)$, the expectation is with respect to the distribution of \mathbf{x} .

We first state and prove two lemmas. We use $\|\cdot\|$ to denote the Euclidean norm, and $\|\cdot\|_{\mathcal{H}}$ to denote a supnorm metric.

Lemma A.1. Assume assumptions A2, A4, and A5 hold. Then

$$\begin{aligned} \|\hat{F} - F_0\|_{\mathcal{H}} &\doteq \sup_t \sup_{\mathbf{x}} |\hat{F}(t|\mathbf{x}) - F_0(t|\mathbf{x})| \\ &= O_p((\log n)^{1/2} n^{-1/4 - \gamma_0/2}), \quad (\text{A.3}) \end{aligned}$$

where $0 < \gamma_0 < 1/4$.

Proof. This follows directly from theorem 2.1 of Gonzalez-Manteiga and Cadarso-Suarez (1994).

Lemma A.2. For all positive values $\varepsilon_n = o(1)$, we have

$$\begin{aligned} \sup_{\|\boldsymbol{\beta} - \boldsymbol{\beta}_0\| \leq \varepsilon_n, \|F - F_0\|_{\mathcal{H}} \leq \varepsilon_n} \|\mathbf{M}_n(\boldsymbol{\beta}, F) - \mathbf{M}(\boldsymbol{\beta}, F) - \mathbf{M}_n(\boldsymbol{\beta}_0, F_0)\| \\ = o_p(n^{-1/2}). \quad (\text{A.4}) \end{aligned}$$

Proof. Let x_{ij} and m_{ij} denote the j th coordinates of \mathbf{x}_i and \mathbf{m}_i , respectively. For notational simplicity, in the following we omit the subscript i in various expressions such as $\mathbf{x}_i, x_{ij}, T_i, C_i, F_i$, and G_i . Let $K_j, j = 1, \dots, 5$, be some positive constants. Note that for $j = 1, \dots, p$, $|m_j(\boldsymbol{\beta}, F) - m_j(\boldsymbol{\beta}', F')|^2 \leq B_1 + B_2 + B_3$, where

$$\begin{aligned} B_1 &= x_j^2 |I(C > \mathbf{x}^T \boldsymbol{\beta}, T < \mathbf{x}^T \boldsymbol{\beta}) - I(C > \mathbf{x}^T \boldsymbol{\beta}', T < \mathbf{x}^T \boldsymbol{\beta}')|, \\ B_2 &= x_j^2 |I(C < \mathbf{x}^T \boldsymbol{\beta}, T < C) - I(C < \mathbf{x}^T \boldsymbol{\beta}', T < C)|, \\ B_3 &= x_j^2 \left| I(C < \mathbf{x}^T \boldsymbol{\beta}, T > C) \left[1 - \frac{1-\tau}{1-F(C)} I\{F(C) < \tau\} \right] \right. \\ &\quad \left. - I(C < \mathbf{x}^T \boldsymbol{\beta}', T > C) \left[1 - \frac{1-\tau}{1-F'(C)} I\{F'(C) < \tau\} \right] \right|. \end{aligned}$$

It's easy to verify that

$$\begin{aligned} \sup_{\boldsymbol{\beta}': \|\boldsymbol{\beta} - \boldsymbol{\beta}'\| \leq \varepsilon_n} |I(\mathbf{x}^T \boldsymbol{\beta} < C) - I(\mathbf{x}^T \boldsymbol{\beta}' < C)| \\ \leq \|\mathbf{x}\| |I(\mathbf{x}^T \boldsymbol{\beta} - \varepsilon_n < C) - I(\mathbf{x}^T \boldsymbol{\beta} + \varepsilon_n < C)|. \end{aligned}$$

Therefore, by assumptions A1 and A2,

$$\begin{aligned} E \left(\sup_{\boldsymbol{\beta}': \|\boldsymbol{\beta} - \boldsymbol{\beta}'\| \leq \varepsilon_n} B_1 \right) \\ \leq E \left[\sup_{\boldsymbol{\beta}': \|\boldsymbol{\beta} - \boldsymbol{\beta}'\| \leq \varepsilon_n} x_j^2 \{ |I(C > \mathbf{x}^T \boldsymbol{\beta}') - I(C > \mathbf{x}^T \boldsymbol{\beta})| \right. \\ \left. + |I(T < \mathbf{x}^T \boldsymbol{\beta}') - I(T < \mathbf{x}^T \boldsymbol{\beta})| \right] \\ \leq E[\|\mathbf{x}\|^3 \{G(\mathbf{x}^T \boldsymbol{\beta} + \varepsilon_n) - G(\mathbf{x}^T \boldsymbol{\beta} - \varepsilon_n)\} \\ + \|\mathbf{x}\|^3 \{F_0(\mathbf{x}^T \boldsymbol{\beta} + \varepsilon_n) - F_0(\mathbf{x}^T \boldsymbol{\beta} - \varepsilon_n)\}] \\ \leq K_1 \varepsilon_n. \end{aligned}$$

Following the similar arguments, we can show that $E(\sup_{\boldsymbol{\beta}': \|\boldsymbol{\beta} - \boldsymbol{\beta}'\| \leq \varepsilon_n} B_2) \leq K_2 \varepsilon_n$. Note that

$$\begin{aligned} B_3 &\leq x_j^2 I(C < \mathbf{x}^T \boldsymbol{\beta}) \\ &\quad \times \left| \frac{1-\tau}{1-F'(C)} I\{F'(C) < \tau\} - \frac{1-\tau}{1-F(C)} I\{F(C) < \tau\} \right| \\ &\quad + x_j^2 \left[1 - \frac{1-\tau}{1-F'(C)} I\{F'(C) < \tau\} \right] \\ &\quad \times |I(C < \mathbf{x}^T \boldsymbol{\beta}) - I(C < \mathbf{x}^T \boldsymbol{\beta}')| \\ &\leq x_j^2 \left| \frac{1-\tau}{1-F'(C)} I\{F'(C) < \tau\} - \frac{1-\tau}{1-F(C)} I\{F(C) < \tau\} \right| \\ &\quad + x_j^2 |I(C < \mathbf{x}^T \boldsymbol{\beta}) - I(C < \mathbf{x}^T \boldsymbol{\beta}')| \\ &\doteq B_{31} + B_{32}. \end{aligned}$$

It is easy to see that

$$\begin{aligned} B_{31} &= x_j^2 I\{F(C) < \tau, F'(C) < \tau\} \frac{|F(C) - F'(C)|(1-\tau)}{\{1 - F'(C)\}(1 - F(C))} \\ &\quad + x_j^2 I\{F(C) < \tau < F'(C)\} \frac{1-\tau}{1-F(C)} \end{aligned}$$

$$\begin{aligned}
& + x_j^2 I\{F'(C) < \tau < F(C)\} \frac{1 - \tau}{1 - F'(C)} \\
& \leq \frac{x_j^2}{1 - \tau} |F(C) - F'(C)| + x_j^2 I\{F(C) < \tau < F'(C)\} \\
& \quad + x_j^2 I\{F'(C) < \tau < F(C)\}.
\end{aligned}$$

Since

$$\begin{aligned}
& E \left[\sup_{F': \|F - F'\|_{\mathcal{H}} \leq \varepsilon_n} I\{F(C) < \tau < F'(C)\} \right] \\
& \leq P\{F(C) < \tau < F(C) + \varepsilon_n\} \\
& = G\{F^{-1}(\tau)\} - G\{F^{-1}(\tau - \varepsilon_n)\} \leq K_3 \varepsilon_n.
\end{aligned}$$

Therefore, by assumption A1, we have $E(\sup_{F': \|F - F'\|_{\mathcal{H}} \leq \varepsilon_n} B_{31}) \leq K_4 \varepsilon_n$. Following similar arguments used in proving B_1 , we obtain $E(\sup_{\beta': \|\beta - \beta'\| \leq \varepsilon_n} B_{32}) \leq K_5 \varepsilon_n$. Therefore, condition (3.2) of Chen, Linton, and Van Keilegom (2003) holds with $r = 2$ and $s_j = 1/2$, and condition (3.3) is satisfied by remark 3(ii) of their paper. Thus, Lemma A.2 holds by applying theorem 3 of Chen, Linton, and Van Keilegom (2003).

Proof of Theorem 1

Note that $F_0(t|\mathbf{x}) < \tau$ is equivalent to $t < \mathbf{x}^T \boldsymbol{\beta}_0$, and $F_0(\mathbf{x}^T \boldsymbol{\beta}_0|\mathbf{x}) = \tau$. Therefore, when plugging in the true $\boldsymbol{\beta}_0$ and F_0 into \mathbf{M} , we get

$$\mathbf{M}(\boldsymbol{\beta}_0, F_0) = E(x[\tau - \tau\{1 - G(\mathbf{x}^T \boldsymbol{\beta}_0|\mathbf{x})\} - \tau G(\mathbf{x}^T \boldsymbol{\beta}_0|\mathbf{x})]) = 0.$$

Because $\boldsymbol{\beta}_0$ is the solution of $\mathbf{M}(\boldsymbol{\beta}, F_0) = 0$ with $\mathbf{M}(\boldsymbol{\beta}, F_0)$ being a monotone function of $\boldsymbol{\beta}$, it is in the interior of a compact parameter space \mathcal{B} . Therefore, the consistency of $\hat{\boldsymbol{\beta}}$ is the direct conclusion of theorem 1 of Chen, Linton, and Van Keilegom (2003), and we only need verify the conditions (1.1)–(1.2) and (1.5') in their paper, as (1.3) is trivially satisfied and (1.4) follows from Lemma A.1.

(1.1) By the subgradient condition of quantile regression (Koenker 2005), there exists a vector \mathbf{v} with coordinates $|v_i| \leq 1$ such that

$$\begin{aligned}
\|\mathbf{M}_n(\hat{\boldsymbol{\beta}}, \hat{w})\| &= n^{-1} \|(\mathbf{x}_i v_i : i \in \Xi)\| \\
&= o_p(n^{-1/2}), \quad \text{by assumption A1, (A.5)}
\end{aligned}$$

where Ξ denotes a p -element subset of $\{1, 2, \dots, n\}$.

(1.2) For any $\varepsilon > 0$ and $\boldsymbol{\beta} \in \mathcal{B}$,

$$\begin{aligned}
& \inf_{\|\boldsymbol{\beta} - \boldsymbol{\beta}_0\| > \varepsilon} \|\mathbf{M}(\boldsymbol{\beta}, F_0)\| \\
&= \inf_{\|\boldsymbol{\beta} - \boldsymbol{\beta}_0\| > \varepsilon} \|\mathbf{M}(\boldsymbol{\beta}_0, F_0) - \mathbf{M}(\boldsymbol{\beta}, F_0)\| \\
&= \inf_{\|\boldsymbol{\beta} - \boldsymbol{\beta}_0\| > \varepsilon} \|E[\mathbf{x}\{H(\mathbf{x}^T \boldsymbol{\beta}|\mathbf{x}) - H(\mathbf{x}^T \boldsymbol{\beta}_0|\mathbf{x})\} \\
& \quad + R(\boldsymbol{\beta}, F_0|\mathbf{x}) - R(\boldsymbol{\beta}_0, F_0|\mathbf{x})]\| \\
&\geq \inf_{\|\boldsymbol{\beta} - \boldsymbol{\beta}_0\| > \varepsilon} \|E[\mathbf{x}\mathbf{x}^T(\boldsymbol{\beta} - \boldsymbol{\beta}_0)\{1 - G(\xi^*|\mathbf{x})\}f_0(\xi^*|\mathbf{x})]\|,
\end{aligned}$$

which is strictly positive under assumptions A1 and A3. Here ξ^* is some value between $\mathbf{x}^T \boldsymbol{\beta}$ and $\mathbf{x}^T \boldsymbol{\beta}_0$.

(1.5') Let $\{a_n\}$ be a sequence of positive numbers approaching zero as $n \rightarrow \infty$. Note that $E\{\|\mathbf{x}_i w_i I(Y_i \leq \mathbf{x}_i^T \boldsymbol{\beta}_0)\|^2\} \leq E(\|\mathbf{x}_i\|^2) \leq K_{\mathbf{x}}$, under assumption A1. It then follows from Chebyshev's inequality that

$$\sup_{\boldsymbol{\beta} \in \mathcal{B}, \|F - F_0\|_{\mathcal{H}} \leq a_n} \|\mathbf{M}_n(\boldsymbol{\beta}, F) - \mathbf{M}(\boldsymbol{\beta}, F)\| = o_p(1).$$

Proof of Theorem 2

The asymptotic normality of $\hat{\boldsymbol{\beta}}$ relies on the results of theorem 2 in Chen, Linton, and Van Keilegom (2003). We need prove the conditions (2.1)–(2.4), (2.5'), and (2.6') in their paper. The conditions (2.1), (2.4), and (2.5') hold directly by (A.5), Lemmas A.1 and A.2, respectively. Note that for any C_i lying above the τ th conditional quantile $\mathbf{x}_i^T \boldsymbol{\beta}_0$, the quantile fit will not be affected if we assign the entire weight to either (\mathbf{x}_i, C_i) or $(\mathbf{x}_i, Y^{+\infty})$. Bearing this in mind, we obtain

$$\begin{aligned}
\Gamma_1(\boldsymbol{\beta}_0, F_0) &= \frac{\partial \mathbf{M}(\boldsymbol{\beta}, F_0)}{\partial \boldsymbol{\beta}} \Big|_{\boldsymbol{\beta} = \boldsymbol{\beta}_0} \\
&= -E[\mathbf{x}\mathbf{x}^T \{1 - G(\mathbf{x}^T \boldsymbol{\beta}_0|\mathbf{x})\} f_0(\mathbf{x}^T \boldsymbol{\beta}_0|\mathbf{x})],
\end{aligned}$$

which is continuous at $\boldsymbol{\beta}_0$ and of full rank under assumption A3. For all $\boldsymbol{\beta} \in \mathcal{B}$, we define the functional derivative of $\mathbf{M}(\boldsymbol{\beta}, F)$ at F_0 in the direction $[F - F_0]$ as

$$\begin{aligned}
& \Gamma_2(\boldsymbol{\beta}, F_0)[F - F_0] \\
&= E \left(\lim_{\varepsilon \rightarrow 0} \frac{1}{\varepsilon} [\mathbf{M}(\boldsymbol{\beta}, F_0 + \varepsilon(F - F_0)) - \mathbf{M}(\boldsymbol{\beta}, F_0)] \right) \\
&= (1 - \tau) E \left[\mathbf{x} \int_{-\infty}^{\mathbf{x}^T \boldsymbol{\beta}} \frac{F(t|\mathbf{x}) - F_0(t|\mathbf{x})}{1 - F_0(t|\mathbf{x})} I\{F_0(t|\mathbf{x}) < \tau\} g(t|\mathbf{x}) dt \right. \\
& \quad \left. + E\{\mathbf{x}A(\boldsymbol{\beta})\} \right], \tag{A.6}
\end{aligned}$$

where $A(\boldsymbol{\beta}) = \lim_{\varepsilon \rightarrow 0} \varepsilon^{-1} \int_{-\infty}^{\mathbf{x}^T \boldsymbol{\beta}} g(t|\mathbf{x}) [I\{F_\varepsilon(t|\mathbf{x}) < \tau\} - I\{F_0(t|\mathbf{x}) < \tau\}] dt$ and $F_\varepsilon(t|\mathbf{x}) = F_0(t|\mathbf{x}) + \varepsilon(F(t|\mathbf{x}) - F_0(t|\mathbf{x}))$. Firstly, it's easy to verify that for $\boldsymbol{\beta}$ such that $\mathbf{x}^T \boldsymbol{\beta} \leq \mathbf{x}^T \boldsymbol{\beta}_0$, $A(\boldsymbol{\beta}) = 0$. For $\boldsymbol{\beta}$ such that $\mathbf{x}^T \boldsymbol{\beta} > \mathbf{x}^T \boldsymbol{\beta}_0$, $F_\varepsilon^{-1}(\tau|\mathbf{x}) < \mathbf{x}^T \boldsymbol{\beta}$ for ε sufficiently small. In this case,

$$\begin{aligned}
A(\boldsymbol{\beta}) &= \lim_{\varepsilon \rightarrow 0} \varepsilon^{-1} \int_{-\infty}^{\mathbf{x}^T \boldsymbol{\beta}} g(t|\mathbf{x}) [I\{F_\varepsilon(t|\mathbf{x}) < \tau\} - I\{F_0(t|\mathbf{x}) < \tau\}] dt \\
&= \lim_{\varepsilon \rightarrow 0} \varepsilon^{-1} \int_{-\infty}^{\mathbf{x}^T \boldsymbol{\beta}} g(t|\mathbf{x}) [I\{\mathbf{x}^T \boldsymbol{\beta}_0 \leq t \leq \mathbf{x}^T \boldsymbol{\beta}\} \\
& \quad - I\{F_\varepsilon^{-1}(\tau|\mathbf{x}) \leq t \leq \mathbf{x}^T \boldsymbol{\beta}\}] dt \\
&= \lim_{\varepsilon \rightarrow 0} \varepsilon^{-1} [G\{F_\varepsilon^{-1}(\tau|\mathbf{x})|\mathbf{x}\} - G(\mathbf{x}^T \boldsymbol{\beta}_0|\mathbf{x})].
\end{aligned}$$

For $\boldsymbol{\beta} = \boldsymbol{\beta}_0$, note that $I\{F_0(t|\mathbf{x}) < \tau\} = 1$ for $t \in (-\infty, \mathbf{x}^T \boldsymbol{\beta}_0)$, so

$$\begin{aligned}
A(\boldsymbol{\beta}_0) &= - \lim_{\varepsilon \rightarrow 0} \varepsilon^{-1} \int_{-\infty}^{\mathbf{x}^T \boldsymbol{\beta}_0} g(t|\mathbf{x}) I\{F_\varepsilon(t|\mathbf{x}) > \tau\} dt \\
&= \lim_{\varepsilon \rightarrow 0} \varepsilon^{-1} [G\{F_\varepsilon^{-1}(\tau|\mathbf{x})|\mathbf{x}\} - G(\mathbf{x}^T \boldsymbol{\beta}_0|\mathbf{x})].
\end{aligned}$$

By expanding $G\{F_\varepsilon^{-1}(\tau|\mathbf{x})\}$ (treated as a function of ε) around $\varepsilon = 0$, and using the fact that $\frac{d}{d\varepsilon} F_\varepsilon^{-1}(\tau|\mathbf{x})|_{\varepsilon=0} = \frac{\tau - F(\mathbf{x}^T \boldsymbol{\beta}_0)}{f_0(\mathbf{x}^T \boldsymbol{\beta}_0)}$ (example 20.5 in van der Vaart 1998), we obtain

$$G(F_\varepsilon^{-1}(\tau|\mathbf{x})|\mathbf{x}) = G(\mathbf{x}^T \boldsymbol{\beta}_0|\mathbf{x}) + g(\mathbf{x}^T \boldsymbol{\beta}_0|\mathbf{x}) \frac{\tau - F(\mathbf{x}^T \boldsymbol{\beta}_0|\mathbf{x})}{f_0(\mathbf{x}^T \boldsymbol{\beta}_0|\mathbf{x})} \varepsilon + O(\varepsilon^2).$$

Therefore, for $\boldsymbol{\beta}$ such that $\mathbf{x}^T \boldsymbol{\beta} \geq \mathbf{x}^T \boldsymbol{\beta}_0$,

$$A(\boldsymbol{\beta}) = g(\mathbf{x}^T \boldsymbol{\beta}_0|\mathbf{x}) \frac{F_0(F_0^{-1}(\tau)) - F(F_0^{-1}(\tau))}{f_0(\mathbf{x}^T \boldsymbol{\beta}_0|\mathbf{x})}. \tag{A.7}$$

Following the routine Taylor expansion, we can verify condition (2.3) under assumptions A1 and A2. We now verify condition (2.6). Com-

bining (A.6) and (A.7), we have

$$\begin{aligned} & \Gamma_2(\boldsymbol{\beta}_0, F_0)[\hat{F} - F_0] \\ &= (1 - \tau)E \left[\mathbf{x} \int_{-\infty}^{\mathbf{x}^T \boldsymbol{\beta}_0} \frac{\hat{F}(t|\mathbf{x}) - F_0(t|\mathbf{x})}{1 - F_0(t|\mathbf{x})} g(t|\mathbf{x}) dt \right. \\ & \quad \left. + \mathbf{x} \frac{g(\mathbf{x}^T \boldsymbol{\beta}_0|\mathbf{x})}{f_0(\mathbf{x}^T \boldsymbol{\beta}_0|\mathbf{x})} \{F_0(\mathbf{x}^T \boldsymbol{\beta}_0|\mathbf{x}) - \hat{F}(\mathbf{x}^T \boldsymbol{\beta}_0|\mathbf{x})\} \right] \\ &= (1 - \tau) \left[\int_{-\infty}^{+\infty} \mathbf{x} \bar{h}(\mathbf{x}) \int_{-\infty}^{\mathbf{x}^T \boldsymbol{\beta}_0} \frac{g(t|\mathbf{x})}{1 - F_0(t|\mathbf{x})} \right. \\ & \quad \times \{\hat{F}(t|\mathbf{x}) - F_0(t|\mathbf{x})\} dt d\mathbf{x} \\ & \quad \left. - \int_{-\infty}^{+\infty} \mathbf{x} \bar{h}(\mathbf{x}) \frac{g(\mathbf{x}^T \boldsymbol{\beta}_0|\mathbf{x})}{f_0(\mathbf{x}^T \boldsymbol{\beta}_0|\mathbf{x})} \right. \\ & \quad \left. \times \{\hat{F}(\mathbf{x}^T \boldsymbol{\beta}_0|\mathbf{x}) - F_0(\mathbf{x}^T \boldsymbol{\beta}_0|\mathbf{x})\} d\mathbf{x} \right]. \quad (\text{A.8}) \end{aligned}$$

It follows from theorem 2.3 of Gonzalez-Manteiga and Cadarso-Suarez (1994) that under assumptions A2–A5,

$$\begin{aligned} \hat{F}(t|\mathbf{x}) - F_0(t|\mathbf{x}) &= (nh_n)^{-1} \sum_{j=1}^n K\left(\frac{\mathbf{x} - \mathbf{x}_j}{h_n}\right) \xi(Y_j, \delta_j, t, \mathbf{x}) \\ & \quad + O_p\left(\left(\frac{\log n}{nh_n}\right)^{3/4} + h_n^2\right), \end{aligned}$$

where

$$\begin{aligned} & \xi(Y_j, \delta_j, t, \mathbf{x}) \\ &= \{1 - F_0(t|\mathbf{x})\} \left[- \int_0^{\min(Y_j, t)} \frac{f_0(s|\mathbf{x}) ds}{\{1 - F_0(s|\mathbf{x})\}^2 \{1 - G(s|\mathbf{x})\}} \right. \\ & \quad \left. + \frac{I(Y_j \leq t, \delta_j = 1)}{\{1 - F_0(Y_j|\mathbf{x})\} \{1 - G(Y_j|\mathbf{x})\}} \right] \end{aligned}$$

are independent random variables with mean zero and finite variances for any given t and \mathbf{x} .

We now plug this linear representation of $\hat{F}(t) - F_0(t)$ in $\Gamma_2(\boldsymbol{\beta}_0, F_0)[\hat{F} - F_0]$. Using standard change of variables and Taylor expansion arguments, and the assumption that $\int K(u) du = 1$, we obtain

$$\begin{aligned} & \Gamma_2(\boldsymbol{\beta}_0, F_0)[\hat{F} - F_0] \\ &= (1 - \tau)n^{-1} \sum_{i=1}^n \boldsymbol{\phi}(Y_i, \delta_i, \mathbf{x}_i^T \boldsymbol{\beta}_0, \mathbf{x}_i) + o_p(n^{-1/2}), \quad (\text{A.9}) \end{aligned}$$

where

$$\begin{aligned} & \boldsymbol{\phi}(Y_i, \delta_i, \mathbf{x}_i^T \boldsymbol{\beta}_0, \mathbf{x}_i) \\ &= \mathbf{x}_i \bar{h}(\mathbf{x}_i) \left\{ \int_{-\infty}^{\mathbf{x}_i^T \boldsymbol{\beta}_0} \frac{g(t|\mathbf{x}_i) \xi(Y_i, \delta_i, t, \mathbf{x}_i)}{1 - F_0(t|\mathbf{x}_i)} dt \right. \\ & \quad \left. - \frac{g(\mathbf{x}_i^T \boldsymbol{\beta}_0|\mathbf{x}_i)}{f_0(\mathbf{x}_i^T \boldsymbol{\beta}_0|\mathbf{x}_i)} \xi(Y_i, \delta_i, \mathbf{x}_i^T \boldsymbol{\beta}_0, \mathbf{x}_i) \right\} \quad (\text{A.10}) \end{aligned}$$

is a random vector with mean 0 and $E\{\|\boldsymbol{\phi}(Y_i, \delta_i, \mathbf{x}_i^T \boldsymbol{\beta}_0, \mathbf{x}_i)\|^2\} < \infty$ under assumptions A1–A3. Recall that $\mathbf{M}_n(\boldsymbol{\beta}_0, F_0) = n^{-1} \times \sum_{i=1}^n \mathbf{m}_i(\boldsymbol{\beta}_0, F_0)$ with $\mathbf{m}_i(\boldsymbol{\beta}_0, F_0)$ being independent mean 0 random vectors. From (A.2), it's easy to verify that

$$\begin{aligned} & \text{Cov}\{\mathbf{m}_i(\boldsymbol{\beta}_0, F_0)\} \\ &= E_{\{\mathbf{x}, C\}} \left[\mathbf{xx}^T \left\{ I(C \leq \mathbf{x}^T \boldsymbol{\beta}_0) \frac{F_0(C|\mathbf{x})(1 - \tau)^2}{1 - F_0(C|\mathbf{x})} \right. \right. \end{aligned}$$

$$\left. \left. + I(C > \mathbf{x}^T \boldsymbol{\beta}_0) \tau(1 - \tau) \right\} \right]$$

$$\doteq d_1.$$

Combining (A.1) and (A.9), and applying the central limit theorem gives

$$n^{1/2} \{\mathbf{M}_n(\boldsymbol{\beta}_0, w_0) + \Gamma_2(\boldsymbol{\beta}_0, F_0)[\hat{F} - F_0]\} \xrightarrow{D} N(0, \mathbf{V}),$$

where

$$\mathbf{V} = \text{Cov}\{\mathbf{m}_i(\boldsymbol{\beta}_0, F_0) + (1 - \tau)\boldsymbol{\phi}(Y_i, \delta_i, \mathbf{x}_i^T \boldsymbol{\beta}_0, \mathbf{x}_i)\} \doteq d_1 + d_2 + d_3,$$

$$d_2 = (1 - \tau)E\{\mathbf{m}_i(\boldsymbol{\beta}_0, F_0)\boldsymbol{\phi}^T(Y_i, \delta_i, \mathbf{x}_i^T \boldsymbol{\beta}_0, \mathbf{x}_i)\},$$

$$d_3 = (1 - \tau)^2 E\{\boldsymbol{\phi}(Y_i, \delta_i, \mathbf{x}_i^T \boldsymbol{\beta}_0, \mathbf{x}_i)\boldsymbol{\phi}^T(Y_i, \delta_i, \mathbf{x}_i^T \boldsymbol{\beta}_0, \mathbf{x}_i)\}.$$

The proof for (12) is thus complete by theorem 2 of Chen, Linton, and Van Keilegom (2003).

[Received April 2008. Revised December 2008.]

REFERENCES

- Bang, H., and Tsiatis, A. A. (2002), "Median Regression With Censored Cost Data," *Biometrics*, 58, 643–649.
- Beran, R. (1981), "Nonparametric Regression With Randomly Censored Survival Data", technical report, University of California, Berkeley.
- Buchinsky, M. (1994), "Changes in the U.S. Wage Structure 1963–1987: Applications of Quantile Regression," *Econometrica*, 62, 405–458.
- Chen, X., Linton, O., and Van Keilegom, I. (2003), "Estimation of Semiparametric Models When the Criterion Function Is Not Smooth," *Econometrica*, 71, 1591–1608.
- Dabrowska, D. M. (1989), "Uniform Consistency of Kernel Conditional Kaplan–Meier Estimate," *The Annals of Statistics*, 17, 1157–1167.
- Efron, B. (1967), "The Two-Sample Problem With Censored Data," in *Proceedings Fifth Berkeley Symposium in Mathematical Statistics*, Vol. IV, eds. L. Le Cam and J. Neyman, New York: Prentice-Hall, pp. 831–853.
- Fitzenberger, B. (1997), "A Guide to Censored Quantile Regressions," in *Handbook of Statistics, Vol. 15: Robust Inference*, eds. G. S. Maddala and C. R. Rao, Amsterdam: North-Holland, pp. 405–437.
- Fitzenberger, B., and Winker, P. (2007), "Improving the Computation of Censored Quantile Regressions," *Computational Statistics and Data Analysis*, 52, 88–108.
- Gannoun, A., Saracco, J., and Yu, K. (2007), "Comparison of Kernel Estimators of Conditional Distribution Function and Quantile Regression Under Censoring," *Statistical Modelling*, 7, 329–344.
- Gonzalez-Manteiga, W., and Cadarso-Suarez, C. (1994), "Asymptotic Properties of a Generalized Kaplan–Meier Estimator With Some Applications," *Journal of Nonparametric Statistics*, 4, 65–78.
- Hastie, T., Tibshirani, R., and Friedman, J. (2001), *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, New York: Springer.
- He, X., and Shao, Q. M. (1996), "A General Bahadur Representation of M-Estimators and Its Application to Linear Regression With Nonstochastic Designs," *The Annals of Statistics*, 24, 2608–2630.
- Honoré, B., Khan, S., and Powell, J. L. (2002), "Quantile Regression Under Random Censoring," *Journal of Econometrics*, 109, 67–105.
- Koenker, R. (2005), *Quantile Regression*, London: Cambridge University Press.
- Koenker, R., and Geling, O. (2001), "Reappraising Medfly Longevity: A Quantile Regression Survival Analysis," *Journal of the American Statistical Association*, 96, 458–468.
- Koenker, R., and Park, B. (1996), "An Interior Point Algorithm for Nonlinear Quantile Regression," *Journal of Econometrics*, 71, 265–283.
- Leconte, E., Poiraud-Casanova, S., and Thomas-Agnan, C. (2002), "Smooth Conditional Distribution Function and Quantiles Under Random Censorship," *Lifetime Data Analysis*, 8, 229–246.
- Lindgren, A. (1997), "Quantile Regression With Censored Data Using Generalized L_1 Minimization," *Computational Statistics and Data Analysis*, 23, 509–524.
- McKeague, I. W., Subramanian, S., and Sun, Y. (2001), "Median Regression and the Missing Information Principle," *Journal of Nonparametric Statistics*, 13, 709–727.
- Neocleous, T., Vanden Branden, K., and Portnoy, S. (2006), Correction to "Censored Regression Quantiles," by S. Portnoy, 98 (2003), 1001–1012, *Journal of the American Statistical Association*, 101, 860–861.
- Pakes, A., and Pollard, D. (1989), "Simulation and the Asymptotics of Optimization Estimators," *Econometrica*, 57, 1027–1057.

- Peng, L., and Huang, Y. (2008), "Survival Analysis With Quantile Regression Models," *Journal of the American Statistical Association*, 103, 637–649.
- Pohar, M., and Stare, J. (2006), "Relative Survival Analysis in R," *Computer Methods and Programs in Biomedicine*, 81, 272–278.
- Portnoy, S. (2003), "Censored Regression Quantiles," *Journal of the American Statistical Association*, 98, 1001–1012.
- Powell, J. L. (1984), "Least Absolute Deviations Estimation for the Censored Regression Model," *Journal of Econometrics*, 25, 303–325.
- (1986), "Censored Regression Quantiles," *Journal of Econometrics*, 32, 143–155.
- van der Vaart, A. W. (1998), *Asymptotic Statistics*, Cambridge: Cambridge University Press.
- Ying, Z., Jung, S. H., and Wei, L. J. (1995), "Survival Analysis With Median Regression Models," *Journal of the American Statistical Association*, 90, 178–184.