*The Annals of Statistics*
2022, Vol. 50, No. 5, 2899–2924
https://doi.org/10.1214/22-AOS2214


# SCALABLE ESTIMATION AND INFERENCE FOR CENSORED QUANTILE REGRESSION PROCESS

BY XUMING HE[1,a], XIAOOU PAN[2,c], KEAN MING TAN[1,b] AND WEN-XIN ZHOU[2,d]

[1]*Department of Statistics, University of Michigan,* [a]*xmhe@umich.edu,* [b]*keanming@umich.edu*
[2]*Department of Mathematical Sciences, University of California, San Diego,* [c]*xip024@ucsd.edu,* [d]*wez243@ucsd.edu*

Censored quantile regression (CQR) has become a valuable tool to study the heterogeneous association between a possibly censored outcome and a set of covariates, yet computation and statistical inference for CQR have remained a challenge for large-scale data with many covariates. In this paper, we focus on a smoothed martingale-based sequential estimating equations approach, to which scalable gradient-based algorithms can be applied. Theoretically, we provide a unified analysis of the smoothed sequential estimator and its penalized counterpart in increasing dimensions. When the covariate dimension grows with the sample size at a sublinear rate, we establish the uniform convergence rate (over a range of quantile indexes) and provide a rigorous justification for the validity of a multiplier bootstrap procedure for inference. In high-dimensional sparse settings, our results considerably improve the existing work on CQR by relaxing an exponential term of sparsity. We also demonstrate the advantage of the smoothed CQR over existing methods with both simulated experiments and data applications.

**1. Introduction.** Censored data are prevalent in many applications where the response variable of interest is partially observed, mostly due to loss of follow-up. For instance, in a lung cancer study considered by Shedden et al. [50], 46.6% of the lung cancer patients' survival time are censored, due to either early withdrawal from the study or death because of other reasons that are unrelated to lung cancer. Commonly used methods to study the association between the censored response and explanatory variables (covariates) are through the use of Cox proportional hazards model and the accelerated failure time (AFT) model [1, 31]. Both models assume homogeneous covariate effects and are not applicable to cases in which the lower and upper quantiles of the conditional distribution of the censored response, potentially with different covariate effects, are of interest. Moreover, in many scientific studies, higher or lower quantiles of the response variable are more of interest than the mean. To capture heterogeneous covariate effects and to better predict the response at different quantile levels, various censored quantile regression (CQR) methods have been developed under different assumptions on the censoring mechanism [6, 9–11, 25, 38, 45, 48, 49, 59, 63, 64]. We refer the reader to Chapters 6 and 7 in [34] as well as [43] for a comprehensive review of censored quantile regression.

We consider the random right censoring mechanism, in which the censoring points are unknown for the uncensored observations. Statistical methods for CQR were first proposed under the stringent assumption that the uncensored response variable (not observable due to censoring) is marginally independent of the censoring variable; see, for example, [25, 64]. Under a more relaxed conditional independence assumption, conditioned on the covariates, [45] generalized the Kaplan–Meier estimator for estimating the (univariate) survival function to the regression setting, based on Efron's [14] redistribution-of-mass construction. From a

Received January 2022; revised April 2022.
*MSC2020 subject classifications.* Primary 62J05, 62J07; secondary 62F40.
*Key words and phrases.* Censored quantile regression, smoothing, high-dimensional survival data, nonasymptotic theory, weighted bootstrap.

different perspective, [44] employed a martingale-based approach for fitting CQR, and the resulting method has been shown to be closely related to [45]'s method [40, 42]. Both [45]'s and [44]'s methods, along with their variants, involve solving a series of quantile regression problems that can be reformulated as linear programs, solvable by the simplex or interior point method [3, 36, 46]. Statistical properties of the aforementioned methods have been well studied, assuming that the number of covariates, $p$, is fixed [40, 42, 44, 47]. To this date, the impact of dimensionality in the increasing-$p$ regime, in which $p$ is allowed to increase with the number of observations, has remained unclear in the presence of censored outcomes.

In the high-dimensional setting in which $p > n$, convex and nonconvex penalty functions are often employed to perform variable selection and to achieve a trade-off between statistical bias and model complexity. While penalized Cox proportional hazards and AFT models have been well studied [5, 7, 16, 27], existing work on penalized CQR under the framework of [45] and [44] in the high-dimensional setting is still lagging. Large-sample properties of penalized CQR estimators were first derived under the fixed-$p$ setting ($p < n$), mainly due to the technical challenges introduced by the sequential nature of the procedure [52, 58, 60]. More recently, [67] studied a penalized CQR estimator, extending the method of [44] to the high-dimensional setting ($p > n$). They showed that the estimation error (under $\ell_2$-norm) of the $\ell_1$-penalized CQR estimator is upper bounded by $\mathcal{O}(\exp(Cs)\sqrt{s\log(p)/n})$ with high probability, where $C > 0$ is a dimension-free constant. Compared to the $\ell_1$-penalized QR for uncensored data [4], whose convergence rate is of order $\mathcal{O}(\sqrt{s\log(p)/n})$, there is a substantial gap in terms of the impact of the sparsity parameter $s$.

In addition to the above theoretical issues, our study is motivated by the computational hardness of CQR under the framework of [45] and [44] for problems with large dimension. Recall that this framework involves fitting a series of quantile regressions sequentially over a dense grid of quantile indexes, each of which is solvable by the Frisch–Newton algorithm with computational complexity that grows as a cubic function of $p$ [46]. Moreover, under the regime in which $p < n$, the asymptotic covariance matrix of the estimator is rather complicated, and thus resampling methods are often used to perform statistical inference [44, 45]. A sample-based inference procedure (without resampling) for Peng–Huang's estimator [44] is available by adapting the plug-in covariance estimation method from [53]. In the high-dimensional setting ($p > n$), computation of the $\ell_1$-penalized QR is based on either reformulation as linear programs [37] or alternating direction method of multiplier algorithms [22, 65]. These algorithms are generic and applicable to a broad spectrum of problems but lack scalability. Since the $\ell_1$-penalized CQR not only requires the estimation of the whole quantile regression process, but also relies on cross-validation to select the sequence of (mostly different) penalty levels, the state-of-the-art methods [17, 67] can be highly inefficient when applied to large-$p$ problems.

To illustrate the computational challenge for CQR, we compare the $\ell_1$-penalized CQR proposed by Zheng et al. [67] and our proposed method by analyzing a gene expression data set studied in [50]. In this study, 22,283 genes from 442 lung adenocarcinomas are incorporated to predict the survival time in lung cancer, with 46.6% subjects that are censored. We implement both methods with quantile grid set as $\{0.1, 0.11, \ldots, 0.7\}$, and use a predetermined sequence of regularization parameters. For Zheng et al. [67], we use the `rqPen` package to compute the $\ell_1$-penalized QR estimator at each quantile level [51]. The computational time and maximum allocated memory are reported in Table 1. The reference machine for this experiment is a worker node with 2.5 GHz 32-core processor and 512 GB of memory in a high-performance computing cluster.

In this paper, we develop a smoothed framework for CQR that is scalable to problems with large dimension $p$ in both low- and high-dimensional settings. Our proposed method is motivated by the smoothed estimating equation approach that has surfaced mostly in the

TABLE 1

*Computational runtime and maximum allocated memory for fitting $\ell_1$-penalized CQR and the proposed method on the gene expression data with censored response in* [50]. *One gigabyte* (*GB*) *equals* 1024 *megabytes* (*MB*)

| Methods | Runtime | Allocated memory |
|---|---|---|
| $\ell_1$-penalized CQR | 170 hours+ | 38 GB |
| Proposed method | 2 minutes | 926 MB |

econometrics literature [12, 18, 23, 30, 61, 62], which can be applied to the stochastic integral based sequential estimation procedure proposed by Peng and Huang [44] for CQR. We show in Section 2.2 that the smoothed sequential estimating equations method can be reformulated as solving a sequence of optimization problems with (at least) twice differentiable and convex loss functions for which gradient-based algorithms are available. Large-scale statistical inference can then be performed efficiently via multiplier/weighted bootstrap. In the high-dimensional setting, we propose and analyze $\ell_1$-penalized smoothed CQR estimators obtained by sequentially minimizing smoothed convex loss functions plus $\ell_1$-penalty, which we solve using a scalable and efficient majorize-minimization-type algorithm, as evidenced in Table 1.

Theoretically, we provide a unified analysis for the proposed smoothed estimator in both low- and high-dimensional settings. In the low-dimensional case where the dimension is allowed to increase with the sample size, we establish the uniform rate of convergence and a uniform Bahadur-type representation for the smoothed CQR estimator. We also provide a rigorous justification for the validity of a weighted/multiplier bootstrap procedure with explicit error bounds as functions of $(n, p)$. To our knowledge, these are the first results for censored quantile regression in the increasing-$p$ regime with $p < n$. The main challenges are as follows. To fit the QR process with censored response variables, the stochastic integral based approach entails a sequence of estimating equations that correspond to a prespecified grid of quantile indexes. A sequence of pointwise estimators can then be sequentially obtained by solving these equations. The sequential nature of this procedure poses technical challenges because at each quantile level, the objective function (or the estimating equation) depends on all of the previous estimates. To establish convergence rates for the estimated regression process, a delicate analysis beyond what is used in [23] is required to deal with the accumulated estimation error sequentially. The mesh width of the grid should converge to zero at a proper rate in order to balance the accumulated estimation error and discretization error. In the high-dimensional setting, we show that with suitably chosen penalty levels and bandwidth, the $\ell_1$-penalized smoothed CQR estimator has a uniform convergence rate of $\mathcal{O}(\sqrt{s\log(p)/n})$, provided the sample size satisfies $n \gtrsim s^3\log(p)$. The technical arguments used in this case are also very different from those in [67] and subsequent work [17], and as a result, our conclusion improves that of Zheng et al. [67] by relaxing the exponential term $\exp(Cs)$ in the convergence rate to a linear term in $s$. Such an improvement is significant when the effective model size $s$ is allowed to grow with $n$ and $p$ in the context of censored quantile regression.

The rest of the article is organized as follows. In Section 2, we provide a formal formulation of the CQR. We then briefly review the martingale-based estimating equation estimator proposed by Peng and Huang [44] in Section 2.1. The proposed smoothed CQR is detailed in Section 2.2, along with the multiplier bootstrap method for large-scale inference in Section 2.3. We then provide a comprehensive theoretical analysis for the smoothed CQR estimator in Section 3 and its bootstrap counterpart. In Section 4, we generalize the smoothed CQR to the high-dimensional setting by incorporating a penalty function to the smoothed

CQR loss and study the theoretical properties of the regularized estimator. Extensive numerical studies and data applications are in Sections 5 and 6. The R code that implements the proposed method is available at https://github.com/XiaoouPan/scqr.

*Notation.* For any two real numbers $a$ and $b$, we write $a \wedge b = \min\{a, b\}$ and $a \vee b = \max\{a, b\}$. Given a pair of vectors $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^p$, we use $\boldsymbol{u}^{\mathrm{T}}\boldsymbol{v}$ and $\langle \boldsymbol{u}, \boldsymbol{v} \rangle$ interchangeably to denote their inner product. For a positive semidefinite matrix $\Sigma \in \mathbb{R}^{p \times p}$, we define the $\Sigma$-induced $\ell_2$-norm $\|\boldsymbol{u}\|_\Sigma = \|\Sigma^{1/2}\boldsymbol{u}\|_2$ for any $\boldsymbol{u} \in \mathbb{R}^p$. For every $r \geq 0$, we use $\mathbb{B}^p(r) = \{\boldsymbol{\beta} \in \mathbb{R}^p : \|\boldsymbol{\beta}\|_2 \leq r\}$ and $\mathbb{S}^{p-1}(r) = \{\boldsymbol{\beta} \in \mathbb{R}^p : \|\boldsymbol{\beta}\|_2 = r\}$ to denote the Euclidean ball and sphere, respectively, with radius $r$. In particular, we write $\mathbb{S}^{p-1} = \mathbb{S}^{p-1}(1)$. Given an event/subset $\mathcal{A}$, $\mathbb{1}\{\mathcal{A}\}$ or $\mathbb{1}_{\mathcal{A}}$ represents the indicator function of this event/subset. For two nonnegative arrays $\{a_n\}_{n \geq 1}$ and $\{b_n\}_{n \geq 1}$, we write $a_n \lesssim b_n$ if $a_n \leq C b_n$ for some constant $C > 0$ independent of $n$, $a_n \gtrsim b_n$ if $b_n \lesssim a_n$, and $a_n \asymp b_n$ if $a_n \lesssim b_n$ and $a_n \gtrsim b_n$.

**2. Censored quantile regression.** Let $z \in \mathbb{R}$ be a response variable of interest, and $\boldsymbol{x} = (x_1, \ldots, x_p)^{\mathrm{T}}$ be a $p$-vector ($p \geq 2$) of random covariates with $x_1 \equiv 1$. In this work, we focus on a global conditional quantile model on $z$ described as follows. Given a closed interval $[\tau_L, \tau_U] \subseteq (0, 1)$, assume that the $\tau$th conditional quantile of $z$ given $\boldsymbol{x}$ takes the form

$$F^{-1}_{z|\boldsymbol{x}}(\tau) = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{\beta}^*(\tau) \quad \text{for any } \tau \in [\tau_L, \tau_U], \tag{1}$$

where $\boldsymbol{\beta}^*(\tau) \in \mathbb{R}^p$, formulated as a function of $\tau$, is the unknown vector of regression coefficients.

We assume that $z$ is subject to right censoring by $C$, a random variable that is conditionally independent of $z$ given the covariates $\boldsymbol{x}$. Let $y = z \wedge C$ the censored outcome, and $\Delta = \mathbb{1}(z \leq C)$ be an event indicator. The observed sample $\{y_i, \Delta_i, \boldsymbol{x}_i\}_{i=1}^n$ consists of independent and identically distributed (i.i.d.) replicates of the triplet $(y, \Delta, \boldsymbol{x})$. In addition, we assume at the outset that the lowest quantile of interest $\tau_L$ satisfies $\mathbb{P}\{y \leq \boldsymbol{x}^{\mathrm{T}}\boldsymbol{\beta}^*(\tau_L), \Delta = 0\} = 0$. This condition, interpreted as no censoring below the $\tau_L$th quantile, is commonly imposed in the context of CQR; see, for example, Condition C in [45] and Assumption 3.1 in [67]. Moreover, our quantiles of interest are confined up to $\tau_U < 1$ subject to some identifiability concerns, which is a subtle issue for CQR problems. Briefly speaking, the model (1) may become nonidentifiable as $\tau$ moves toward 1, due to large amount of censored information in the upper tail. In practice, determining $\tau_U$ is usually a compromise between inference range of interest and data censoring rate, and $\tau_L$ can be chosen to be close to 0 if censoring occurs at early stages. Theoretically, the above assumption on $\tau_L$ helps us simplify the technical analysis.

The above model is broadly defined, yet it is inspired by approaching survival data with quantile regression [35]. To briefly illustrate, let $T$ be a nonnegative random variable representing the failure time to an event. The conditional quantile model (1) on $z = \log(T)$ can be viewed as a generalization of the standard AFT model in the sense that coefficients not only shift the location but also affect the shape and dispersion of the conditional distributions.

2.1. *Martingale-based estimating equation estimator.* Under the global linear model (1), two well-known methods are the recursively reweighted estimator of [45] and the stochastic integral based estimating equation estimator of [44]. Both methods are grid-based algorithms that iteratively solve a sequence of (weighted) check function minimization problems over a predetermined grid of $\tau$-values. Motivated by the recent success of smoothing methods for uncensored quantile regressions [18, 23, 54], we propose a smoothed estimating equation approach for CQR in the next subsection. We start with a brief introduction of [44]'s method that is built upon the martingale structure of randomly censored data.

To this end, denote by $\Lambda_{z|\boldsymbol{x}}(t) = -\log\{1 - \mathbb{P}(z \le t|\boldsymbol{x})\}$ the cumulative conditional hazard function of $z$ given $\boldsymbol{x}$, and define the counting processes $N_i(t) = \mathbb{1}\{y_i \le t, \Delta_i = 1\}$ and $N_{0i}(t) = \mathbb{1}\{y_i \le t, \Delta_i = 0\}$ for $i = 1, \ldots, n$, where $\Delta_i = \mathbb{1}(z_i \le C_i)$. Define $\mathcal{F}_i(s) = \sigma\{N_i(u), N_{0i}(u) : u \le s\}$ as the $\sigma$-algebra generated by the foregoing processes. Note that $\{\mathcal{F}_i(s) : s \in \mathbb{R}\}$ is an increasing family of sub-$\sigma$-algebras, also known as filtration, and $N_i(t)$ is an adapted submartingale. By the unique Doob–Meyer decomposition, one can construct an $\mathcal{F}_i(t)$-martingale $M_i(t) = N_i(t) - \Lambda_{z_i|\boldsymbol{x}_i}(y_i \wedge t)$ satisfying $\mathbb{E}\{M_i(t)|\boldsymbol{x}_i\} = 0$; see Section 1.3 of [19] for details. Taking $t = \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}^*(\tau)$ for each $i$, the martingale property implies

$$\mathbb{E}\Bigg[\sum_{i=1}^{n}\big\{N_i\big(\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}^*(\tau)\big) - \Lambda_{z_i|\boldsymbol{x}_i}\big(y_i \wedge \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}^*(\tau)\big)\big\}\boldsymbol{x}_i\Bigg] = \mathbf{0}.$$

This lays the foundation for the stochastic integral based estimating equation approach. The monotonicity of the function $\tau \mapsto \boldsymbol{x}^{\mathrm{T}}\boldsymbol{\beta}^*(\tau)$, implied by the global linearity in (1), leads to

$$\Lambda_{z_i|\boldsymbol{x}_i}\big(y_i \wedge \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}^*(\tau)\big) = H(\tau) \wedge H\big(\mathbb{P}(z_i \le y_i|\boldsymbol{x}_i)\big) = \int_0^{\tau}\mathbb{1}\big\{y_i \ge \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}^*(u)\big\}\,\mathrm{d}H(u)$$

for $\tau \in [\tau_L, \tau_U]$, where $H(u) := -\log(1 - u)$ for $0 < u < 1$. This motivates Peng and Huang's estimator [44], which solves the following estimating equation:

$$\frac{1}{n}\sum_{i=1}^{n}\Bigg[N_i\big(\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}(\tau)\big) - \int_0^{\tau}\mathbb{1}\big\{y_i \ge \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}(u)\big\}\,\mathrm{d}H(u)\Bigg]\boldsymbol{x}_i = \mathbf{0}, \quad \text{for every } \tau_L \le \tau \le \tau_U.$$

However, the exact solution to the above equation is not directly obtainable. By adapting Euler's forward method for an ordinary differential equation, [44] proposed a grid-based sequential estimating procedure as follows. Let $\tau_L = \tau_0 < \tau_1 < \cdots < \tau_m = \tau_U$ be a grid of quantile indices. Noting that $\mathbb{P}\{y \le \boldsymbol{x}^{\mathrm{T}}\boldsymbol{\beta}^*(\tau_0), \Delta = 0\} = 0$, we have $\mathbb{E}\int_0^{\tau_0}\mathbb{1}\{y_i \ge \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}^*(u)\}\,\mathrm{d}H(u) = \tau_0$, and hence $\boldsymbol{\beta}^*(\tau_0)$ can be estimated by solving the usual quantile equation $(1/n)\sum_{i=1}^{n}\{N_i(\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}) - \tau_0\}\boldsymbol{x}_i = \mathbf{0}$. Denote $\widetilde{\boldsymbol{\beta}}(\tau_0)$ as the solution to the above equation. At grid points $\tau_k$, $k = 1, \ldots, m$, the estimators $\widetilde{\boldsymbol{\beta}}(\tau_k)$ are sequentially obtained by solving

$$\frac{1}{n}\sum_{i=1}^{n}\Bigg[N_i\big(\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}\big) - \sum_{j=0}^{k-1}\int_{\tau_j}^{\tau_{j+1}}\mathbb{1}\big\{y_i \ge \boldsymbol{x}_i^{\mathrm{T}}\widetilde{\boldsymbol{\beta}}(\tau_j)\big\}\,\mathrm{d}H(u) - \tau_0\Bigg]\boldsymbol{x}_i = \mathbf{0}. \tag{2}$$

The resulting estimated function $\widetilde{\boldsymbol{\beta}}(\cdot) : [\tau_L, \tau_U] \mapsto \mathbb{R}^p$ is right continuous and piecewise constant that jumps only at each grid point. Computationally, solving the above equation is equivalent to minimizing an $\ell_1$-type convex objective function after introducing a sufficiently large pseudo point to the data. The minimizer, however, is not always uniquely defined. To avoid this lack of uniqueness as well as grid dependence, [28] introduced a more general (population) integral equation, and then proposed a Progressive Localized Minimization (PLMIN) algorithm to solve its empirical version exactly. This algorithm automatically determines the breakpoints of the solution, and thus is grid-free. Under a continuity condition on the density functions (see, e.g., condition (C2) in [28]), the estimating functions used in [44] and [28] are asymptotically equivalent.

2.2. *A smoothed estimating equation approach.* Due to the discontinuity stemming from the indicator function in the counting process $N_i(\cdot)$, exact solutions to the estimating equations (2) may not exist. In fact, $\widetilde{\boldsymbol{\beta}}(\tau_j)$ for $j = 0, \ldots, m$ are defined as the general solutions to generalized estimating equations [20], which correspond to subgradients of some convex yet nondifferentiable functions. Computationally, one may reformulate these equations as a sequence of linear programs, solvable by the Frisch–Newton algorithm described in [46]. The

computation complexity grows rapidly when the dimensionality $p$ increases. To mitigate the computational burden of the existing methods, we employ a smoothed estimating equation (SEE) approach for fitting large-scale censored quantile regression models.

Let $K(\cdot)$ be a symmetric and nonnegative kernel function and let $\bar{K}(u) = \int_{-\infty}^{u} K(x)\,\mathrm{d}x$, which is a nondecreasing function that is between 0 and 1. The nonsmooth indicator function $\mathbb{1}(u \geq 0)$ can thus be approximated by $\bar{K}(u/h)$ for some $h > 0$ in the sense that as $h \to 0$, $\bar{K}(u/h) \to 1$ for $u \geq 0$ and $\bar{K}(u/h) \to 0$ for $u < 0$. Hereinafter, $h > 0$ will be referred to as a bandwidth. As aforementioned, let $\tau_L = \tau_0 < \tau_1 < \cdots < \tau_m = \tau_U$ be a grid of quantile indices for some $m \geq 1$. Given a kernel function $K(\cdot)$ and a bandwidth $h > 0$, write

$$K_h(u) = h^{-1}K(u/h) \quad \text{and} \quad \bar{K}_h(u) = \bar{K}(u/h) = \int_{-\infty}^{u/h} K(v)\mathrm{d}v, \quad u \in \mathbb{R},$$

so that $\bar{K}_h'(u) = K_h(u)$. We now propose a smooth SEE approach for CQR.

1. At $\tau = \tau_0$, we estimate $\boldsymbol{\beta}^*(\tau_0)$ by $\widehat{\boldsymbol{\beta}}(\tau_0)$, obtained from solving $\widehat{Q}_0(\boldsymbol{\beta}) = \mathbf{0}$, where

$$\widehat{Q}_0(\boldsymbol{\beta}) := \frac{1}{n}\sum_{i=1}^{n}\{\Delta_i \bar{K}_h(-r_i(\boldsymbol{\beta})) - \tau_0\}\boldsymbol{x}_i \quad \text{and} \quad r_i(\boldsymbol{\beta}) = y_i - \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}. \tag{3}$$

2. At grid points $\tau_k$ for $k = 1, \ldots, m$, set $\widehat{\boldsymbol{\beta}}(\tau) = \widehat{\boldsymbol{\beta}}(\tau_{k-1})$ for any $\tau \in (\tau_{k-1}, \tau_k)$, and then obtain estimators $\widehat{\boldsymbol{\beta}}(\tau_k)$ of $\boldsymbol{\beta}^*(\tau_k)$ by solving $\widehat{Q}_k(\boldsymbol{\beta}) = \mathbf{0}$, where

$$\widehat{Q}_k(\boldsymbol{\beta}) := \frac{1}{n}\sum_{i=1}^{n}\left[\Delta_i \bar{K}_h(-r_i(\boldsymbol{\beta})) - \sum_{j=0}^{k-1}\bar{K}_h(r_i(\widehat{\boldsymbol{\beta}}(\tau_j))\{H(\tau_{j+1}) - H(\tau_j)\} - \tau_0\right]\boldsymbol{x}_i. \tag{4}$$

Note that the resulting estimator $\widehat{\boldsymbol{\beta}}(\cdot) : [\tau_L, \tau_U] \mapsto \mathbb{R}^p$ is right continuous and piecewise constant with jumps only at grids. For notational convenience, throughout the remainder of this paper we write

$$\boldsymbol{\beta}_k^* = \boldsymbol{\beta}^*(\tau_k) \quad \text{and} \quad \widehat{\boldsymbol{\beta}}_k = \widehat{\boldsymbol{\beta}}(\tau_k), \quad k = 0, 1, \ldots, m.$$

Before proceeding, it is worth noticing that the above smoothed estimating equations method is closely related to the convolution smoothing approach studied in [18] and [23]. Consider the check function $\rho_\tau(u) = \tau\{u - \mathbb{1}(u < 0)\}$, and its convolution smoothed counterpart

$$\ell_{\tau,h}(u) = (\rho_\tau * K_h)(u) = \int_{-\infty}^{\infty}\rho_\tau(v)K_h(v-u)\,\mathrm{d}v,$$

where $*$ denotes the convolution operator. Given censored data $\{(y_i, \Delta_i, \boldsymbol{x}_i)\}_{i=1}^n$, define the empirical smoothed loss

$$\widehat{L}_0(\boldsymbol{\beta}) = \frac{1}{n}\sum_{i=1}^{n}\{\Delta_i \ell_{\tau_0,h}(y_i - \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}) + \tau_0(\Delta_i - 1)\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}\}, \tag{5}$$

whose gradient and Hessian are

$$\nabla\widehat{L}_0(\boldsymbol{\beta}) = \widehat{Q}_0(\boldsymbol{\beta}) \quad \text{and} \quad \nabla^2\widehat{L}_0(\boldsymbol{\beta}) = \frac{1}{n}\sum_{i=1}^{n}\Delta_i K_h(r_i(\boldsymbol{\beta}))\boldsymbol{x}_i\boldsymbol{x}_i^{\mathrm{T}},$$

respectively. Hence, the foregoing estimator $\widehat{\boldsymbol{\beta}}_0$ can be equivalently defined as the solution to the (unconstrained) optimization problem $\min_{\boldsymbol{\beta}\in\mathbb{R}^p}\widehat{L}_0(\boldsymbol{\beta})$. When a nonnegative kernel is used, the objective function $\widehat{L}_0(\cdot)$ is convex, and thus any minimizer satisfies the first-order

condition. At subsequent grid points $\tau_k$ for $k = 1, \ldots, m$, the estimator $\widehat{\boldsymbol{\beta}}_k$ can also be viewed as an $M$-estimator that solves

$$\min_{\boldsymbol{\beta}\in\mathbb{R}^p} \left\{ \widehat{L}_k(\boldsymbol{\beta}) := \widehat{L}_0(\boldsymbol{\beta}) - \left\langle \frac{1}{n}\sum_{i=1}^{n}\sum_{j=0}^{k-1} \bar{K}_h\big(y_i - \boldsymbol{x}_i^{\mathrm{T}}\widehat{\boldsymbol{\beta}}_j\big)\{H(\tau_{j+1}) - H(\tau_j)\}\boldsymbol{x}_i, \boldsymbol{\beta}\right\rangle \right\}. \tag{6}$$

Notably, kernel smoothing produces continuously differentiable estimating functions $\widehat{Q}_k(\cdot)$ $(k = 0, \ldots, m)$, or equivalently, convex and twice-differentiable loss functions $\widehat{L}_k(\cdot)$, which have the same positive semidefinite Hessian matrix $\nabla^2\widehat{L}_k(\boldsymbol{\beta}) = (1/n)\sum_{i=1}^n \Delta_i K_h(\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta} - y_i)\boldsymbol{x}_i\boldsymbol{x}_i^{\mathrm{T}}$. As we shall see, the empirical loss functions $\widehat{L}_k(\cdot)$ are not only globally convex but also locally strongly convex (with high probability). This property ensures the existence of global solutions to the sequential estimation problems, which can be efficiently solved by a quasi-Newton algorithm described in Section A.1 of the Supplementary Material [24].

2.3. *Inference with bootstrapped process*. In this subsection, we construct componentwise confidence intervals for $\boldsymbol{\beta}^*(\tau)$ at some quantile index $\tau$ of interest by bootstrapping the quantile process. Recall that $\widehat{\boldsymbol{\beta}}_k$'s are the solutions to the equations $\widehat{Q}_k(\boldsymbol{\beta}) = \mathbf{0}$, where $\widehat{Q}_k(\cdot)$ $(k = 0, 1, \ldots, m)$ are defined in (3) and (4). Analogously, we construct bootstrap estimators $\widehat{\boldsymbol{\beta}}_k^\flat$ following a sequential procedure based on the bootstrapped SEEs obtained by perturbing $\widehat{Q}_k(\cdot)$ with random weights. Independent of the observed data $\{y_i, \Delta_i, \boldsymbol{x}_i\}_{i=1}^n$, let $W_1, \ldots, W_n$ be exchangeable nonnegative random variables, satisfying $\mathbb{E}(W_i) = 1$ and $\operatorname{var}(W_i) > 0$. The bootstrap estimators can be constructed as follows:

1. Set $\widehat{\boldsymbol{\beta}}_0^\flat$ as the solution of $\widehat{Q}_0^\flat(\boldsymbol{\beta}) = \mathbf{0}$, where

$$\widehat{Q}_0^\flat(\boldsymbol{\beta}) := \frac{1}{n}\sum_{i=1}^{n} W_i\big\{\Delta_i \bar{K}_h\big(-r_i(\boldsymbol{\beta})\big) - \tau_0\big\}\boldsymbol{x}_i \quad \text{with } r_i(\boldsymbol{\beta}) = y_i - \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}. \tag{7}$$

2. For $k = 1, \ldots, m$, compute $\widehat{\boldsymbol{\beta}}_k^\flat$ sequentially by solving $\widehat{Q}_k^\flat(\boldsymbol{\beta}) = \mathbf{0}$, where

$$\widehat{Q}_k^\flat(\boldsymbol{\beta}) := \frac{1}{n}\sum_{i=1}^{n} W_i\left[\Delta_i \bar{K}_h\big(-r_i(\boldsymbol{\beta})\big) - \sum_{\ell=0}^{k-1} \bar{K}_h\big(r_i\big(\widehat{\boldsymbol{\beta}}_\ell^\flat\big)\big)\{H(\tau_{\ell+1} - H(\tau_\ell)\} - \tau_0\right]\boldsymbol{x}_i. \tag{8}$$

3. Define the bootstrap estimate of the coefficient process $\widehat{\boldsymbol{\beta}}^\flat(\cdot) : [\tau_L, \tau_U] \mapsto \mathbb{R}^p$ as $\widehat{\boldsymbol{\beta}}^\flat(\tau) = \widehat{\boldsymbol{\beta}}_{k-1}^\flat$ for $\tau \in [\tau_{k-1}, \tau_k)$ and $k = 1, \ldots, m$.

For a prescribed nominal level, we can construct componentwise percentile or normal-based confidence intervals for $\boldsymbol{\beta}_j^*(\tau)$ $(j = 1, \ldots, p)$. The above multiplier bootstrap estimator $\widehat{\boldsymbol{\beta}}^\flat(\cdot) : [\tau_L, \tau_U] \mapsto \mathbb{R}^p$ of the coefficient process behaves similarly as $\widehat{\boldsymbol{\beta}}(\cdot)$, in the sense that they are both right continuous and piecewise constant with jumps only at the grids. The multiplier bootstrap method, which dates back at least to [2], is motivated by the following simple yet important observation. Let $\mathbb{E}^*(\cdot)$ be conditional expectation given the data, that is, $\mathbb{E}^*(\cdot) = \mathbb{E}(\cdot|\{y_i, \Delta_i, \boldsymbol{x}_i\}_{i=1}^n)$. Since $\mathbb{E}(W_i) = 1$, we have $\mathbb{E}^*\{\widehat{Q}_0^\flat(\boldsymbol{\beta})\} = \widehat{Q}_0(\boldsymbol{\beta})$ and $\mathbb{E}^*\{\widehat{Q}_k^\flat(\boldsymbol{\beta})\} \approx \widehat{Q}_k(\boldsymbol{\beta})$ for $k = 1, \ldots, m$. This means that in the bootstrap world, $\widehat{Q}_k^\flat(\boldsymbol{\beta})$ can be viewed as an empirical version of $\widehat{Q}_k(\boldsymbol{\beta})$, and thus $\widehat{\boldsymbol{\beta}}_k^\flat$ can be regarded as the bootstrap estimator of $\widehat{\boldsymbol{\beta}}_k$.

We complete this section with a brief discussion of other resampling methods for quantile regression. Given the random weights $\{W_i\}_{i=1}^n$ independent of data, another available approach is to minimize the randomly perturbed objective functions [29, 44]. In the current

setting, it seems more natural to directly bootstrap the estimating equations. In terms of bootstrapping estimating equations with uncensored data, [41]'s method is based on the assumption that the estimating equation is exactly or asymptotically pivotal, and [26]'s proposal is based on resampling with replacement. A generalized weighted bootstrap and its asymptotic theory has been rigorously studied in [8] and [39]. For censored quantile regression, the sequential SEEs (4) are not directly formulated as empirical averages of independent random quantities, nor do they satisfy the required assumptions in the literature; see Section 2 of [41], Section 2 of [26] and Section 3 of [8]. Hence, the validity of weighted bootstrap for CQR is of independent interest, and will be examined in Section 3.3.

REMARK 2.1. In practice, random weights $\{W_i\}_{i=1}^n$ can be generated from one of the following distributions: (i) $(W_1, \ldots, W_n) \sim \text{Multinomial}(n, 1/n, \ldots, 1/n)$. This leads to Efron's nonparametric bootstrap, for which the random weights are exchangeable but not independent; (ii) $W_1, \ldots, W_n \sim \text{Exp}(1)$ are i.i.d. exponentially distributed random variables; and (iii) $W_i = e_i + 1$, where $e_i$'s are i.i.d. Rademacher random variables, defined by $\mathbb{P}(e_i = 1) = \mathbb{P}(e_i = 0) = 1/2$. We refer to this as the Rademacher multiplier bootstrap. Its theoretical properties will be investigated in Section 3.3.

## 3. Theoretical analysis.

3.1. *Regularity conditions*. We first impose some technical assumptions required for the results in Sections 3.2 and 3.3.

CONDITION 3.1 (Kernel function). Let $K(\cdot)$ be a symmetric, Lipschitz continuous and nonnegative kernel function, that is, $K(u) = K(-u)$, $K(u) \geq 0$ for all $u \in \mathbb{R}$ and $\int_{-\infty}^{\infty} K(u)\,\mathrm{d}u = 1$. Moreover, $\kappa_u = \sup_{u\in\mathbb{R}} K(u) < \infty$, $\kappa_l = \min_{|u|\leq c} K(u) > 0$ for some $c > 0$. We define its higher-order absolute moments as $\kappa_\ell = \int_{-\infty}^{\infty} |u|^\ell K(u)\mathrm{d}u$ for any positive integer $\ell$.

CONDITION 3.2 (Random design). The random covariate vector $\boldsymbol{x} = (x_1, \ldots, x_p)^{\mathrm{T}} \in \mathcal{X} \subseteq \mathbb{R}^p$ is compactly supported with $\zeta_p := \sup_{\boldsymbol{x}\in\mathcal{X}} \|\Sigma^{-1/2}\boldsymbol{x}\|_2 < \infty$, where $\Sigma = \mathbb{E}(\boldsymbol{x}\boldsymbol{x}^{\mathrm{T}})$ is positive definite.

CONDITION 3.3 (Conditional densities). Assume $(z, \boldsymbol{x})$ follows the global conditional quantile model (1). Define the conditional cumulative distribution functions $F_z(u|\boldsymbol{x}) = \mathbb{P}(z \leq u|\boldsymbol{x})$, $F_y(u|\boldsymbol{x}) = \mathbb{P}(y \leq u|\boldsymbol{x})$ and $G(u|\boldsymbol{x}) = \mathbb{P}(y \leq u, \Delta = 1|\boldsymbol{x})$, where $y = z \wedge C$ and $C$ is independent of $z$ given $\boldsymbol{x}$. Assume that the conditional densities $f_z(u|\boldsymbol{x}) = F_z'(u|\boldsymbol{x})$, $f_y(u|\boldsymbol{x}) = F_y'(u|\boldsymbol{x})$ and $g(u|\boldsymbol{x}) = G'(u|\boldsymbol{x})$ exist, and satisfy almost surely (over $\boldsymbol{x}$) that

$$\inf_{\tau\in[\tau_L,\tau_U]} \min\{f_y(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\beta}^*(\tau)|\boldsymbol{x}), f_z(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\beta}^*(\tau)|\boldsymbol{x})\} \geq \underline{f} > 0, \qquad \sup_{u\in\mathbb{R}} f_y(u|\boldsymbol{x}) \leq \overline{f},$$

$$0 < \underline{g} \leq \inf_{|u-\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\beta}^*(\tau)|\leq 1/2, \tau\in[\tau_L,\tau_U]} g(u|\boldsymbol{x}) \leq \sup_{u\in\mathbb{R}} g(u|\boldsymbol{x}) \leq \overline{g}.$$

Moreover, there exists a constant $l_1 > 0$ such that for any $u \in \mathbb{R}$,

$$\sup_{\boldsymbol{x}\in\mathbb{R}^p, \tau\in[\tau_L,\tau_U]} \left|f_y(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\beta}^*(\tau) + u|\boldsymbol{x}) - f_y(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\beta}^*(\tau)|\boldsymbol{x})\right| \leq l_1|u|,$$

$$\sup_{\boldsymbol{x}\in\mathbb{R}^p, \tau\in[\tau_L,\tau_U]} \left|g(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\beta}^*(\tau) + u|\boldsymbol{x}) - g(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\beta}^*(\tau)|\boldsymbol{x})\right| \leq l_1|u|.$$

CONDITION 3.4 (Grid size). The grid of quantile levels $\tau_L = \tau_0 < \tau_1 < \cdots < \tau_m = \tau_U$ satisfies $n^{-1} \le \delta_* \le \delta^* \lesssim n^{-1/2}$, where $\delta^* = \max_{1\le k\le m}(\tau_k - \tau_{k-1})$ and $\delta_* = \min_{1\le k\le m}(\tau_k - \tau_{k-1})$.

Condition 3.1 holds for most commonly used kernel functions, including: (a) uniform kernel $K(u) = (1/2)\mathbb{1}(|u| \le 1)$, (b) Gaussian kernel $K(u) = (2\pi)^{-1/2}e^{-u^2/2}$, (c) logistic kernel $K(u) = e^{-u}/(1+e^{-u})^2$, (d) Epanechnikov/parabolic kernel $K(u) = (3/4)(1-u^2)\mathbb{1}(|u| \le 1)$ and (e) triangular kernel $K(u) = (1-|u|)\mathbb{1}(|u| \le 1)$. To simplify the analysis, we take $c = 1$ in Condition 3.1; otherwise, if $c < 1$ and $K(\pm 1) = 0$, we can simply use a rescaled kernel $K_c(u) := cK(cu)$, so that $\min_{|u|\le 1} K_c(u) = c\min_{|u|\le c} K(u)$. The compactness of $\mathcal{X}$ in Condition 3.2 is a common requirement for a global linear quantile regression model (quantile regression process) [32]. If the support of the covariate space—the set of $x_j$'s that occur with positive probability—is unbounded, at some points there will be "crossings" of the conditional quantile functions, unless these functions are parallel, which corresponds to a pure location-shift model. The quantity $\zeta_p$ plays an important role in the theoretical results. Alternatively, one may assume $\|\Sigma^{-1/2}\boldsymbol{x}\|_\infty \le B_0$ (almost surely) as in [67], which in turn implies $\zeta_p \le B_0 p^{1/2}$ in the worst-case scenario. In general, it is reasonable to assume that $\zeta_p \asymp p^{1/2}$. In addition to $\zeta_p$, define the moment parameters

$$m_q = \sup_{\boldsymbol{u}\in\mathbb{S}^{p-1}} \mathbb{E}\big(|\boldsymbol{u}^{\mathrm{T}}\Sigma^{-1/2}\boldsymbol{x}|^q\big) \quad \text{for } q = 3, 4, \tag{9}$$

which satisfy the worst case bounds $m_3 \le \zeta_p$ and $m_4 \le \zeta_p^2$.

Conditions 3.2 and 3.3 ensure that the coefficient function $\boldsymbol{\beta}^*(\cdot)$ is Lipschitz continuous. Since $\boldsymbol{\beta}^*(\tau)$ solves the equation $\mathbb{E}[\{\tau - \mathbb{1}(z \le \boldsymbol{x}^{\mathrm{T}}\boldsymbol{\beta})\}\boldsymbol{x}] = \mathbf{0}$, we have $\frac{d}{d\tau}\boldsymbol{\beta}^*(\tau) = \mathbb{E}\{f_z(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\beta}^*(\tau)|\boldsymbol{x})\boldsymbol{x}\boldsymbol{x}^{\mathrm{T}}\}^{-1}\mathbb{E}(\boldsymbol{x})$. Under Condition 3.2, it holds

$$\max_{\tau\in[\tau_L,\tau_U]}\left\|\frac{d}{d\tau}\Sigma^{1/2}\boldsymbol{\beta}^*(\tau)\right\|_2 \le \underline{f}^{-1}\max_{\tau\in[\tau_L,\tau_U]}\big\|\mathbb{E}\big(\Sigma^{-1/2}\boldsymbol{x}\big)\big\|_2 \le \underline{f}^{-1},$$

which, together with the mean value theorem, implies

$$\big\|\boldsymbol{\beta}^*(\tau) - \boldsymbol{\beta}^*\big(\tau'\big)\big\|_\Sigma \le \underline{f}^{-1}|\tau - \tau'| \quad \text{for any } \tau, \tau' \in [\tau_L, \tau_U]. \tag{10}$$

By the definitions in Condition 3.3, $G(u|\boldsymbol{x}) \le F(u|\boldsymbol{x})$ for any $u > 0$. Recall that we have assumed no censored observations at low quantile levels $\tau \le \tau_L$. Hence, $G(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\beta}^*(\tau_L)|\boldsymbol{x}) = F(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\beta}^*(\tau_L)|\boldsymbol{x}) = \tau_L$, and $G(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\beta}^*(\tau)|\boldsymbol{x}) \le \tau \le F(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\beta}^*(\tau)|\boldsymbol{x})$ for $\tau_L < \tau \le \tau_U$. Condition 3.4 assures a fine grid by controlling the gap between two contiguous points, so that the approximation/discretization error does not exceed the statistical error.

3.2. *Uniform rate of convergence and Bahadur representation.* In this section, we characterize the statistical properties of the SEE estimators for censored quantile regression with growing dimensions. That is, the dimension $p = p_n$ is subject to the growth condition $p \asymp n^a$ for some $a \in (0, 1)$. Our first result provides the uniform rate of convergence for the estimated coefficient function $\widehat{\boldsymbol{\beta}}(\cdot)$ under mild bandwidth constraints.

THEOREM 3.1 (Uniform consistency). *Assume Conditions 3.1–3.4 hold, and choose the bandwidth $h = h_n \asymp \{(p + \log n)/n\}^\gamma$ for some $\gamma \in [1/4, 1/2)$. Further let $n \gtrsim \{\zeta_p^2(p + \log n)^{1/2-\gamma}\}^{1/(1-\gamma)}$. Then the SEE estimator $\widehat{\boldsymbol{\beta}}(\cdot) : [\tau_L, \tau_U] \mapsto \mathbb{R}^p$ satisfies*

$$\sup_{\tau\in[\tau_L,\tau_U]}\big\|\widehat{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}^*(\tau)\big\|_\Sigma \lesssim \left(\frac{1-\tau_L}{1-\tau_U}\right)^{C_0\overline{f}/\underline{g}}\underline{g}^{-1}\sqrt{\frac{p+\log n}{n}} \tag{11}$$

*with probability at least $1 - C_1 n^{-1}$, where $C_0, C_1 > 0$ are constants independent of $(n, p)$.*

Since the deviation bound in (11) depends explicitly on $n$, $p$ as well as other model parameters, this nonasymptotic result implies the classical asymptotic consistency by letting $n \to \infty$ with $p$ fixed. From an asymptotic perspective, Theorem 3.1 implies that the smoothed estimator with a bandwidth $h = h_n \asymp \{\log(n)/n\}^{\gamma}$ for some $\gamma \in [1/4, 1/2)$ satisfies $\sup_{\tau_L \le \tau \le \tau_U} \|\widehat{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}^*(\tau)\|_2 \to 0$ in probability as $n \to \infty$.

Recall that in the sequential estimation procedure described in Section 2.2, the $j$th estimator $\widehat{\boldsymbol{\beta}}_j$ $(j \ge 1)$ depends implicitly on its predecessors through the estimating function (4). In other words, the accumulative estimation errors of $\widehat{\boldsymbol{\beta}}(\tau)$ for $\tau_L \le \tau < \tau_j$ may have a nonnegligible impact on $\widehat{\boldsymbol{\beta}}_j = \widehat{\boldsymbol{\beta}}(\tau_j)$. The next result explicitly quantifies this accumulative error. For $\tau \in [\tau_L, \tau_U]$, define $p \times p$ matrices

$$\mathbf{J}(\tau) = \mathbb{E}\{g(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\beta}^*(\tau)|\boldsymbol{x})\boldsymbol{x}\boldsymbol{x}^{\mathrm{T}}\} \quad \text{and} \quad \mathbf{H}(\tau) = \mathbb{E}\{f_y(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\beta}^*(\tau)|\boldsymbol{x})\boldsymbol{x}\boldsymbol{x}^{\mathrm{T}}\}, \tag{12}$$

both of which are positive definite under Conditions 3.2 and 3.3. Moreover, define the integrated covariate effect and its estimate

$$\boldsymbol{\beta}^*_{\mathrm{int}}(\tau) := \mathbf{J}(\tau)\boldsymbol{\beta}^*(\tau) + \int_{\tau_L}^{\tau} \mathbf{H}(u)\boldsymbol{\beta}^*(u)\,\mathrm{d}H(u)$$

$$\text{and} \quad \widehat{\boldsymbol{\beta}}_{\mathrm{int}}(\tau) := \mathbf{J}(\tau)\widehat{\boldsymbol{\beta}}(\tau) + \int_{\tau_L}^{\tau} \mathbf{H}(u)\widehat{\boldsymbol{\beta}}(u)\,\mathrm{d}H(u),$$

respectively, so that $\widehat{\boldsymbol{e}}(\tau) := \widehat{\boldsymbol{\beta}}_{\mathrm{int}}(\tau) - \boldsymbol{\beta}^*_{\mathrm{int}}(\tau)$ can be interpreted as the accumulated error in the sequential estimation procedure up to $\tau$. That is,

$$\widehat{\boldsymbol{e}}(\tau) = \underbrace{\mathbf{J}(\tau)\{\widehat{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}^*(\tau)\}}_{\text{current step}} + \underbrace{\int_{\tau_L}^{\tau} \mathbf{H}(u)\{\widehat{\boldsymbol{\beta}}(u) - \boldsymbol{\beta}^*(u)\}\,\mathrm{d}H(u)}_{\text{preceding steps}}. \tag{13}$$

The following theorem provides a uniform Bahadur representation for $\widehat{\boldsymbol{e}}(\cdot)$.

THEOREM 3.2 (Uniform Bahadur representation). *Assume that the same set of conditions in Theorem 3.1 hold. Moreover, assume $\delta^* \asymp n^{-(1/2+\alpha)}$ for some $\alpha \in (0, 1/2)$. Then the SEE estimator $\widehat{\boldsymbol{\beta}}(\cdot) : [\tau_L, \tau_U] \mapsto \mathbb{R}^p$ satisfies*

$$\widehat{\boldsymbol{e}}(\tau) = \widehat{\boldsymbol{\beta}}_{\mathrm{int}}(\tau) - \boldsymbol{\beta}^*_{\mathrm{int}}(\tau) = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{U}_i(\tau) + \boldsymbol{r}_n(\tau), \tag{14}$$

*where*

$$\boldsymbol{U}_i(\tau) := \left\{\tau_L + \int_{\tau_L}^{\tau} \bar{K}_h\big(y_i - \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}^*(u)\big)\,\mathrm{d}H(u) - \Delta_i \bar{K}_h\big(\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}^*(\tau) - y_i\big)\right\}\boldsymbol{x}_i \tag{15}$$

*satisfies $\sup_{\tau\in[\tau_L,\tau_U]} \|\mathbb{E}\boldsymbol{U}_i(\tau)\|_{\Sigma^{-1}} \lesssim h^2$, and the remainder process $\boldsymbol{r}_n(\cdot) : [\tau_L, \tau_U] \mapsto \mathbb{R}^p$ is such that*

$$\sup_{\tau\in[\tau_L,\tau_U]} \|\boldsymbol{r}_n(\tau)\|_{\Sigma^{-1}} \lesssim m_4^{1/2}\frac{p+\log n}{nh^{1/2}} + m_3\frac{p+\log n}{n} + h\sqrt{\frac{p+\log n}{n}} + n^{-1/2-\alpha} \tag{16}$$

*with probability at least $1 - C_2 n^{-1}$ for some absolute constant $C_2 > 0$, where $m_q$ $(q = 3, 4)$ are given in (9).*

REMARK 3.1. Together, the above uniform Bahadur representation and the production integration theory [21] establish the asymptotic distribution of $\widehat{\boldsymbol{\beta}}(\cdot)$. Define

$$\boldsymbol{\theta}^*(\tau) = \mathbf{J}(\tau)\boldsymbol{\beta}^*(\tau), \qquad \widehat{\boldsymbol{\theta}}(\tau) = \mathbf{J}(\tau)\widehat{\boldsymbol{\beta}}(\tau) \quad \text{and}$$

$$\boldsymbol{\Psi}(\tau) = \frac{1}{1-\tau}\mathbf{H}(\tau)\mathbf{J}(\tau)^{-1}, \quad \tau \in [\tau_L, \tau_U].$$

Then equation (13) reads $\widehat{\boldsymbol{e}}(\tau) = \widehat{\boldsymbol{\theta}}(\tau) - \boldsymbol{\theta}^*(\tau) + \int_{\tau_L}^{\tau} \boldsymbol{\Psi}(u)\{\widehat{\boldsymbol{\theta}}(u) - \boldsymbol{\theta}^*(u)\}\mathrm{d}u$. Combined with Theorem 3.2, this implies

$$
\begin{aligned}
&n^{1/2}\{\widehat{\boldsymbol{\theta}}(\tau) - \boldsymbol{\theta}^*(\tau)\} + \int_{\tau_L}^{\tau} \boldsymbol{\Psi}(u) n^{1/2}\{\widehat{\boldsymbol{\theta}}(u) - \boldsymbol{\theta}^*(u)\}\,\mathrm{d}u \\
&\quad = \frac{1}{n^{1/2}} \sum_{i=1}^{n} \{\boldsymbol{U}_i(\tau) - \mathbb{E}\boldsymbol{U}_i(\tau)\} + \bar{\boldsymbol{r}}_n(\tau), \quad \tau \in [\tau_L, \tau_U],
\end{aligned} \tag{17}
$$

where the rescaled remainder $\bar{\boldsymbol{r}}_n(\cdot)$ satisfies $\sup_{\tau\in[\tau_L,\tau_U]} \|\bar{\boldsymbol{r}}_n(\tau)\|_2 = o_{\mathbb{P}}(1)$, with a properly chosen bandwidth that will be discussed in Remark 3.2. Note that equation (17) is a stochastic differential equation for $n^{1/2}\{\widehat{\boldsymbol{\theta}}(\tau) - \boldsymbol{\theta}^*(\tau)\}$ [44]. From the classical production integration theory ([21] and Section II.6 of [1]), it follows that

$$
n^{1/2}\{\widehat{\boldsymbol{\theta}}(\tau) - \boldsymbol{\theta}^*(\tau)\} = \boldsymbol{\phi}\left(\frac{1}{n^{1/2}} \sum_{i=1}^{n} \{\boldsymbol{U}_i(\tau) - \mathbb{E}\boldsymbol{U}_i(\tau)\}\right) + o_{\mathbb{P}}(1), \tag{18}
$$

where $\boldsymbol{\phi}$ is a linear operator from $\mathcal{F}$ to $\mathcal{F}$ defined as

$$
\boldsymbol{\phi}(\boldsymbol{g})(\tau) = \Pi_{u\in[\tau_L,\tau]}\{\mathbf{I}_p - \boldsymbol{\Psi}(u)\mathrm{d}u\}\boldsymbol{g}(\tau_L) + \int_{\tau_L}^{\tau} \Pi_{u\in(s,\tau]}\{\mathbf{I}_p - \boldsymbol{\Psi}(u)\mathrm{d}u\}\,\mathrm{d}\boldsymbol{g}(s) \tag{19}
$$

for $\boldsymbol{g} \in \mathcal{F} := \{\boldsymbol{f} : [\tau_L, \tau_U] \to \mathbb{R}^p | \boldsymbol{f} \text{ is left continuous with right limit}\}$, and $\Pi$ denotes the product-limit; see Definition 1 in [21]. After careful proofreading, we believe that the above form of $\boldsymbol{\phi}(\cdot)$ corrects an error (possibly a typo) in the proof of Theorem 2 in [44]; see the arguments between (B.1) and (B.3) therein. Specifically, the linear operator $\boldsymbol{\phi}$ in [44] reads

$$
\boldsymbol{\phi}(\boldsymbol{g})(\tau) = \Pi_{u\in[\tau_L,\tau]}\{\mathbf{I}_p + \boldsymbol{\Psi}(u)\mathrm{d}u\}\boldsymbol{g}(\tau_L) + \int_{\tau_L}^{\tau} \Pi_{u\in(s,\tau]}\{\mathbf{I}_p + \boldsymbol{\Psi}(u)\mathrm{d}u\}\,\mathrm{d}\boldsymbol{g}(s).
$$

The asymptotic distribution of $n^{1/2}\{\widehat{\boldsymbol{\theta}}(\tau) - \boldsymbol{\theta}^*(\tau)\}$ or its linear functional is thus determined by that of

$$
\boldsymbol{\phi}\left(\frac{1}{n^{1/2}} \sum_{i=1}^{n} \{\boldsymbol{U}_i(\tau) - \mathbb{E}\boldsymbol{U}_i(\tau)\}\right) \quad \text{and} \quad \frac{1}{n^{1/2}} \sum_{i=1}^{n} \{\boldsymbol{U}_i(\tau) - \mathbb{E}\boldsymbol{U}_i(\tau)\}.
$$

REMARK 3.2 (Order of bandwidth). We further discuss the order of bandwidth $h$, as a function of $(n, p)$, required in Theorem 3.2 and Remark 3.1. Following (17), if the moment parameters $m_3$ (absolute skewness) and $m_4$ (kurtosis) are dimension-free, the Bahadur linearization remainder $\bar{\boldsymbol{r}}_n(\cdot)$ satisfies with high probability that $\sup_{\tau\in[\tau_L,\tau_U]} \|\bar{\boldsymbol{r}}_n(\tau)\|_{\Sigma^{-1}} \lesssim n^{1/2}h^2 + (p + \log n)/(nh)^{1/2} + n^{-\alpha}$. Set the bandwidth $h \asymp \{(p + \log n)/n\}^{\gamma}$ for some $\gamma \in [1/4, 1/2)$, this implies

$$
\sup_{\tau\in[\tau_L,\tau_U]} \|\bar{\boldsymbol{r}}_n(\tau)\|_{\Sigma^{-1}} \lesssim \frac{(p + \log n)^{2\gamma}}{n^{2\gamma-1/2}} + \frac{(p + \log n)^{1-\gamma/2}}{n^{1/2-\gamma/2}} + \frac{1}{n^{\alpha}} = o_{\mathbb{P}}(1),
$$

provided that $p = o(n^{1-1/(4\gamma)} \wedge n^{(1-\gamma)/(2-\gamma)})$. In particular, letting $1 - 1/(4\gamma) = (1 - \gamma)/(2 - \gamma)$ yields $\gamma = 2/5$. We therefore choose the bandwidth $h \asymp \{(p + \log n)/n\}^{2/5}$, so that all the asymptotic results (from uniform rate of convergence to Bahadur representation) hold under the growth condition $p = o(n^{3/8})$ of dimensionality $p$ in sample size $n$.

Theorem 3.2 explicitly characterizes the leading term of the integrated estimation error (13), along with a high probability bound on the remainder process. As discussed in Remark 3.1, the asymptotic distributions of $n^{1/2}\{\widehat{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}^*(\tau)\}$ or its linear functional can be established based on the stochastic integral representation (18), which further depends on the

centered random process $n^{-1/2}\sum_{i=1}^{n}\{\boldsymbol{U}_i(\cdot)-\mathbb{E}\boldsymbol{U}_i(\cdot)\}$. Let $\{\boldsymbol{a}_n\}_{n=1}^{\infty}$ be a sequence of deterministic vectors in $\mathbb{R}^p$, and define

$$\mathbb{G}_n(\tau):=\frac{1}{n^{1/2}}\sum_{i=1}^{n}\langle \boldsymbol{a}_n/\|\boldsymbol{a}_n\|_{\Sigma},\boldsymbol{U}_i(\tau)-\mathbb{E}\boldsymbol{U}_i(\tau)\rangle,\quad \tau\in[\tau_L,\tau_U]. \tag{20}$$

The asymptotic behavior of $\{\mathbb{G}_n(\tau):\tau\in[\tau_L,\tau_U]\}$ is provided in the following result.

THEOREM 3.3 (Weak convergence). *Assume Conditions 3.1–3.4 hold with* $\delta^*\asymp n^{-(1/2+\alpha)}$ *for some* $\alpha\in(0,1/2)$. *Moreover, assume* $h\asymp\{(p+\log n)/n\}^{2/5}$ *and* $p=o(n^{3/8})$ *as* $n\to\infty$. *For any deterministic sequence of vectors* $\{\boldsymbol{a}_n\}_{n\geq 1}$, *if the following limit*

$$H(\tau,\tau'):=\lim_{n\to\infty}\frac{1}{\|\boldsymbol{a}_n\|_{\Sigma}^2}\boldsymbol{a}_n^{\mathrm{T}}\mathbb{E}\{\boldsymbol{U}_i(\tau)\boldsymbol{U}_i(\tau')^{\mathrm{T}}\}\boldsymbol{a}_n \tag{21}$$

*exists for any* $\tau,\tau'\in[\tau_L,\tau_U]$ *with* $\boldsymbol{U}_i(\cdot)$ *defined in* (15), *then*

$$\mathbb{G}_n(\cdot)\rightsquigarrow\mathbb{G}(\cdot)\quad \text{in } \ell^{\infty}([\tau_L,\tau_U]), \tag{22}$$

*where* $\mathbb{G}_n(\cdot)$ *is given in* (20), *and* $\mathbb{G}(\cdot)$ *is a tight zero-mean Gaussian process with covariance function* $H(\cdot,\cdot)$ *and has almost surely continuous sample paths*.

Regarding the relative efficiency of the proposed SEE estimator compared to its nonsmoothed counterpart [44], note that the (integrated) kernel $\bar{K}_h(u)$ converges to $\mathbb{1}(u\geq 0)$ as $h\to 0$. Hence, the smoothed process $n^{-1/2}\sum_{i=1}^{n}\boldsymbol{U}_i(\tau)$ with $\boldsymbol{U}_i(\tau)$ given in (15) has the same asymptotic distribution as

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Big\{\tau_L+\int_{\tau_L}^{\tau}\mathbb{1}\big(y_i\geq\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}^*(u)\big)\,\mathrm{d}H(u)-\Delta_i\mathbb{1}\big(y_i\leq\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}^*(\tau)\big)\Big\}\boldsymbol{x}_i.$$

As a result, the covariance function $H(\cdot,\cdot)$ defined in (21) coincides with that in [44]; see the proof of Theorem 2 therein. In other words, the SEE estimator and Peng and Huang's estimator converge to the same Gaussian process as $n\to\infty$ with $p$ fixed, and thence the asymptotic relative efficiency is 1. The technical devices required to deal with the fixed-$p$ and growing-$p$ cases are quite different. For the former, the consistency follows from the Glivenko–Cantelli theorem, and the weak convergence is a consequence of Donsker's theorem. To establish nonasymptotic results, we rely on a localized analysis as well as a (local) restricted strong convexity of the smoothed objective function that holds with high probability. The weak convergence is based on the nonasymptotic uniform Bahadur representation (Theorem 3.2), complemented by showing the convergence of finite-dimensional marginals and the asymptotic tightness.

3.3. *Rademacher multiplier bootstrap inference.* In this section, we establish the theoretical guarantees of the Rademacher multiplier/weighted bootstrap for censored quantile regression as described in Section 2.3. In this case, $W_i=e_i+1$ and $e_i$'s are i.i.d. Rademacher random variables. For the random covaviate vector $\boldsymbol{x}\in\mathbb{R}^p$, we assume that the moment parameters $m_3$ and $m_4$ defined in (9) are dimension-free. We first present the (conditional) uniform consistency of the bootstrapped process $\{\widehat{\boldsymbol{\beta}}^{\flat}(\tau):\tau\in[\tau_L,\tau_U]\}$ given the observed data $\mathbb{D}_n=\{(y_i,\Delta_i,\boldsymbol{x}_i\}_{i=1}^n$. Let $\mathbb{P}^*(\cdot)=\mathbb{P}(\cdot|\mathbb{D}_n)$ be the conditional probability given $\mathbb{D}_n$.

THEOREM 3.4 (Conditional uniform consistency). *Assume Conditions 3.1–3.4 hold, and let the bandwidth satisfy* $h=h_n\asymp\{(p+\log n)/n\}^{\gamma}$ *for some* $\gamma\in[1/4,1/2)$. *Then there*

*exists an event* $\mathcal{E} = \mathcal{E}(\mathbb{D}_n)$ *with* $\mathbb{P}(\mathcal{E}) \geq 1 - C_3 n^{-1}$ *such that conditional on* $\mathcal{E}$, *the bound* (11) *holds, and the bootstrapped process* $\widehat{\boldsymbol{\beta}}^\flat(\cdot) : [\tau_L, \tau_U] \mapsto \mathbb{R}^p$ *satisfies*

$$\sup_{\tau \in [\tau_L, \tau_U]} \|\widehat{\boldsymbol{\beta}}^\flat(\tau) - \widehat{\boldsymbol{\beta}}(\tau)\|_\Sigma \lesssim \sqrt{\frac{p + \log n}{n}}, \tag{23}$$

*with* $\mathbb{P}^*$*-probability at least* $1 - C_3 n^{-1}$, *provided* $\zeta_p^2 (p + \log n)^{1/2-\gamma} (p \log n)^{1/2} \lesssim n^{1-\gamma}$. *Here,* $C_3 > 0$ *is an absolute constant.*

Analogously to (13), define the bootstrapped integrated error as

$$\widehat{\boldsymbol{e}}^\flat(\tau) := \mathbf{J}(\tau)\{\widehat{\boldsymbol{\beta}}^\flat(\tau) - \widehat{\boldsymbol{\beta}}(\tau)\} + \int_{\tau_L}^{\tau} \mathbf{H}(u)\{\widehat{\boldsymbol{\beta}}^\flat(u) - \widehat{\boldsymbol{\beta}}(u)\}\, \mathrm{d}H(u), \tag{24}$$

where $\mathbf{J}(\cdot)$ and $\mathbf{H}(\cdot)$ are given in (12). We then develop a linear representation for $\widehat{\boldsymbol{e}}^\flat(\tau)$, which can be viewed as a parallel version of Theorem 3.2 in the bootstrap world.

THEOREM 3.5 (Conditional uniform Bahadur representation). *Assume the conditions in Theorem 3.4 hold, and that the kernel* $K(\cdot)$ *in Condition 3.1 is Lipschitz continuous. Moreover, assume* $\delta^* \lesssim n^{-(1/2+\alpha)}$ *for some* $\alpha > 0$. *Then there exists an event* $\mathcal{F} = \mathcal{F}(\mathbb{D}_n)$ *with* $\mathbb{P}(\mathcal{F}) \geq 1 - C_4 n^{-1}$ *such that conditional on* $\mathcal{F}$, (14)–(16) *hold, and the bootstrapped process* $\widehat{\boldsymbol{\beta}}^\flat(\cdot) : [\tau_L, \tau_U] \mapsto \mathbb{R}^p$ *satisfies*

$$\widehat{\boldsymbol{e}}^\flat(\tau) = \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{U}_i^\flat(\tau) + \boldsymbol{r}_n^\flat(\tau), \tag{25}$$

*where* $\boldsymbol{U}_i^\flat(\tau) = e_i \boldsymbol{U}_i(\tau)$ *with* $\boldsymbol{U}_i(\tau)$ *defined in* (15), *and*

$$\begin{aligned} &\sup_{\tau \in [\tau_L, \tau_U]} \|\boldsymbol{r}_n^\flat(\tau)\|_{\Sigma^{-1}} \\ &\lesssim m_4^{1/2} \frac{p + \log n}{n h^{1/2}} + h \sqrt{\frac{p + \log n}{n}} + \zeta_p^2 \frac{(p + \log n)(p \log n)^{1/2}}{n^{3/2} h} + n^{-1/2-\alpha} \end{aligned} \tag{26}$$

*with* $\mathbb{P}^*$*-probability at least* $1 - C_4 n^{-1}$.

Theorem 3.5 shows that the bootstrap integrated error $\widehat{\boldsymbol{e}}^\flat(\cdot)$ can be approximated, up to a higher order remainder, by the linear process $\{(1/n) \sum_{i=1}^n e_i \boldsymbol{U}_i(\tau) : \tau \in [\tau_L, \tau_U]\}$, where $e_i$'s are independent Rademacher random variables, and $\mathbb{E}^* \boldsymbol{U}_i^\flat(\tau) = \mathbf{0}$. Provided that $h \asymp \{(p + \log n)/n\}^{2/5}$ and $p$ satisfies the growth condition $p = o(n^{3/8})$ as in Theorem 3.3, then applying the same analysis in Remark 3.1 gives us the following stochastic integral representation: with probability (over $\mathbb{D}_n$) approaching one, $\sup_{\tau \in [\tau_L, \tau_U]} \|\boldsymbol{r}_n^\flat(\tau)\|_{\Sigma^{-1}} = o_{\mathbb{P}^*}(1)$, and

$$n^{1/2} \mathbf{J}(\tau)\{\widehat{\boldsymbol{\beta}}^\flat(\tau) - \widehat{\boldsymbol{\beta}}(\tau)\} = \boldsymbol{\phi}\left( \frac{1}{n^{1/2}} \sum_{i=1}^{n} \boldsymbol{U}_i^\flat(\tau) \right) + o_{\mathbb{P}^*}(1), \tag{27}$$

where $\boldsymbol{\phi}$ is the linear operator defined in (19). Note that $\mathbb{E}^*\{\boldsymbol{U}_i^\flat(s) \boldsymbol{U}_i^\flat(t)^{\mathrm{T}}\} = \boldsymbol{U}_i(s) \boldsymbol{U}_i(t)^{\mathrm{T}}$ for any $s, t \in [\tau_L, \tau_U]$. It can be shown that on $[\tau_L, \tau_U]$, $n^{-1/2} \sum_{i=1}^n \{\boldsymbol{U}_i(\cdot) - \mathbb{E}\boldsymbol{U}_i(\cdot)\}$ has the same asymptotic distribution as $n^{-1/2} \sum_{i=1}^n \boldsymbol{U}_i^\flat(\cdot)$ conditionally on the data $\mathbb{D}_n$; see Theorem 3.3 and Theorem 3.6 below. This, together with (18) and (27), validates to some level the

use of the bootstrap process $\widehat{\boldsymbol{\beta}}^\flat(\cdot)$ in the inference. To illustrate this, consider the following bootstrap counterpart of the process $\mathbb{G}_n(\cdot)$ defined in (20):

$$\mathbb{G}_n^\flat(\tau) := \frac{1}{n^{1/2}} \sum_{i=1}^n \langle \boldsymbol{a}_n / \|\boldsymbol{a}_n\|_\Sigma, \boldsymbol{U}_i^\flat(\tau) \rangle, \quad \tau \in [\tau_L, \tau_U]. \tag{28}$$

THEOREM 3.6 (Validation of bootstrap process). *Assume Conditions 3.1–3.4 hold with* $\delta^* \lesssim n^{-(1/2+\alpha)}$ *for* $\alpha \in (0, 1/2)$, $h \asymp \{(p+\log n)/n\}^{2/5}$ *and* $p = o(n^{3/8})$. *In addition, assume the kernel* $K(\cdot)$ *is Lipschitz continuous. Then, for any sequence of (deterministic) vectors* $\{\boldsymbol{a}_n\}_{n=1}^\infty$, *there exists a sequence of events* $\{\mathcal{F}_n = \mathcal{F}_n(\mathbb{D}_n)\}_{n=1}^\infty$ *such that* $\mathbb{P}(\mathcal{F}_n) \to 1$, *and conditional on* $\{\mathcal{F}_n\}_{n=1}^\infty$, (25) *holds and the conditional distribution of* $\mathbb{G}_n^\flat(\cdot)$ *given* $\mathbb{D}_n$ *is asymptotically equivalent to the unconditional distribution of* $\mathbb{G}_n(\cdot)$ *established in* (22).

**4. Regularized censored quantile regression.** We extend the proposed SEE approach to high-dimensional sparse QR models with random censoring. The goal is to identify the set of relevant predictors, defined as

$$\mathcal{S}^* = \bigcup_{\tau \in [\tau_L, \tau_U]} \operatorname{supp}(\boldsymbol{\beta}^*(\tau)), \tag{29}$$

assuming that its cardinality $s := |\mathcal{S}^*|$ is much smaller than the ambient dimension $p$—the total number of predictors, but may grow with sample size $n$. Recall the sequentially defined smoothed loss functions $\widehat{L}_k(\cdot)$ $(k = 0, 1, \ldots, m)$ in (5) and (6). When $p < n$, finding the solution to the SEE $\widehat{Q}_k(\boldsymbol{\beta}) = 0$ is equivalent to solving the optimization problem $\min_{\boldsymbol{\beta} \in \mathbb{R}^p} \widehat{L}_k(\boldsymbol{\beta})$. For fitting sparse models in high dimensions, we start with the $\ell_1$-penalized approach [4, 57]. At quantile levels $\tau_L = \tau_0 < \tau_1 < \cdots < \tau_m = \tau_U$, we define $\ell_1$-penalized smoothed CQR estimators $\widehat{\boldsymbol{\beta}}_k := \widehat{\boldsymbol{\beta}}(\tau_k)$ sequentially as

$$\widehat{\boldsymbol{\beta}}(\tau_k) \in \operatorname*{argmin}_{\boldsymbol{\beta} \in \mathbb{R}^p} \{\widehat{L}_k(\boldsymbol{\beta}) + \lambda_k \cdot \|\boldsymbol{\beta}\|_1\}, \tag{30}$$

for $k \in \{0, \ldots, m\}$, where $0 < \lambda_1 \le \cdots \le \lambda_m$ are regularization parameters. Define $\widehat{\boldsymbol{\beta}}(\tau) = \widehat{\boldsymbol{\beta}}(\tau_{k-1})$ for $\tau \in (\tau_{k-1}, \tau_k)$. It is worth noticing that for each $k \ge 1$, $\widehat{L}_k(\cdot)$ is essentially a shifted or perturbed version of $\widehat{L}_0(\cdot)$, that is, $\widehat{L}_k(\boldsymbol{\beta}) = \widehat{L}_0(\boldsymbol{\beta}) - (1/n) \sum_{i=1}^n \sum_{j=0}^{k-1} \bar{K}_h(y_i - \boldsymbol{x}_i^{\mathrm{T}} \widehat{\boldsymbol{\beta}}_j)\{H(\tau_{j+1}) - H(\tau_j)\}\boldsymbol{x}_i^{\mathrm{T}} \boldsymbol{\beta}$, where $H(u) = -\log(1-u)$. All of these empirical loss functions are convex, and have the same Hessian matrix.

CONDITION 4.1 (Random design in high dimensions). The (random) covariate vector $\boldsymbol{x} = (x_1, \ldots, x_p)^{\mathrm{T}} \in \mathcal{X} \subseteq \mathbb{R}^p$ $(x_1 \equiv 1)$ is compactly supported with $\max_{1 \le j \le p} |x_j| \le B_0$ almost surely for some $B_0 \ge 1$. For convenience, assume $B_0 = 1$. The normalized vector $\Sigma^{-1/2}\boldsymbol{x}$ has uniformly bounded kurtosis, that is, $m_4$ defined in (9) is a dimension-free constant, where $\Sigma = \mathbb{E}(\boldsymbol{x}\boldsymbol{x}^{\mathrm{T}})$ is positive definite.

THEOREM 4.1. *Assume Conditions 3.1, 3.3, 3.4 and Condition 4.1 hold. Under the sample size scaling* $n \gtrsim s^3 \log p$, *let the bandwidth* $h$ *and penalty levels* $\lambda_k$*'s satisfy* $s\sqrt{\log(p)/n} \lesssim h \lesssim \{s \log(p)/n\}^{1/4}$ *and* $\lambda_k \asymp \{1 + \log(\frac{1-\tau_L}{1-\tau_k})\}\sqrt{\log(p)/n}$ *for* $k = 0, 1, \ldots, m$. *Then there exist constants* $C_0, C_1, C_2 > 0$ *independent of* $(s, p, n)$ *such that*

$$\sup_{\tau_L \le \tau \le \tau_U} \|\widehat{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}^*(\tau)\|_\Sigma \le C_1 \left(\frac{1-\tau_L}{1-\tau_U}\right)^{C_0 \overline{f}/\underline{g}} \underline{g}^{-1} \log\left(\frac{1-\tau_L}{1-\tau_U}\right) \sqrt{\frac{s \log p}{\underline{\gamma} n}}$$

*with probability at least* $1 - C_2 p^{-1}$, *where* $\underline{\gamma} = \lambda_{\min}(\Sigma)$ *is the minimal eigenvalue of* $\Sigma$.

Theorem 4.1 provides the rate of convergence for the $\ell_1$-penalized smoothed CQR estimator $\widehat{\boldsymbol{\beta}}(\cdot)$ uniformly in the set of quantile indices $\tau \in [\tau_L, \tau_U]$. Under a similar set of assumptions, [67] established the uniform convergence rate for the $\ell_1$-penalized (nonsmoothed) CQR estimator, which is of order $\exp(Cs)\sqrt{s\log(p \vee n)/n}$. We conjecture that the additional exponential term $\exp(Cs)$ is a consequence of the marginal smoothness condition posed in [67] (see Condition (C4) therein), and can be relaxed as in our Theorem 4.1. In fact, our analysis relies on the global Lipschitz property (10), which follows directly from the model assumption (1) and a lower bound on the conditional density.

REMARK 4.1 (Comments on the tuning parameters $h$ and $\{\lambda_k\}_{k=0}^m$). To achieve the same convergence rate $\sqrt{s\log(p)/n}$ for the $\ell_1$-penalized QR estimator with noncensored data [4], the bandwidth $h$ is required to be in the range specified in Theorem 4.1; for example, one may choose $h \asymp \{s\log(p)/n\}^{1/4}$. Since such a choice depends on the unknown sparsity, in practice we simply choose $h$ to be of order $\{\log(p)/n\}^{1/4}$. Since the numerical performance is rather insensitive to the choice of bandwidth, we use the default value $h = \max\{0.05, 0.5\{\log(p)/n\}^{1/4}\}$ as suggested in [54] although it can also be tuned by cross-validation.

The penalty levels $\lambda_k$'s play a more pivotal role in obtaining a reasonable fit for the whole CQR process. Our theoretical analysis suggests that $\{\lambda_k\}_{k=0}^m$ should be chosen as a slowly growing sequence along the $\tau$-grid. Numerical results also confirm that a single $\lambda$ value, even after proper tuning, cannot guarantee a quality estimation of the entire regression process. On the other hand, it is computationally prohibitive to determine each $\lambda_k$ $(k = 0, 1, \ldots, m)$ via cross-validation. By examining the proof of Theorem 4.1, we see that once $\lambda_0$ is specified, the subsequent $\lambda_k$'s satisfy $\lambda_k = \{1 + \log(\frac{1-\tau_L}{1-\tau_k})\}\lambda_0$ for $k = 1, \ldots, m$. Therefore, to implement the proposed sequential procedure, we only treat $\lambda_0$ as a tuning parameter, and use the above formula to determine the rest of $\lambda_k$'s.

REMARK 4.2 (Adaptive $\ell_1$-penalization). It has been recognized that the $\ell_1$-penalized estimator, with the penalty level determined via cross-validation, typically has small prediction error but has a nonnegligible estimation bias and tends to overfit with many false discoveries. To reduce the estimation error and false positives, a popular strategy is to use reweighted $\ell_1$-penalization via either adaptive Lasso [68] or the local linear approximation (LLA) method for folded-concave penalties [15, 54, 69]. Let $w(\cdot)$ be a nonincreasing and nonnegative function defined on $[0, \infty)$. Fix $k$, let $\widehat{\boldsymbol{\beta}}_k^{(0)} = \widehat{\boldsymbol{\beta}}(\tau_k)$ be the $\ell_1$-penalized censored QR estimator at quantile level $\tau_k$. For $t = 1, \ldots, T$, we iteratively update the previous estimate $\widehat{\boldsymbol{\beta}}_k^{(t-1)}$ by solving

$$\widehat{\boldsymbol{\beta}}_k^{(t)} = (\widehat{\beta}_{k,1}^{(t)}, \ldots, \widehat{\beta}_{k,p}^{(t)})^{\mathrm{T}} \in \operatorname*{argmin}_{\boldsymbol{\beta}\in\mathbb{R}^p} \left\{ \widehat{L}_k(\boldsymbol{\beta}) + \lambda_k \cdot \sum_{j=1}^{p} w\big(|\widehat{\beta}_{k,j}^{(t-1)}|/\lambda_k\big)|\beta_j| \right\}.$$

When $T = 1$ and $w(u) = u^{-1}$ for $u > 0$ (or $(u + \epsilon)^{-1}$ for a small constant $\epsilon > 0$), this corresponds to an adaptive Lasso-type estimator [68]; when $w(u) = \mathbb{1}(u \le 1) + \frac{(a-u)_+}{a-1}\mathbb{1}(u > 1)$ for $u \ge 0$ and some $a > 2$, this corresponds to the LLA method using the smoothly clipped absolute deviation (SCAD) penalty [15]; when $w(u) = (1 - u/a)_+$ for $u \ge 0$ and some $a \ge 1$, this corresponds to the LLA method using the minimax concave penalty (MCP) [66].

**5. Numerical studies.** We apply the proposed methods in Sections 2 and 4 on simulated data sets and compare to that of Peng and Huang [44] and Zheng et al. [67] for both low- and high-dimensional settings in Sections 5.1 and 5.2, respectively. The proposed method

involves selecting a smoothing parameter $h$: for $p < n$, we set $h = \{(p + \log n)/n\}^{2/5} \vee 0.05$; for $p > n$, guided by Remark 4.1, we set $h = \{0.05 \vee 0.5\{\log(p)/n\}^{1/4}\}$. We found that the performance of our proposed method is insensitive to the choice of bandwidth, as also observed in [18] and [23]. We implemented Peng and Huang [44] using the `crq` function with `method = "PengHuang"` from the `quantreg` package [33]. On the other hand, Zheng et al. [67] is implemented using the barebones function `LASSO.fit` from `rqPen` [51] instead of the function `rq(..., method = "lasso")` in the package `quantreg`. This is because the function `rq(..., method = "lasso")` reports some numerical issues (e.g., singular design error) frequently in our numerical studies. All of the numerical studies are performed on a worker node with 32 CPUs, 2.5 GHz processor and 512 GB of memory in a high-performance computing cluster.

5.1. *Censored quantile regression*: *Estimation and inference*. We assess the performance of our proposed method in the low-dimensional setting with $n = 5000$ and $p = 100$. We start with generating the random covariates $\widetilde{\boldsymbol{x}}_i \in \mathbb{R}^p$ from a mixture of different distributions to represent different types of variables commonly encountered in many data sets. In particular, we generate the first 45 covariates from $\mathcal{N}(\mathbf{0}, \Sigma = (\sigma_{jk})_{1 \le j,k \le 45})$, where $\sigma_{jk} = 0.5^{|j-k|}$ for $1 \le j, k \le 45$, the second 45 covariates from a multivariate uniform distribution on the cube $[-2, 2]^{45}$ with the same covariance matrix $\Sigma$ using the `R` package `MultiRNG`, and the last 10 covariates from a Bernoulli distribution. Note that the three blocks of covariates generated are independent across the blocks. The response variables $z_i \in \mathbb{R}$ are then generated from the following models, both of which satisfy the global assumption in (1).

(i) Homoscedastic model: $z_i = \langle \widetilde{\boldsymbol{x}}_i, \boldsymbol{\gamma} \rangle + \varepsilon_i$ for $i = 1, \ldots, n$, where $\gamma_j \sim \text{Uniform}(-2, 2)$ for $j = 1, \ldots, p$. Let $Q_{t_2}(\tau)$ be the $\tau$-quantile of the $t_2$-distribution, and let $\boldsymbol{x}_i = (1, \widetilde{\boldsymbol{x}}_i^{\mathrm{T}})^{\mathrm{T}}$. Then the above model can be equivalently formulated as

$$z_i = \langle \boldsymbol{x}_i, \boldsymbol{\beta}^*(\tau) \rangle, \quad i = 1, \ldots, n, \quad \text{where } \boldsymbol{\beta}^*(\tau) = \big(Q_{t_2}(\tau), \boldsymbol{\gamma}^{\mathrm{T}}\big)^{\mathrm{T}} \in \mathbb{R}^{p+1}. \tag{31}$$

Under the above model, the covariate effects remain the same across all quantile levels.

(ii) Heteroscedastic model: $z_i = \langle \widetilde{\boldsymbol{x}}_i, \boldsymbol{\gamma} \rangle + |\widetilde{x}_{i,1}| \cdot \varepsilon_i$ for $i = 1, \ldots, n$, where $\gamma_1 = 0$ and $\gamma_j \sim \text{Uniform}(-2, 2)$ for $j = 2, \ldots, p$. Let $\boldsymbol{x}_i = (1, |\widetilde{x}_{i,1}|, \widetilde{\boldsymbol{x}}_{i,-1}^{\mathrm{T}})^{\mathrm{T}}$, where $\widetilde{\boldsymbol{x}}_{i,-1} \in \mathbb{R}^{p-1}$ is obtained by removing the first element of $\widetilde{\boldsymbol{x}}_i$. The model is equivalent to

$$\begin{aligned} z_i &= \langle \boldsymbol{x}_i, \boldsymbol{\beta}^*(\tau) \rangle, \quad i = 1, \ldots, n, \\ &\text{where } \boldsymbol{\beta}^*(\tau) = \big(0, Q_{t_2}(\tau), \gamma_2, \ldots, \gamma_p\big)^{\mathrm{T}} \in \mathbb{R}^{p+1}. \end{aligned} \tag{32}$$

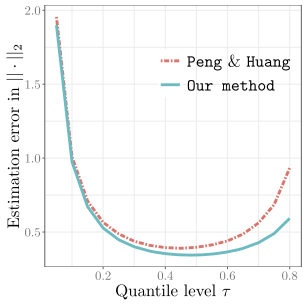
In this model, the first covariate has varying marginal effects for different quantile levels. Specifically, the effect of $|\widetilde{x}_1|$ on the $\tau$th quantile of $z$ is $F_{t_2}^{-1}(\tau)$, which is negligible when $\tau \approx 0.5$, but grows stronger as $\tau$ moves toward 0 or 1.

For both types of models, the random censoring variables are generated from a Gaussian mixture distribution, that is,

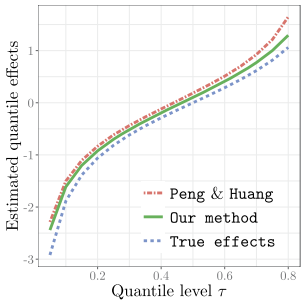
$$C_i \sim \mathbb{1}\{w_i = 1\}\mathcal{N}(0, 16) + \mathbb{1}\{w_i = 2\}\mathcal{N}(5, 1) + \mathbb{1}\{w_i = 3\}\mathcal{N}(10, 0.25) \tag{33}$$

for $i = 1, \ldots, n$, where $w_i$ is sampled from $\{1, 2, 3\}$ with equal probability, and $y_i = z_i \wedge C_i$ is the censored outcome. The corresponding censoring rate varies from 25% to 50%.

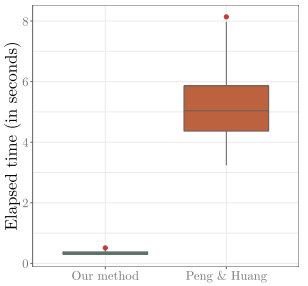
We implement both methods with a quantile grid of $\{\tau_k\}_{k=0}^m = \{0.05, 0.1, \ldots, 0.75, 0.8\}$. At each quantile level $\tau_k$, we use the estimation error under the $\ell_2$ norm, $\|\widehat{\boldsymbol{\beta}}(\tau_k) - \boldsymbol{\beta}^*(\tau_k)\|_2$, as a general measure of accuracy. We also calculate the run-time in seconds for both methods. Results, averaged across 500 independent replications, are reported in Figure 1. Fig-



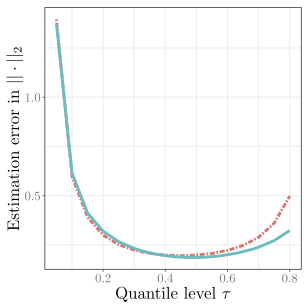

(a) $\ell_2$-error under model (31)



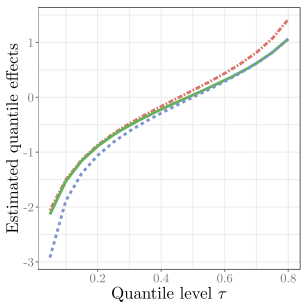

(b) Estimated quantile effects under model (31)



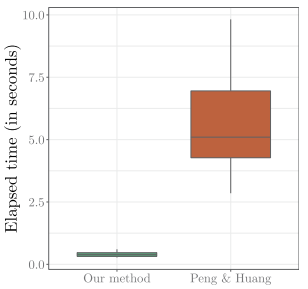

(c) Runtime under model (31)




(d) $\ell_2$-error under model (32)




(e) Estimated quantile effects under model (32)

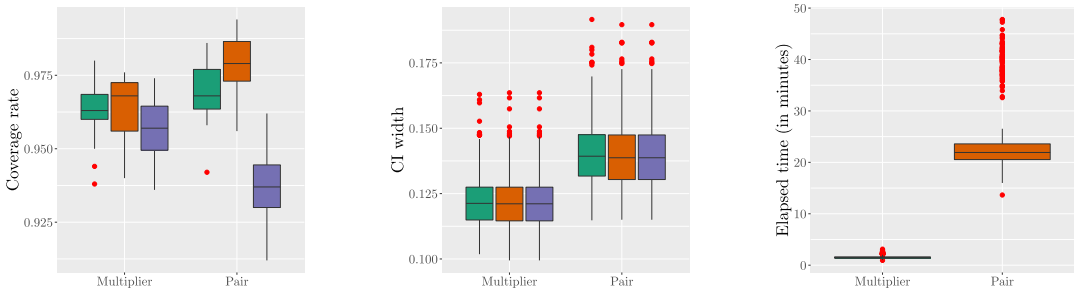



(f) Runtime under model (32)

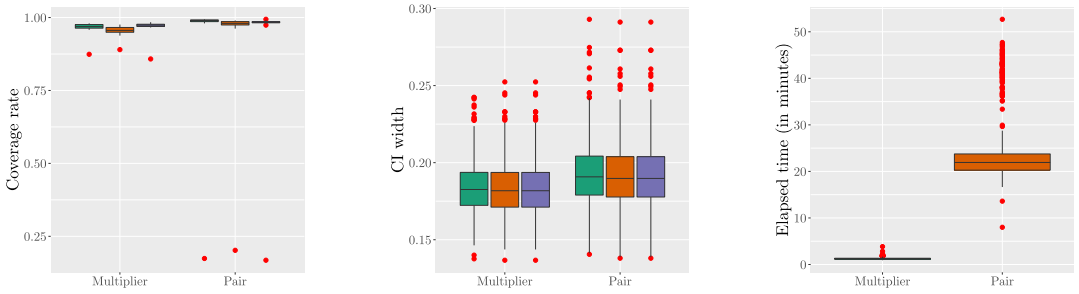
FIG. 1. *Numerical comparisons among CQR and our smoothed CQR for models* (31)–(32) *along the quantile grid. The left panels* (*a*) *and* (*d*) *display the* $\ell_2$*-induced estimation errors* $\|\widehat{\boldsymbol{\beta}}(\tau_k) - \boldsymbol{\beta}^*(\tau_k)\|_2$. *The middle panels* (*b*) *and* (*e*) *present the estimated quantile effects, which are* $\widehat{\beta}_0(\tau_k)$ *in model* (31) *and* $\widehat{\beta}_1(\tau_k)$ *in model* (32) *accordingly. The blue dashed lines in the middle panels represent the true quantile effects* $Q_{t_2}(\tau)$. *The right panels* (*c*) *and* (*f*) *record the empirical running time of the processes along the grid points.*

ures 1(a) and (d) contain the estimation error under the $\ell_2$ norm across all quantile levels; Figures 1(b) and (d) contains the regression coefficient that varies across quantile levels, that is, $\{\beta_0(\tau_k)\}_{k=0}^m$ for model (31) and $\{\beta_1(\tau_k)\}_{k=0}^m$ for model (32); and Figures 1(c) and (f) contain the computation time for fitting the entire QR process. We see that the two methods perform very closely at low quantile levels, and the smoothed approach is particularly advantageous at high quantile levels. Computationally, our implementation of the smoothed method is about 10 to 20 times faster than Peng and Huang [44]'s method, implemented by the `crq` function in `quantreg`. The numerical results on smaller-scale data sets are presented in Appendix F.1 of the Supplementary Material.

Next, we consider both the proposed multiplier bootstrap detailed in Section 2.3 and the classical paired bootstrap for performing statistical inference at $\tau = 0.5$. Three types of 95% confidence intervals (CIs) are constructed with $B = 1000$ bootstrap samples: the percentile CI, the pivotal CI and the normal CI. Coverage proportions for all of the covariates, confidence interval width for the first covariate, and computational time for the entire bootstrap process, averaged over 500 replications, are plotted in Figure 2. Under the homogeneous setting (31), all types of confidence intervals produced by multiplier bootstrap maintain the nominal level, while the normal intervals by pair resampling suffer from under coverage. In the heterogeneous setting (32), although outliers that correspond to the confidence intervals for the first covariate exist for both methods, multiplier bootstrap manages to mitigate this issue. Furthermore, compared to pair resampling, multiplier bootstrap constructs narrower confidence intervals with slightly smaller standard deviations. Finally, the computational advantage of multiplier bootstrap for smoothed CQR is evident in Figures 2(c) and (f).




(a) Coverage under model (31) (b) CI width under model (31) (c) Runtime under model (31)



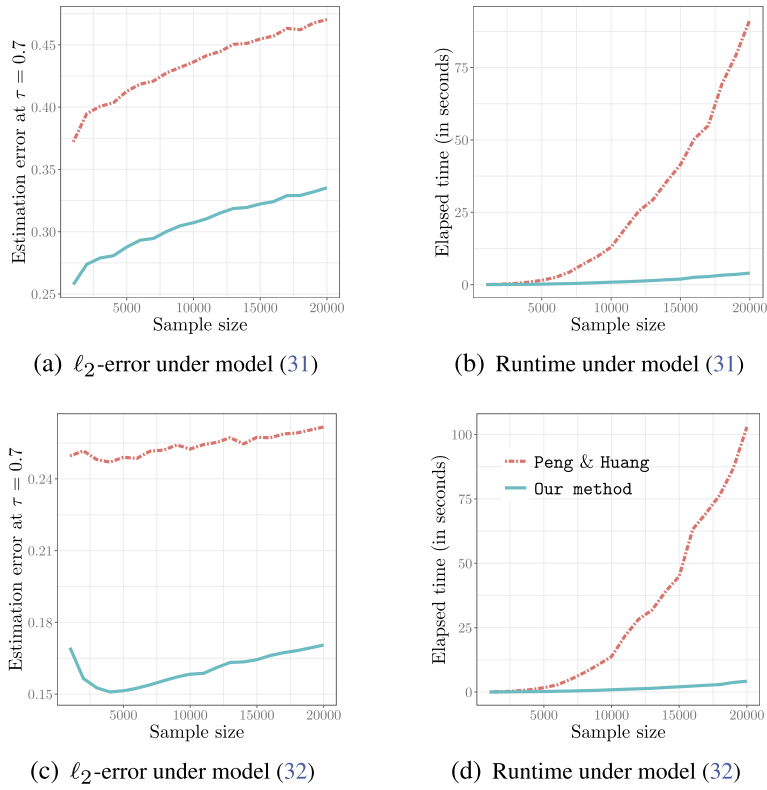

(d) Coverage under model (32) (e) CI width under model (32) (f) Runtime under model (32)

FIG. 2. *Box plots of the empirical coverage, confidence interval width, and running time for two resampling-based methods. "Multiplier" refers to the proposed multiplier bootstrap method, and "Pair" refers to pair resampling with replacement in the regression setting. In panels (a), (b), (d) and (e), within each method, different colors of boxes represent different types of confidence interval*: (*i*) *percentile interval* ■, (*ii*) *pivotal interval* ■ *and* (*iii*) *normal interval.* ■.

To better appreciate the computational advantage of smoothed CQR, we further consider large-scale simulation settings by setting $n \in \{1000, 2000, \ldots, 20{,}000\}$ and $p = n/100$. We use the same data generating processes as in (31)–(33), except that the covariates $\widetilde{\boldsymbol{x}}_i$ are now generated from $\mathcal{N}(\mathbf{0}_p, \Sigma)$ with $\Sigma = (0.5^{|j-k|})_{1\le j,k\le p}$. The censoring rate varies from 30% to 45%. In this case, we restrict attention to the estimation error and runtime of the two methods when $\tau = 0.7$. The results, averaged over 500 repetitions, are presented in Figure 3. We see from Figure 3 that the computation gain of the proposed method over Peng and Huang [44] is dramatic, without compromising the statistical accuracy. The estimation errors at $\tau \in \{0.3, 0.5\}$, as functions of the sample size, are displayed in Figure F.3 in the Supplementary Material.

5.2. *High-dimensional censored quantile regression.* In this section, we examine the numerical performance of the regularized smoothed CQR method with different penalties, which will also be compared with its nonsmoothed counterpart [67]. For the smoothed method, we consider both the $\ell_1$ and folded-concave penalties (SCAD and MCP). The latter is implemented by the LLA algorithm as described in Remark 4.2. The computational details are described in Section A.2 of the Supplementary Material.

Penalized CQR involves selecting a sequence of regularization parameters $\{\lambda_k\}_{k=0}^m$ that correspond to the predetermined $\tau$-grid $\{\tau_k\}_{k=0}^m$. Guided by Theorem 4.1 and Remark 4.1, we adopt a sequence of dilating $\lambda_k$'s with $\lambda_k = \{1 + \log(\frac{1-\tau_L}{1-\tau_k})\}\lambda_0$ for $k = 1, \ldots, m$, where $\lambda_0$ is chosen via the $K$-fold cross-validation ($K = 3$ in our studies). To accommodate censoring,




(a) $\ell_2$-error under model (31) (b) Runtime under model (31)

(c) $\ell_2$-error under model (32) (d) Runtime under model (32)

FIG. 3. *Numerical comparisons between CQR and smoothed CQR under models* (31) *and* (32) *with increasing* $(n, p)$ *subject to* $p = n/100$. *The left panels* (*a*) *and* (*c*) *display the* $\ell_2$*-error at* $\tau = 0.7$ *versus sample size. The right panels* (*b*) *and* (*d*) *present the runtime* (*in second*) *versus sample size.*

the cross-validation criterion is based on the the empirical mean of deviance residuals [55],

$$R(\lambda) := \frac{1}{n}\frac{1}{m+1}\sum_{i=1}^{n}\sum_{k=0}^{m}\sqrt{-2\{M_i(\tau_k,\lambda) + \Delta_i \log(\Delta_i - M_i(\tau_k,\lambda))\}} \tag{34}$$

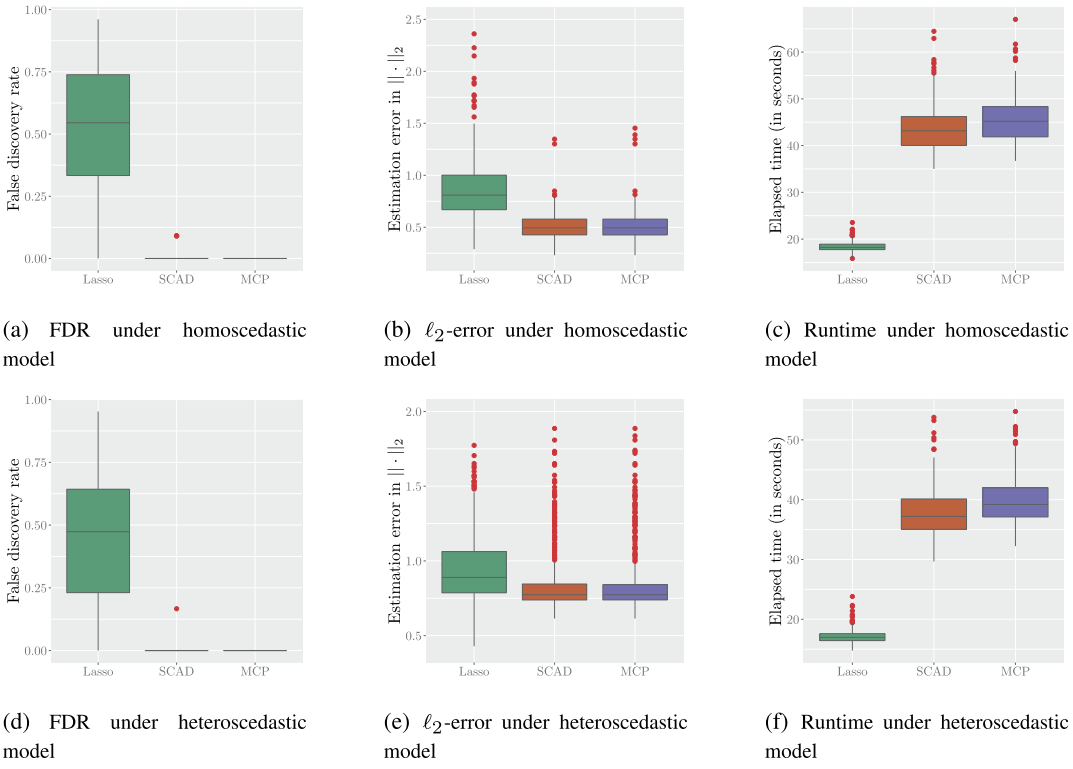
on the validation set, where

$$M_i(\tau_k,\lambda) = \mathbb{1}\{y_i \le \boldsymbol{x}_i^{\mathrm{T}}\widehat{\boldsymbol{\beta}}(\tau_k,\lambda), \Delta_i = 1\} - \int_{\tau_0}^{\tau_k} \mathbb{1}\{y_i \ge \boldsymbol{x}_i^{\mathrm{T}}\widehat{\boldsymbol{\beta}}(u,\lambda)\}\,\mathrm{d}H(u) - \tau_0$$

for $k = 0, \ldots, m$ are the martingale residuals and $\widehat{\boldsymbol{\beta}}(\tau, \lambda)$ refers to the estimated $\boldsymbol{\beta}(\tau)$ with a dilating $\lambda_k$'s starting with $\lambda_0 = \lambda$. The deviance (34) produces a more symmetric distribution through a transformation on the skewed martingale residuals, and is also used in [67] and [17]. In our simulations, we choose $\lambda_0$ from 50 candidates equally spaced on the interval $[0.01, 0.2]$.

In all of our numerical studies, we generate covariates $\widetilde{\boldsymbol{x}}_i \in \mathbb{R}^p$ from $\mathcal{N}(\mathbf{0}, \Sigma)$, where $\Sigma$ is as defined in Section 5.1, and the random errors $\varepsilon_i \sim t_2$. The response variables $z_i$ are generated from models (31)–(32), but with different $\boldsymbol{\gamma}$. For model (31), we consider a sparse $\boldsymbol{\gamma}$ with global sparsity $s = 10$ by setting $\gamma_j \sim \text{Uniform}(1, 1.5)$ for $j = 1, \ldots, 10$, and the rest to be zero. For model (32), $\boldsymbol{\gamma}$ is generated similarly except with $\gamma_1 = 0$. The random censoring variables are generated from (33), with overall censoring rates approximately 25%–30%.

Since the estimated active set depends on the entire quantile process, all numerical experiments are conducted via an estimation-after-selection procedure [67]. That is, in stage one, we perform regularized smoothed CQR to obtain the set $\widehat{\mathcal{S}} = \bigcup_{\tau\in\{\tau_0,\ldots,\tau_m\}} \text{supp}(\widehat{\boldsymbol{\beta}}(\tau))$. In stage




(a) FDR under homoscedastic model

(b) $\ell_2$-error under homoscedastic model

(c) Runtime under homoscedastic model

(d) FDR under heteroscedastic model

(e) $\ell_2$-error under heteroscedastic model

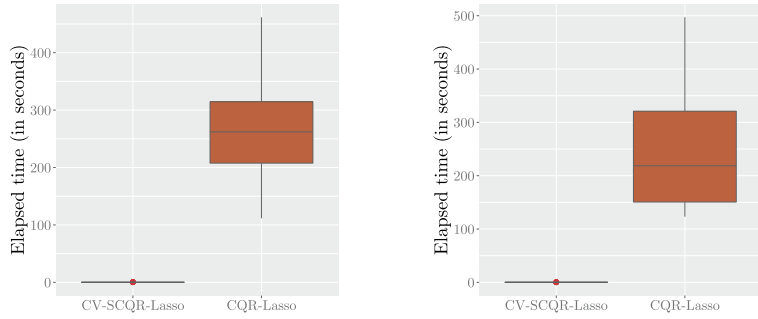
(f) Runtime under heteroscedastic model

FIG. 4. *Box plots of the false discovery rate, $\ell_2$-error and runtime for the $\ell_1$, SCAD and MCP regularized smoothed CQR. The true positive rates (TPR) are not visually informative, and thus are reported as follows. For the homoscedastic model, the average TPR are* 1.00 *for Lasso*, 0.9996 *for SCAD, and* 0.9992 *for MCP*; *for the heteroscedastic model, the average TPR are* 0.9872 *for Lasso*, 0.919 *for SCAD and* 0.917 *for MCP. The censoring rates vary between* 25% *and* 30%.

two, we perform smoothed CQR using the covariates in $\widehat{\mathcal{S}}$. Recall that $\mathcal{S}$ is the true active set defined in (29), and let $\mathcal{S}^c$ be its complement. To assess the numerical performance of our proposed method, we report (1) the true positive rate (TPR), $\text{TPR} = |\mathcal{S} \cap \widehat{\mathcal{S}}|/|\mathcal{S}|$; (2) the false discovery rate (FDR), $\text{FDR} = |\mathcal{S}^c \cap \widehat{\mathcal{S}}|/|\widehat{\mathcal{S}}|$; (3) average $\ell_2$-error, $(1/m)\sum_{k=0}^{m} \|\widehat{\boldsymbol{\beta}}(\tau_k) - \boldsymbol{\beta}(\tau_k)\|_2$; and (4) elapsed time for running the estimation-after-selection process, including cross-validation.

Results for the proposed method using different penalty functions, averaged over 500 replications when $(n, p) = (400, 1000)$, are reported in Figure 4. As expected, $\ell_1$-penalized method tends to select larger models with many spurious variables, and thus has higher false discovery rates than SCAD and MCP. Under the heterogeneous model, both SCAD and MCP sometimes miss the first true signal and have lower TPR than Lasso. This is due to the fact that the first signal corresponds to the evolving quantile effect $Q_{t_2}(\tau)$ that vanishes as $\tau$ approaches 0.5 and, therefore, is more likely to be missed by folded concave regularization.

To better demonstrate the computational efficiency of the proposed SEE method on large-scale data, we consider the $\ell_1$-penalized CQR (CQR-Lasso) method [67] as a benchmark. As discussed in [67], CQR-Lasso can be reformulated as a sequence of $\ell_1$-penalized median regressions with two pseudo observations, to which existing packages for penalized QR can be applied. Moreover, [67] used cross-validation to choose $\lambda_0$ (the initial penalty level) and the increment $c > 0$ by a two-dimensional grid search. In principle, we can apply this tuning scheme to both CQR-Lasso and its smoothed counterpart to achieve better variable selection performance. From a computational point of view, we apply a simpler tuning method




(a) Runtime under homoscedastic model

(b) Runtime under heteroscedastic model

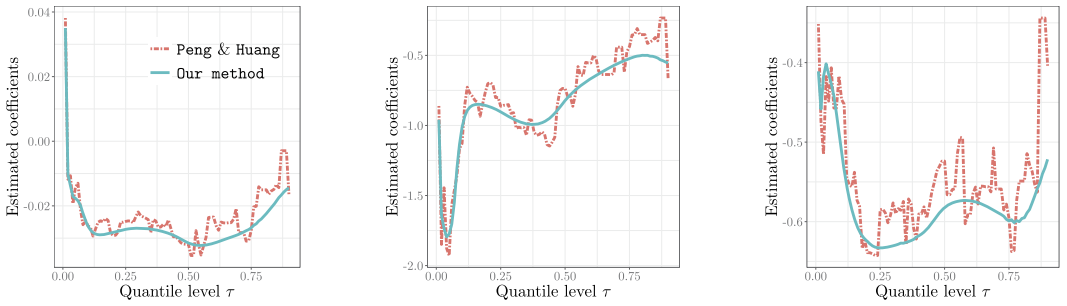
FIG. 5. *Box plots of runtime for $\ell_1$-penalized CQR (CQR-Lasso) and cross-validated $\ell_1$-penalized smoothed CQR (CV-SCQR-Lasso). The censoring rates vary from 25% to 30%. CQR-Lasso is implemented using the* `LASSO.fit` *function in the package* `rqPen`. *The results on FDR and TPR are shown in Figure F.4 in the Supplementary Material.*

by only choosing $\lambda_0$ via cross-validation and focus on speed comparisons. To be specific, we first compute the cross-validated $\ell_1$-penalized smoothed CQR (SCQR-Lasso) and record its runtime, and then compute the CQR-Lasso estimator using the same selected $\lambda$-sequence and record the runtime. For SCQR-Lasso, we apply the LAMM algorithm, described in Appendix A.2 of the Supplementary Material, to compute each $\widehat{\boldsymbol{\beta}}(\tau_k)$ defined in (30); for CQR-Lasso, we use the `LASSO.fit` function in `rqPen` to fit the penalized median regression at each quantile level. The box plots of running time (in second) over 500 replications are displayed in Figure 5. On average, our implementation of the cross-validated SCQR-Lasso is more than 10 times faster than the CQR-Lasso implementation without cross-validation (18 seconds versus 250 seconds). The box plots of false discovery rates are shown in Figure F.4 in the Supplementary Material. The code for the proposed method and our implementation of [67]'s method is available at https://github.com/XiaoouPan/scqr.

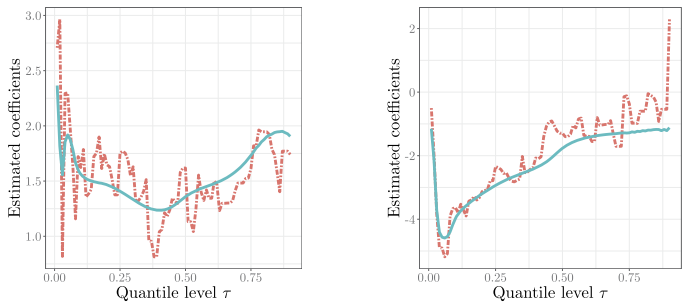
**6. Data applications.** As stated in the Introduction, the data applications are conducted on a worker node with 2.5 GHz 32-core processor and 512 GB of RAM in a high-performance computing cluster.

6.1. *Primary biliary cirrhosis data.* We apply the proposed method to the Mayo primary biliary cirrhosis data set [13], a double-blinded randomized trial conducted by Mayo Clinic between 1974 and 1984. Primary biliary cirrhosis is a rare but fatal chronic liver disease. Our response of interest is the survival time on logarithmic scale, and an observation is censored if the patient stays alive by the end of the research. Five variables are included into our modeling: age in days, the presence of edema, serum bilirubin in mg/dl, albumin in gm/dl and prothrombin time in seconds, with logarithmic transformations applied to the last three variables. These features are statistically significant in a multivariate Cox proportional hazards model [13]. After removing data with missing covariates, the data set contains 416 patients and a censoring rate of 61.5%.

We apply both the classical and the proposed smoothed CQR methods to this data set. The former is implemented by `crq(..., method = "PengHuang")` in the `quantreg` package over the quantile grid $\{0.01, 0.02, \ldots, 0.90\}$. The bandwidth parameter of our method is set to be $h = \max\{0.05, \{(p + \log n)/n\}^{2/5}\}$. The estimated regression coefficients are plotted in Figure 6 as functions of quantile levels. It is worth noting that our method leads




(a) Variable: age (b) Variable: edema (c) Variable: log(bili)




(d) Variable: log(albumin) (e) Variable: log(protime)

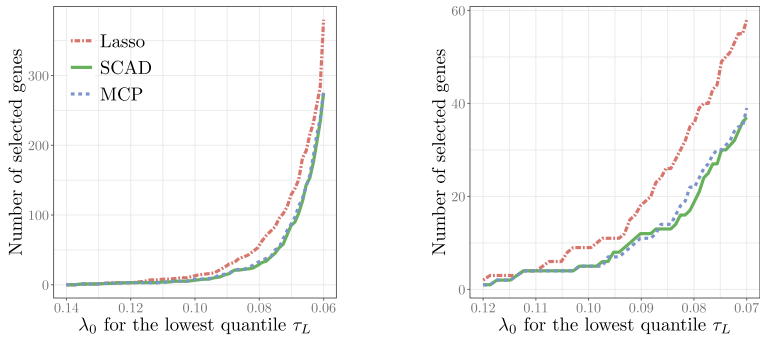
FIG. 6. *Estimated regression coefficients over* $\tau \in [0.01, 0.9]$ *for five variables in the Mayo primary biliary cirrhosis data. Specifically, "age" stands for age in days, "edema" indicates the presence of edema, "bili" represents serum bilirubin in mg/dl, "albumin" means albumin in gm/dl and "protime" refers to prothrombin time in seconds. Peng and Huang's estimators are obtained via* `crq` *function in the* `quantreg` *package with* `method = "PengHuang."`

to a fairly smooth estimated coefficient process, while there is much higher variability in the usual CQR estimator [44]. Arguably, this could be an advantage of the smoothed method because it produces more interpretable results.

Among the five covariates, age exhibits modest effects along the process, while albumin and prothrombin time possess varying effects with opposite signs, especially for short survivors. Our findings echo the conclusions made in [28], and offer an alternative perspective to this data set apart from [56], in which the regression coefficients are assumed to be different across the quantile levels.

6.2. *Microarray data for lung adenocarcinoma.* We now apply the proposed regularized smoothed CQR method to a gene expression-based data from a large retrospective study for survival prediction in lung cancer [50]. The data set provides gene expression profiling using microarray technologies, and has been briefly introduced in Section 1. After removing observations with missing values, we have 22,283 genes from 442 lung adenocarcinomas samples, with a censoring proportion of 46.6%.

To demonstrate the scalability of our method, we first run regularized CQR on the whole data set without any processing steps. Then, to roughly denoise the large data set and to better interpret the results, we follow the preprocessing procedure carried out in [67] by selecting 3000 genes with the largest variances, and further investigate the impact of these genes on lung cancer survival time. For both analysis strategies, the quantile grid is set to be $\{0.1, 0.11, \ldots, 0.7\}$, the bandwidth is set as $h = \{0.05 \vee 0.5\{\log(p)/n\}^{1/4}\}$ and the tuning parameter is gradually dilating with $\lambda_k = \{1 + \log(\frac{1-\tau_L}{1-\tau_k})\}\lambda_0$ for $k = 1, \ldots, m$, where $\lambda_0$ ranges over 50 reasonable candidates. Figure 7 contains the number of detected genes across




(a) Selected genes from the full data with $p = 22,283$

(b) Selected genes from the preprocessed data with $p = 3,000$

FIG. 7. *Number of selected genes from regularized smoothed CQR with Lasso, SCAD and MCP penalties, when $\lambda_0$ gradually decreases over a reasonable range. The left and right panels contain results for the entire data with $p = 22{,}283$ and the preprocessed data with $p = 3000$, respectively.*

various values of $\lambda_0$. Moreover, we report the first ten identified genes (denoted by their Affymetrix probe IDs) in Table 2.

Our proposed method is computationally scalable and takes only 2 to 4 minutes for fitting the large microarray data set with $p = 22{,}283$. As a reference, it takes 10.82 hours to run the $\ell_1$-penalized CQR even on the preprocessed data with $p = 3000$. In addition, with the same data, the genes detected by Lasso and nonconvex penalties substantially overlap, and some are commonly identified regardless of the preprocessing step, for example, "201250_s_at" and "200750_s_at." These genes may be potentially revealing and enlightening for survival prediction in lung cancer, and the intrinsic biological explanation can be a gripping topic for genetics research.

TABLE 2

*The leading 10 identified genes (presented by their Affymetrix probe IDs) using regularized smoothed CQR with Lasso, SCAD and MCP penalties, as $\lambda_0$ gradually diminishes. The methods are applied to the whole data with $p = 22{,}283$, and the preprocessed data with $p = 3000$. The last row indicates the average running time for each $\lambda_0$, and is recorded in minutes. The running time of $\ell_1$-penalized CQR for a single $\lambda_0$ is more than 7 days on the whole data, and 10.82 hours on the preprocessed data*

| | Whole data ($p = 22{,}283$) | | | Preprocessed data ($p = 3000$) | | |
|---|---|---|---|---|---|---|
| | Lasso | SCAD | MCP | Lasso | SCAD | MCP |
| Identified genes | 205394_at | 205394_at | 205394_at | 213911_s_at | 213911_s_at | 213911_s_at |
| | 220658_s_at | 220658_s_at | 220658_s_at | 217938_s_at | 217938_s_at | 217938_s_at |
| | 221249_s_at | 221249_s_at | 221249_s_at | 201890_at | 201890_at | 201890_at |
| | 209825_s_at | 201250_s_at | 201250_s_at | 201250_s_at | 201250_s_at | 201250_s_at |
| | 217938_s_at | 40093_at | 40093_at | 200750_s_at | 200750_s_at | 200750_s_at |
| | 201250_s_at | 204728_s_at | 200750_s_at | 212951_at | 201761_at | 201761_at |
| | 40093_at | 200750_s_at | 204728_s_at | 202503_s_at | 202503_s_at | 202503_s_at |
| | 203967_at | 218193_s_at | 209825_s_at | 209773_s_at | 212951_at | 200786_at |
| | 210052_s_at | 203967_at | 203967_at | 201761_at | 200786_at | 209773_s_at |
| | 218193_s_at | 219787_s_at | 219787_s_at | 204170_s_at | 209773_s_at | 212951_at |
| Time (in minutes) | 2.00 | 4.10 | 4.06 | 0.25 | 0.35 | 0.36 |

**Acknowledgments.** The authors acknowledge two anonymous referees and an Associate Editor for their constructive comments that improved the quality and presentation of this paper.

**Funding.** X. He was supported by NSF Grants DMS-1914496 and DMS-1951980. K. M. Tan was supported by NSF Grants DMS-1949730, DMS-2113356 and NIH Grant RF1-MH122833. W.-X. Zhou acknowledges the support of the NSF Grant DMS-2113409.

## SUPPLEMENTARY MATERIAL

**Supplementary material for "Scalable estimation and inference for censored quantile regression process"** (DOI: 10.1214/22-AOS2214SUPP; .pdf). This supplementary material contains the proofs of all theoretical results in Sections 3 and 4, along with the optimization algorithms and additional simulation studies.

## REFERENCES

[1] ANDERSEN, P. K., BORGAN, Ø., GILL, R. D. and KEIDING, N. (1993). *Statistical Models Based on Counting Processes*. *Springer Series in Statistics*. Springer, New York. MR1198884 https://doi.org/10.1007/978-1-4612-4348-9

[2] BARBE, P. and BERTAIL, P. (1995). *The Weighted Bootstrap*. *Lecture Notes in Statistics* **98**. Springer, New York. MR2195545 https://doi.org/10.1007/978-1-4612-2532-4

[3] BARRODALE, I. and ROBERTS, F. (1974). Solution of an overdetermined system of equations in the $\ell_1$ norm. *Commun. ACM* **17** 319–320.

[4] BELLONI, A. and CHERNOZHUKOV, V. (2011). $\ell_1$-penalized quantile regression in high-dimensional sparse models. *Ann. Statist.* **39** 82–130. MR2797841 https://doi.org/10.1214/10-AOS827

[5] BRADIC, J., FAN, J. and JIANG, J. (2011). Regularization for Cox's proportional hazards model with NP-dimensionality. *Ann. Statist.* **39** 3092–3120. MR3012402 https://doi.org/10.1214/11-AOS911

[6] BUCHINSKY, M. and HAHN, J. (1998). An alternative estimator for the censored quantile regression model. *Econometrica* **66** 653–671. MR1627038 https://doi.org/10.2307/2998578

[7] CAI, T., HUANG, J. and TIAN, L. (2009). Regularized estimation for the accelerated failure time model. *Biometrics* **65** 394–404. MR2751463 https://doi.org/10.1111/j.1541-0420.2008.01074.x

[8] CHATTERJEE, S. and BOSE, A. (2005). Generalized bootstrap for estimating equations. *Ann. Statist.* **33** 414–436. MR2157808 https://doi.org/10.1214/009053604000000904

[9] CHERNOZHUKOV, V. and HONG, H. (2002). Three-step censored quantile regression and extramarital affairs. *J. Amer. Statist. Assoc.* **97** 872–882. MR1941416 https://doi.org/10.1198/016214502388618663

[10] DE BACKER, M., EL GHOUCH, A. and VAN KEILEGOM, I. (2019). An adapted loss function for censored quantile regression. *J. Amer. Statist. Assoc.* **114** 1126–1137. MR4011767 https://doi.org/10.1080/01621459.2018.1469996

[11] DE BACKER, M., EL GHOUCH, A. and VAN KEILEGOM, I. (2020). Linear censored quantile regression: A novel minimum-distance approach. *Scand. J. Stat.* **47** 1275–1306. MR4178194 https://doi.org/10.1111/sjos.12475

[12] DE CASTRO, L., GALVAO, A. F., KAPLAN, D. M. and LIU, X. (2019). Smoothed GMM for quantile models. *J. Econometrics* **213** 121–144. MR4013218 https://doi.org/10.1016/j.jeconom.2019.04.008

[13] DICKSON, E. R., GRAMBSCH, P. M., FLEMING, T. R., FISHER, L. D. and LANGWORTHY, A. (1989). Prognosis in primary biliary cirrhosis: Model for decision making. *Hepatology* **10** 1–7. https://doi.org/10.1002/hep.1840100102

[14] EFRON, B. (1967). The two sample problem with censored data. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume* 4: *Biology and Problems of Health* 831–853.

[15] FAN, J. and LI, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *J. Amer. Statist. Assoc.* **96** 1348–1360. MR1946581 https://doi.org/10.1198/016214501753382273

[16] FAN, J. and LI, R. (2002). Variable selection for Cox's proportional hazards model and frailty model. *Ann. Statist.* **30** 74–99. MR1892656 https://doi.org/10.1214/aos/1015362185

[17] FEI, Z., ZHENG, Q., HONG, H. G. and LI, Y. (2021). Inference for high dimensional censored quantile regression. *J. Amer. Statist. Assoc.* https://doi.org/10.1080/01621459.2021.1957900

[18] FERNANDES, M., GUERRE, E. and HORTA, E. (2021). Smoothing quantile regressions. *J. Bus. Econom. Statist.* **39** 338–357. MR4187194 https://doi.org/10.1080/07350015.2019.1660177

[19] FLEMING, T. R. and HARRINGTON, D. P. (1991). *Counting Processes and Survival Analysis*. *Wiley Series in Probability and Mathematical Statistics*: *Applied Probability and Statistics*. Wiley, New York. MR1100924

[20] FYGENSON, M. and RITOV, Y. (1994). Monotone estimating equations for censored data. *Ann. Statist.* **22** 732–746. MR1292538 https://doi.org/10.1214/aos/1176325493

[21] GILL, R. D. and JOHANSEN, S. (1990). A survey of product-integration with a view toward application in survival analysis. *Ann. Statist.* **18** 1501–1555. MR1074422 https://doi.org/10.1214/aos/1176347865

[22] GU, Y., FAN, J., KONG, L., MA, S. and ZOU, H. (2018). ADMM for high-dimensional sparse penalized quantile regression. *Technometrics* **60** 319–331. MR3847169 https://doi.org/10.1080/00401706.2017.1345703

[23] HE, X., PAN, X., TAN, K. M. and ZHOU, W.-X. (2021). Smoothed quantile regression with large-scale inference. *J. Econometrics* https://doi.org/10.1016/j.jeconom.2021.07.010

[24] HE, X., PAN, X., TAN, K. M. and ZHOU, W.-X. (2022). Supplement to "Scalable estimation and inference for censored quantile regression process." https://doi.org/10.1214/22-AOS2214SUPP

[25] HONORÉ, B., KHAN, S. and POWELL, J. L. (2002). Quantile regression under random censoring. *J. Econometrics* **109** 67–105. MR1899693 https://doi.org/10.1016/S0304-4076(01)00142-7

[26] HU, F. and KALBFLEISCH, J. D. (2000). The estimating function bootstrap. *Canad. J. Statist.* **28** 449–499. MR1793106 https://doi.org/10.2307/3315958

[27] HUANG, J., MA, S. and XIE, H. (2006). Regularized estimation in the accelerated failure time model with high-dimensional covariates. *Biometrics* **62** 813–820. MR2247210 https://doi.org/10.1111/j.1541-0420.2006.00562.x

[28] HUANG, Y. (2010). Quantile calculus and censored regression. *Ann. Statist.* **38** 1607–1637. MR2662354 https://doi.org/10.1214/09-AOS771

[29] JIN, Z., YING, Z. and WEI, L. J. (2001). A simple resampling method by perturbing the minimand. *Biometrika* **88** 381–390. MR1844838 https://doi.org/10.1093/biomet/88.2.381

[30] KAPLAN, D. M. and SUN, Y. (2017). Smoothed estimating equations for instrumental variables quantile regression. *Econometric Theory* **33** 105–157. MR3574862 https://doi.org/10.1017/S0266466615000407

[31] KLEINBAUM, D. G. and KLEIN, M. (2012). *Survival Analysis*: *A Self-Learning Text*, 3rd ed. *Statistics for Biology and Health*. Springer, New York. MR2882858 https://doi.org/10.1007/978-1-4419-6646-9

[32] KOENKER, R. (2005). *Quantile Regression*. *Econometric Society Monographs* **38**. Cambridge Univ. Press, Cambridge. MR2268657 https://doi.org/10.1017/CBO9780511754098

[33] KOENKER, R. (2008). Censored quantile regression redux. *J. Stat. Softw.* **38** 1–25.

[34] KOENKER, R., CHERNOZHUKOV, V., HE, X. and PENG, L. (2017). *Handbook of Quantile Regression*. CRC Press, New York.

[35] KOENKER, R. and GELING, O. (2001). Reappraising medfly longevity: A quantile regression survival analysis. *J. Amer. Statist. Assoc.* **96** 458–468. MR1939348 https://doi.org/10.1198/016214501753168172

[36] KOENKER, R. and MIZERA, I. (2014). Convex optimization in R. *J. Stat. Softw.* **60**.

[37] KOENKER, R. and NG, P. (2005). A Frisch–Newton algorithm for sparse quantile regression. *Acta Math. Appl. Sin. Engl. Ser.* **21** 225–236. MR2141542 https://doi.org/10.1007/s10255-005-0231-1

[38] LENG, C. and TONG, X. (2013). A quantile regression estimator for censored data. *Bernoulli* **19** 344–361. MR3019498 https://doi.org/10.3150/11-BEJ388

[39] MA, S. and KOSOROK, M. R. (2005). Robust semiparametric M-estimation and the weighted bootstrap. *J. Multivariate Anal.* **96** 190–217. MR2202406 https://doi.org/10.1016/j.jmva.2004.09.008

[40] NEOCLEOUS, T., VANDEN BRANDEN, K. and PORTNOY, S. (2006). Correction to: "Censored regression quantiles" [J. Amer. Statist. Assoc. **98**(464) (2003), 1001–1012; MR2041488] by Portnoy. *J. Amer. Statist. Assoc.* **101** 860–861. MR2281250 https://doi.org/10.1198/016214506000000087

[41] PARZEN, M. I., WEI, L. J. and YING, Z. (1994). A resampling method based on pivotal estimating functions. *Biometrika* **81** 341–350. MR1294895 https://doi.org/10.1093/biomet/81.2.341

[42] PENG, L. (2012). Self-consistent estimation of censored quantile regression. *J. Multivariate Anal.* **105** 368–379. MR2877523 https://doi.org/10.1016/j.jmva.2011.10.005

[43] PENG, L. (2021). Quantile regression for survival data. *Annu. Rev. Stat. Appl.* **8** 413–437. MR4243554 https://doi.org/10.1146/annurev-statistics-042720-020233

[44] PENG, L. and HUANG, Y. (2008). Survival analysis with quantile regression models. *J. Amer. Statist. Assoc.* **103** 637–649. MR2435468 https://doi.org/10.1198/016214508000000355

[45] PORTNOY, S. (2003). Censored regression quantiles. *J. Amer. Statist. Assoc.* **98** 1001–1012. MR2041488 https://doi.org/10.1198/016214503000000954

[46] PORTNOY, S. and KOENKER, R. (1997). The Gaussian hare and the Laplacian tortoise: Computability of squared-error versus absolute-error estimators. *Statist. Sci.* **12** 279–300. MR1619189 https://doi.org/10.1214/ss/1030037960

[47] PORTNOY, S. and LIN, G. (2010). Asymptotics for censored regression quantiles. *J. Nonparametr. Stat.* **22** 115–130. MR2598957 https://doi.org/10.1080/10485250903105009

[48] POWELL, J. L. (1984). Least absolute deviations estimation for the censored regression model. *J. Econometrics* **25** 303–325. MR0752444 https://doi.org/10.1016/0304-4076(84)90004-6

[49] POWELL, J. L. (1986). Censored regression quantiles. *J. Econometrics* **32** 143–155. MR0853049 https://doi.org/10.1016/0304-4076(86)90016-3

[50] SHEDDEN, K., TAYLOR, J. M., ENKEMANN, S. A. (2008). Gene expression-based survival prediction in lung adenocarcinoma: A multi-site, blinded validation study. *Nat. Med.* **14** 822–827.

[51] SHERWOOD, B. and MAIDMAN, A. (2020). Package 'rqPen', version 2.2.2. Reference manual. https://cran.r-project.org/web/packages/rqPen/rqPen.pdf.

[52] SHOWS, J. H., LU, W. and ZHANG, H. H. (2010). Sparse estimation and inference for censored median regression. *J. Statist. Plann. Inference* **140** 1903–1917. MR2606727 https://doi.org/10.1016/j.jspi.2010.01.043

[53] SUN, X., PENG, L., HUANG, Y. and LAI, H. J. (2016). Generalizing quantile regression for counting processes with applications to recurrent events. *J. Amer. Statist. Assoc.* **111** 145–156. MR3494649 https://doi.org/10.1080/01621459.2014.995795

[54] TAN, K. M., WANG, L. and ZHOU, W.-X. (2022). High-dimensional quantile regression: Convolution smoothing and concave regularization. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **84** 205–233. MR4400395 https://doi.org/10.1111/rssb.12485

[55] THERNEAU, T. M., GRAMBSCH, P. M. and FLEMING, T. R. (1990). Martingale-based residuals for survival models. *Biometrika* **77** 147–160. MR1049416 https://doi.org/10.1093/biomet/77.1.147

[56] TIAN, L., ZUCKER, D. and WEI, L. J. (2005). On the Cox model with time-varying regression coefficients. *J. Amer. Statist. Assoc.* **100** 172–183. MR2156827 https://doi.org/10.1198/016214504000000845

[57] TIBSHIRANI, R. (1996). Regression shrinkage and selection via the lasso. *J. Roy. Statist. Soc. Ser. B* **58** 267–288. MR1379242

[58] VOLGUSHEV, S., WAGENER, J. and DETTE, H. (2014). Censored quantile regression processes under dependence and penalization. *Electron. J. Stat.* **8** 2405–2447. MR3278338 https://doi.org/10.1214/14-EJS54

[59] WANG, H. J. and WANG, L. (2009). Locally weighted censored quantile regression. *J. Amer. Statist. Assoc.* **104** 1117–1128. MR2562007 https://doi.org/10.1198/jasa.2009.tm08230

[60] WANG, H. J., ZHOU, J. and LI, Y. (2013). Variable selection for censored quantile regresion. *Statist. Sinica* **23** 145–167. MR3076162

[61] WHANG, Y.-J. (2006). Smoothed empirical likelihood methods for quantile regression models. *Econometric Theory* **22** 173–205. MR2230386 https://doi.org/10.1017/S0266466606060087

[62] WU, Y., MA, Y. and YIN, G. (2015). Smoothed and corrected score approach to censored quantile regression with measurement errors. *J. Amer. Statist. Assoc.* **110** 1670–1683. MR3449063 https://doi.org/10.1080/01621459.2014.989323

[63] YANG, X., NARISETTY, N. N. and HE, X. (2018). A new approach to censored quantile regression estimation. *J. Comput. Graph. Statist.* **27** 417–425. MR3816276 https://doi.org/10.1080/10618600.2017.1385469

[64] YING, Z., JUNG, S. H. and WEI, L. J. (1995). Survival analysis with median regression models. *J. Amer. Statist. Assoc.* **90** 178–184. MR1325125

[65] YU, L., LIN, N. and WANG, L. (2017). A parallel algorithm for large-scale nonconvex penalized quantile regression. *J. Comput. Graph. Statist.* **26** 935–939. MR3765357 https://doi.org/10.1080/10618600.2017.1328366

[66] ZHANG, C.-H. (2010). Nearly unbiased variable selection under minimax concave penalty. *Ann. Statist.* **38** 894–942. MR2604701 https://doi.org/10.1214/09-AOS729

[67] ZHENG, Q., PENG, L. and HE, X. (2018). High dimensional censored quantile regression. *Ann. Statist.* **46** 308–343. MR3766954 https://doi.org/10.1214/17-AOS1551

[68] ZOU, H. (2006). The adaptive lasso and its oracle properties. *J. Amer. Statist. Assoc.* **101** 1418–1429. MR2279469 https://doi.org/10.1198/016214506000000735

[69] ZOU, H. and LI, R. (2008). One-step sparse estimates in nonconcave penalized likelihood models. *Ann. Statist.* **36** 1509–1533. MR2435443 https://doi.org/10.1214/009053607000000802