mathematics

MDPI

Article

Smoothing Estimation of Parameters in Censored Quantile Linear Regression Model

Mingquan Wang [1], Xiaohua Ma [1], Xinrui Wang [2], Jun Wang [1], Xiuqing Zhou [1,*] and Qibing Gao [1,*]

[1] School of Mathematical Sciences, Nanjing Normal University, Nanjing 210023, China; wmqhhh@126.com (M.W.); 230902091@njnu.edu.cn (X.M.); 230902090@njnu.edu.cn (J.W.)
[2] College of International Languages and Cultures, Hohai University, Nanjing 211100, China; wangxinrui0824@163.com
* Correspondence: zhouxiuqing@njnu.edu.cn (X.Z.); gaoqibing@njnu.edu.cn (Q.G.)

Abstract: In this paper, we propose a smoothing estimation method for censored quantile regression models. The method associates the convolutional smoothing estimation with the loss function, which is quadratically derivable and globally convex by using a non-negative kernel function. Thus, the parameters of the regression model can be computed by using the gradient-based iterative algorithm. We demonstrate the convergence speed and asymptotic properties of the smoothing estimation for large samples in high dimensions. Numerical simulations show that the smoothing estimation method for censored quantile regression models improves the estimation accuracy, computational speed, and robustness over the classical parameter estimation method. The simulation results also show that the parametric methods perform better than the KM method in estimating the distribution function of the censored variables. Even if there is an error setting in the distribution estimation, the smoothing estimation does not fluctuate too much.

Keywords: censored quantile regression; high-dimensional data; smoothing estimation

MSC: 62N01; 62N02; 62-08

check for updates

Academic Editor: Simeon Reich

Received: 22 November 2024
Revised: 28 December 2024
Accepted: 6 January 2025
Published: 8 January 2025

Citation: Wang, M.; Ma, X.; Wang, X.; Wang, J.; Zhou, X.; Gao, Q. Smoothing Estimation of Parameters in Censored Quantile Linear Regression Model. *Mathematics* **2025**, *13*, 192. https://doi.org/10.3390/math13020192

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Since the regression quantile has similar robustness as the sample quantile, the quantile regression model can well characterize the conditional distribution of the response variable y when the independent variable $\boldsymbol{x}$ is given, thus characterizing the link between the two [1]. Response variable $y \in \mathbb{R}$ and p-dimensional covariates $\boldsymbol{x} = (x_1, ..., x_p)^T$ are given, and $F_{y|\boldsymbol{x}}(\cdot)$ is the conditional distribution function of y concerning $\boldsymbol{x}$. The classical quantile regression model can be expressed as

$$F^{-1}_{y|\boldsymbol{x}}(\tau) = [\boldsymbol{\beta}(\tau)]^T \boldsymbol{x}, \tag{1}$$

where $\boldsymbol{\beta} = \boldsymbol{\beta}(\tau) = (\beta_1(\tau), ..., \beta_p(\tau))^T \in \mathbb{R}^p$ is the regression coefficient. Koenker, R. et al. [2,3] provided a detailed discussion of quantile regression modeling in terms of methodology, theory, and computation. Because the loss function of quantile regression is not smooth, its computational complexity increases dramatically. Considering that the loss function is not derivable, Horowitz [4] used the smooth function to approximate the indicator function, in order to smooth the objective function. This method has also been applied to deal with other quantile regression-related problems, in which de Castro, L. et al. [5] used a smooth moment estimating equation to estimate the parameter in quantile regression;

Chen et al. [6] used a smoothing method to study the quantile regression model with constraints; Galvao, A.F. and Kato, K. [7] studied the smoothing estimation of fixed effects in quantile regression models for panel data; Whang [8] discussed the empirical likelihood estimation of quantile regression models by using a smoothing approach. Although the above literature has solved the problem that the objective function is not derivable, it cannot guarantee the concavity of the objective function. Thus, there is no guarantee that the result is a global optimal solution. Recently, Fernandes et al. [9] proposed a convolutional smoothing method for estimating fixed dimensional parameters of the quantile regression model, based on which the loss function is quadratically derivable and convex, and the method outperforms the original smoothing estimation in terms of estimation accuracy. Under the complete observation data, He [10] used the convolutional smoothing method to estimate the parameters of the high-dimensional quantile regression model, and the loss function after smoothing is a quadratically derivable and convex function. When solving the minimum point of the smoothed objective function numerically, the gradient descent algorithm [10] is used to replace the quantile regression with the least squares estimation, which effectively shortens the computation time and improves the estimation accuracy.

In empirical studies, it is frequently observed that variables of interest are subject to censoring. For instance, in the study [11] on the survival times of AIDS patients, it was found that 43% of the data were right-censored. For parameter estimation in quantile regression models for right-censored data, we refer to Ying, Z. et al. [12], Honoré, B. et al. [13], Portnoy, S. [14], Peng, L. [15], and Yuan, X. et al. [16]. Parameter estimation in quantile regression models for censored data [17–19] has been extensively studied, including right-censored data. In the quantile regression for right-censored data, the problem of non-smooth loss functions still exists, and it has already been studied [20–23]. For the right-censored quantile regression model with fixed dimensional parameters, Peng, L. and Huang, Y. [20], Xu, G. et al. [21], Cai, Z. and Sit, T. [22], Kim, K.H. [23] considered the smoothing estimation methods, respectively. For the high-dimensional large-sample case of the censored quantile regression model, Wu, Y. et al. [24], He, X. et al. [25], and Fei, Z. et al. [26] used a similar method to Fernandes et al. [9] to smooth the estimating equations, which improved the estimation accuracy over classical parameter estimation methods. However, these smoothing estimation methods for high-dimensional censored quantile regression need to grid the quantiles as $0 = \tau_1 < \cdots < \tau_j < \tau_{j+1} \cdots < \tau_k = 1$, to ensure that the approximation error of the estimation equation will not be too large. Furthermore, we need to use the estimates of $\boldsymbol{\beta}(\tau_1), \cdots, \boldsymbol{\beta}(\tau_j)$ before estimating $\boldsymbol{\beta}(\tau_{j+1})$, and the estimation accuracy depends on the number of segmentation points, which also increases the computational complexity.

Considering the censored data, this paper extends the convolutional smoothing method [10] to the high-dimensional censored quantile regression model and proposes the coefficient estimation of the censored quantile regression model based on the convolutional smoothing method. Under certain conditions, the loss function of the smoothed censored quantile regression model is quadratically derivable and globally convex, and the gradient-based iterative algorithm could be used to calculate the regression parameters. In this paper, the bias of the smoothing estimation of the censored quantile regression is characterized under certain conditions, and the speed of convergence, the Bahadur–Kiefer representation, and the Berry–Esseen upper bound of the smoothing estimation of the censored quantile regression are established under high-dimensional and large-sample conditions. Moreover, compared with the classical parameter estimation method, which has different estimation accuracies at different quantile points, the smoothing estimation method basically maintains the same estimation accuracies at each quantile point, and the proposed method greatly saves the computation time under high-dimensional large-sample

condition. In summary, the key contributions and significance of this article are as follows: (1) To our knowledge, this article is the first to apply the convolutional smoothing method to high-dimensional censored data regression analysis. Compared with the non differentiability of the objective function in classical censored quantile regression, the objective function in this method is second-order differentiable, which brings great convenience to the establishment of gradient based calculation algorithms and the discussion of theoretical properties. (2) For high-dimensional data scenarios, under certain regularization conditions, this paper also establishes the asymptotic consistency, asymptotic normality, and other theoretical properties of the proposed smooth estimator, ensuring the good properties of statistical inference.

This paper is organized as follows. Section 2 proposes a convolutional smoothing estimation method for high-dimensional censored quantile regression models, and gives the asymptotic properties of the smoothing estimation. In Section 3, numerical simulations of the smoothing estimation and the classical parameter estimation are carried out for the low and high dimensional cases, and the estimation accuracy and computational speed of the smoothing method in the censored quantile regression analysis are discussed. The discussion and conclusion are given in Sections 4 and 5, and then a detailed proof of the asymptotic property is given in Appendix A.

2. Methods

In this paper, we consider a right-censored quantile regression model; i.e., we assume that the response variable y in the model (1) is right-censored, in which case the observed variables are $z = \min(y, c), \delta = \mathbb{I}(y \leq c)$, where c is the censored variable. The observations are denoted as $\{(z_i, \boldsymbol{x}_i, \delta_i), i = 1, ..., n\}$. Then, the estimation of the parameters in the censored quantile regression model is defined as

$$\hat{\boldsymbol{\beta}}(\tau) \in \min_{\boldsymbol{\beta} \in \mathbb{R}^p} Q(\boldsymbol{\beta}) = \min_{\boldsymbol{\beta} \in \mathbb{R}^p} \frac{1}{n} \sum_{i=1}^{n} \frac{\delta_i}{1 - \hat{G}(z_i)} \rho_\tau(z_i - \boldsymbol{\beta}^T \boldsymbol{x}_i), \tag{2}$$

where $\rho_\tau(u) = u\{\tau - \mathbb{I}(u < 0)\}$ is the τ quantile loss function, and $G(\cdot)$ is the distribution function of the censored variable c_i, and $\hat{G}(\cdot)$ is the estimate of $G(\cdot)$.

Convolutional smoothing estimation for high-dimensional quantile regression models has been studied in the literature [10] by He et al. In this paper, we apply the method to censored data. Let $K(\cdot)$ be the kernel function with integral 1, and h be the window width. Denote the following:

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

Then, the objective function for the smoothing estimation of the censored quantile regression can be written as

$$Q_h(\boldsymbol{\beta}) = \frac{1}{n} \sum_{i=1}^{n} \frac{\delta_i}{1 - \hat{G}(z_i)} \ell_h(z_i - \boldsymbol{\beta}^T \boldsymbol{x}), \tag{3}$$

where $\ell_h(u) = (\rho_\tau * K_h)(u) = \int_{-\infty}^{+\infty} \rho_\tau(v) K_h(v - u) \mathrm{d}v$ and "$*$" denotes the convolution operator. The convolutional smoothing estimate (denoted SCQ) of the censored quantile regression model is defined as the point of minimum of $Q_h(\boldsymbol{\beta})$, $\hat{\boldsymbol{\beta}}_h$. For any $\boldsymbol{\beta} \in \mathbb{R}^p$, write $Q_h^*(\boldsymbol{\beta}) = \mathbb{E} Q_h(\boldsymbol{\beta})$, with

$$\boldsymbol{\beta}_h^*(\tau) \in \mathrm{argmin}_{\boldsymbol{\beta} \in \mathbb{R}^p} Q_h^*(\boldsymbol{\beta}). \tag{4}$$

For ease of presentation and without ambiguity, $\boldsymbol{\beta}^*(\tau) \in \operatorname{argmin}_{\boldsymbol{\beta}\in\mathbb{R}^p} \mathbb{E}Q(\boldsymbol{\beta})$ and $\boldsymbol{\beta}_h^*(\tau)$ are abbreviated as $\boldsymbol{\beta}^*$ and $\boldsymbol{\beta}_h^*$, respectively.

It is easy to find that the loss function of the (3) is quadratically derivable, and its gradient and Hessian matrix can be expressed, respectively, as follows:

$$\nabla Q_h(\boldsymbol{\beta}) = \frac{1}{n}\sum_{i=1}^{n} \frac{\delta_i}{1-\hat{G}(z_i)}\{\mathcal{K}_h(\boldsymbol{\beta}^T\boldsymbol{x}_i - z_i) - \tau\}\boldsymbol{x_i}, \tag{5}$$

$$\nabla^2 Q_h(\boldsymbol{\beta}) = \frac{1}{n}\sum_{i=1}^{n} \frac{\delta_i}{1-\hat{G}(z_i)} K_h(z_i - \boldsymbol{\beta}^T\boldsymbol{x}_i)\boldsymbol{x_i}\boldsymbol{x_i}^T.$$

As long as the kernel function $K(\cdot)$ is non-negative, for any window width $h > 0$, $Q_h(\boldsymbol{\beta})$ is a convex function and $\boldsymbol{\beta}_h^* = \boldsymbol{\beta}_h^*(\tau)$ satisfies the first-order condition $\nabla Q_h(\boldsymbol{\beta}_h^*) = 0$.

Remark 1. *When estimating the distribution of the censored variable c_i, we can use KM estimation to estimate $\hat{G}(\cdot)$. If we assume that the form of the distribution $G(\cdot)$ of the censored variable c_i is known, even if there is a mis-specification of the distribution, we find through the subsequent simulation that the estimation of the distribution of the censored variable by the parametric method is better than that of the distribution of the censored variable by using the KM estimation before smoothing the estimation of the regression parameter in most cases; thus, we rewrite the distribution function of the censored variable c_i as follows $\hat{G}(\cdot) = G(\cdot, \hat{\boldsymbol{\theta}}_n)$. Where the parameter vector $\boldsymbol{\theta}_n \in \mathbb{R}^p$, $\hat{\boldsymbol{\theta}}_n$ is the maximum likelihood estimate of $\boldsymbol{\theta}_n$. The parametric distribution form is used for estimation in both the proof and the hypotheses.*

To give theoretical results, we assume that the covariate $\boldsymbol{x}$ has been centered. Given the vectors $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^p$, $\boldsymbol{u}^T\boldsymbol{v}$ and $\langle \boldsymbol{u}, \boldsymbol{v}\rangle$ both represent their inner product: $a \vee b = \max\{a, b\}$, where a and b are constants. The $\|\cdot\|_q (1 \le q < \infty)$ denotes the ℓ_q-paradigm, i.e., $\|\boldsymbol{u}\|_q = (\sum_{i=1}^p |u_i|^q)^{1/q}$, and $\|\boldsymbol{u}\|_\infty = \max_{1\le i\le p}|u_i|$, where u_i denotes the ith element of the p-dimensional real vector $\boldsymbol{u}$. Given a semi-positive definite matrix $\boldsymbol{\Sigma} \in \mathbb{R}^{p\times p}$, define $\|\boldsymbol{u}\|_{\boldsymbol{\Sigma}} = \|\boldsymbol{\Sigma}^{1/2}\boldsymbol{u}\|_2$ for any vector $\boldsymbol{u} \in \mathbb{R}^p$. For all real numbers $r \ge 0$, define $\mathbb{B}^p(r) = \{\boldsymbol{\beta} \in \mathbb{R}^p : \|\boldsymbol{\beta}\|_2 \le r\}$ and $\mathbb{S}^{p-1}(r) = \{\boldsymbol{\beta} \in \mathbb{R}^p : \|\boldsymbol{\beta}\|_2 = r\}$. For two non-negative sequences $\{a_n\}_{n\ge1}$ and $\{b_n\}_{n\ge1}$, $a_n \lesssim b_n$ denotes the existence of a constant $C > 0$ independent of n such that $a_n \le Cb_n$. $a_n \gtrsim b_n$ is equivalent to $b_n \lesssim a_n$. $a_n \asymp b_n$ is equivalent to $a_n \lesssim b_n$ and $b_n \lesssim a_n$ holds simultaneously. The assumptions required for the theorem are as follows.

***A*1**. The non-negative kernel function $K(\cdot)$ satisfies $K(u) = K(-u)$, with upper bound $\kappa := \sup_{u\in\mathbb{R}} K(u) < +\infty$, and $\kappa_k := \int_{-\infty}^{+\infty} |u|^k K(u)\mathrm{d}u < +\infty$, where $k = 1, 2$.

***A*2**. The regression error term ε on the conditional density function $f_{\varepsilon|\boldsymbol{x}}(\cdot)$ on $\boldsymbol{x}$ satisfies the Lipschtiz condition; i.e., there exists a constant $L > 0$ such that $\forall u, v \in \mathbb{R}$, there is $|f_{\varepsilon|\boldsymbol{x}}(u) - f_{\varepsilon|\boldsymbol{x}}(v)| \le L|u - v|$ which holds almost everywhere. There exist real numbers $\underline{f} > 0$ such that $f_{\varepsilon|\boldsymbol{x}}(0) \ge \underline{f}$ holds almost everywhere for any $\boldsymbol{x}$.

***A*3**. There exist positive constants C_1 and K_j such that

$$\frac{\partial G(z;\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_j} \le K_j(z), E\{K_j^2(z)\} \le C_1 < \infty. (j = 1, ..., d) \tag{6}$$

Denote $z_{(i)}$ as the ith order statistic of $\{z_i\}$, and $\delta_{(i)}$ as the corresponding indicator function. $z_{(i)}$ and $\delta_{(i)}$ satisfy

$$\mathbb{P}(\delta_{(n)} = 1 | z_{(n)}) > 0. \tag{7}$$

A4. The covariate $\boldsymbol{x}$ obeys a subexponential distribution—i.e., there exists $v_0 > 0$ such that for any $\boldsymbol{u} \in \mathbb{S}^{p-1}$ and $t \geq 0$, there are $\mathbb{P}\{|\boldsymbol{u}, \boldsymbol{\omega}| \geq v_0 t\} \leq e^{-t}$, where $\boldsymbol{\Sigma} = \mathbb{E}\{\boldsymbol{x}\boldsymbol{x}^T\}$ is positive definite with $\boldsymbol{\omega} = \boldsymbol{\Sigma}^{-1/2}\boldsymbol{x}$.

With $\boldsymbol{\omega} = \boldsymbol{\Sigma}^{-1/2}\boldsymbol{x}$ and positive integer k, we define $m_k = \sup_{\boldsymbol{u}\in\mathbb{S}^{p-1}} \mathbb{E}|\langle \boldsymbol{u}, \boldsymbol{\omega}\rangle|^k$. The following theorems can be obtained.

Theorem 1 (Upper bound on the estimation error)**.** *Suppose the condition* ***A1***–***A4*** *holds for any real number $t > 0$. If h satisfies the constraints $\underline{f}^{-1} m_3^{1/2} v_0^* \sqrt{P_d(p+t)/n} \lesssim h \lesssim \underline{f} m_3^{-1/2}$, $P_d = \frac{\sum_i \delta_i}{n}$ is the uncensored proportion, then the convolutional smoothing estimate $\widehat{\boldsymbol{\beta}}_h$ satisfies the boundary conditions*

$$\mathbb{P}\left\{ \|\widehat{\boldsymbol{\beta}}_h - \boldsymbol{\beta}^*\|_{\boldsymbol{\Sigma}} \leq \frac{R}{\underline{f}} \left[v_0^* \sqrt{\frac{log(log(\frac{1}{h} \vee 1)) + p + t}{n}} + L\kappa_2 h^2 \right] \right\} \geq 1 - 2e^{-t}, \tag{8}$$

where $v_0^* = C_0 v_0$. C_0 and R are positive constants.

The upper bounds on the estimation error $v_0^* \sqrt{\frac{log(log(\frac{1}{h} \vee 1)) + p + t}{n}}$ and $L\kappa_2 h^2$ can be interpreted as the prediction bias and the speed of convergence. A smaller h leads to smaller deviations after smoothing, but an h that is too small could result in overfitting and slow convergence rates. According to Theorem 1, h satisfies the condition $\sqrt{(p+t)/n} \lesssim h \lesssim 1$. In order to obtain a non-asymptotic Bahadur representation of the smoothing estimation, we tend to replace ***A4*** with ***A4****.

A4*. The covariate $\boldsymbol{x}$ obeys a sub-Gaussian distribution—i.e. there exists $v_1 > 0$ such that for any $\boldsymbol{u} \in \mathbb{S}^{p-1}$ and $t \geq 0$, there is a $\mathbb{P}\{|\boldsymbol{u}, \boldsymbol{\omega}| \geq v_1 t\} \leq 2e^{-t^2/2}$, where $\boldsymbol{\Sigma} = \mathbb{E}\{\boldsymbol{x}\boldsymbol{x}^T\}$ is positive definite with $\boldsymbol{\omega} = \boldsymbol{\Sigma}^{-1/2}\boldsymbol{x}$.

Theorem 2 (non-asymptotic Bahadur representation)**.** *Assuming conditions* ***A1***–***A3*** *and* ***A4**** *hold, $\sup_{u\in\mathbb{R}} f_{\varepsilon|\boldsymbol{x}}(u) \leq \bar{f}$ holds almost everywhere. For any real number $t > 0$, h satisfies the constraint $\underline{f}^{-1} m_3^{1/2} v_1^* \sqrt{P_d(p+t)/n} \lesssim h \lesssim \underline{f} m_3^{-1/2}$. Let $\boldsymbol{J}_h = \nabla^2 Q_h(\boldsymbol{\beta}^*) = \mathbb{E}\{[\frac{\delta}{1-G(z)}]^2 K_h(\varepsilon)\boldsymbol{x}\boldsymbol{x}^T\}$; then,*

$$\begin{aligned} \mathbb{P}\left\{ \left\| \boldsymbol{J}_h(\widehat{\boldsymbol{\beta}}_h - \boldsymbol{\beta}^*) - \frac{1}{n}\sum_{i=1}^{n} \{\tau - \mathcal{K}_h(-\varepsilon_i)\}\boldsymbol{x}_i \right\|_{\boldsymbol{\Sigma}} \right. & \leq R\left(\frac{p+t}{nh^{1/2}} + h^{3/2} \cdot \sqrt{\frac{p+t}{n}} + h^4 \right) \Bigg\} \\ & \geq 1 - 3e^{-t}, \end{aligned} \tag{9}$$

where the real number $R > 0$ is a constant independent of p and n.

Theorem 2 allows for establishing the limiting distribution of the estimators. Based on the non-asymptotic representation in Theorem 2, we establish the Berry–Esseen upper bound for smoothing estimators.

Theorem 3 (Berry–Esseen upper bound)**.** *Assuming that the conditions* ***A1***–***A3*** *and* ***A4**** *hold, $\sup_{u\in\mathbb{R}} f_{\varepsilon|\boldsymbol{x}}(u) \leq \bar{f}$ holds almost everywhere, and h satisfies the condition that for any real number $t > 0$, there exists $\sqrt{\frac{p+t}{n}} \lesssim h \lesssim 1$. Then,*

$$\Lambda_{n,p}(h) := \sup_{x\in\mathbb{R}, \boldsymbol{a}\in\mathbb{R}^p} \left| \mathbb{P}(n^{1/2}\sigma_h^{-1}\langle \boldsymbol{a}, \widehat{\boldsymbol{\beta}}_h - \boldsymbol{\beta}^*\rangle \leq x) - \Phi(x) \right| \lesssim \frac{p + \log n}{(nh)^{1/2}} + n^{1/2}h^2, \tag{10}$$

where $\sigma_h^2 = \sigma_h^2(\boldsymbol{a}) = \boldsymbol{a}^T \boldsymbol{J}_h^{-1} \mathbb{E}\left[\left(\frac{\delta}{1-G(z)}\right)^2 \{\mathcal{K}_h(-\varepsilon) - \tau\}^2 \boldsymbol{x}\boldsymbol{x}^T\right] \boldsymbol{J}_h^{-1} \boldsymbol{a}$, $\Phi(\cdot)$ *denotes the standard normal distribution function.*

Further, if $f_{\varepsilon|\boldsymbol{x}}(\cdot)$ *is quadratically continuously derivable and satisfies* $|f'_{\varepsilon|\boldsymbol{x}}(u) - f'_{\varepsilon|\boldsymbol{x}}(v)| \leq l_2(\boldsymbol{x})|u - v|$ *for any real number* $u, v \in \mathbb{R}$*, where the function* $l_2 : \mathbb{R}^p \to \mathbb{R}^+$ *satisfies* $\mathbb{E}\{l_2^2(\boldsymbol{x})\} \leq C$ *for some positive constant* C*, then*

$$\sup_{x \in \mathbb{R}, \boldsymbol{a} \in \mathbb{R}^p} |\mathbb{P}(n^{1/2}\sigma_h^{-1}\langle \boldsymbol{a}, \widehat{\boldsymbol{\beta}}_h - \boldsymbol{\beta}^* + 0.5\kappa_2 h^2 \boldsymbol{J}_h^{-1} \mathbb{E}\{f'_{\varepsilon|\boldsymbol{x}}(0)\boldsymbol{x}_i\}\rangle) - \Phi(x)| \lesssim \frac{p + \log n}{(nh)^{1/2}} + (p + \log n)^{1/2} h^{3/2} + n^{1/2} h^4. \quad (11)$$

Theorem 3 proves that when h is chosen in the appropriate range and $n, p \to \infty$, the linear combination of $\widehat{\boldsymbol{\beta}}_h$ is estimated to be asymptotically normal. According to Theorem 3, the optimal $h = \{(p + \log n)/n\}^{2/5}$ in the sense that minimizes the right hand of (10), and the error is approximated as $(p + \log n)^{4/5} n^{-3/10}$. If $p^8/n^3 \to 0$, for any given vector $\boldsymbol{a} \in \mathbb{R}^p$, $n^{1/2}\langle \boldsymbol{a}, \widehat{\boldsymbol{\beta}}_h - \boldsymbol{\beta}^* \rangle$ is asymptotically normal.

Remark 2. *The assumptions* ***A1***, ***A2***, ***A4*** *and* ***A4**** *are commonly used in convolutional smoothing estimation of high-dimensional quantile regression models in fully observed data [10]. Condition* ***A3*** *refers to the assumptions concerning the distribution of the censoring variable c. Note that* $\mathbb{P}(\delta_{(n)} = 1|z_{(n)}) = \mathbb{P}(z_{(n)} \leq c|z_{(n)}) = \int_{z_{(n)}}^{+\infty} \mathrm{d}G(s) = 1 - G(z_{(n)})$*, assuming that* $\mathbb{P}(\delta_{(n)} = 1|z_{(n)}) > 0$ *is equivalent to* $G(z_{(n)}) < 1$*—i.e., the probability that the largest observation* $z_{(n)}$ *equals to the true variable of interest is not zero.This is a commonly used condition in statistical inference for censored data [27,28], and this condition can avoid the situation where a large number of observations are censored data. Assuming that* (6) *provides a local smoothing condition for* $G(\cdot)$ *in the neighborhood* θ*, the validity of this assumption could be verified intuitively for many commonly used distribution functions* $G(\cdot)$*.*

3. Numerical Simulation

In this section, the smoothing estimation and classical parameter estimation of the quantile regression model for censored data are numerically simulated, and two cases of low and high dimensionality are considered. Estimation of (2), as proposed in the literature [17], was chosen as the classical parameter estimation of the censored quantile regression model. Notice that the objective function of the classical parameter estimation for censored quantile regression model can be rewritten as $Q(\boldsymbol{\beta}) = \sum_{i=1}^n \rho_\tau[\frac{\delta_i}{1-G(z_i)}(z_i - \boldsymbol{\beta}^{\boldsymbol{T}}\boldsymbol{x})]$. Therefore, when calculating the regression parameters, the censoring problem is transformed into a non-censoring problem, and the objective function of the smoothing estimation for censored quantile regression is rewritten as $Q_h(\boldsymbol{\beta}) = \sum_{i=1}^n \ell_h[\frac{\delta_i}{1-G(z_i)}(z_i - \boldsymbol{\beta}^{\boldsymbol{T}}\boldsymbol{x})]$. The Gaussian kernel and window width $h = \{(p + \log n)/n\}^{2/5}$ are taken as the kernel function and window width, respectively, for the smoothing estimation of the censored quantile regression.

3.1. Model Setting and Evaluation Indicators

In the simulation, the covariates $\boldsymbol{x}_i = (x_{i1}, ..., x_{ip})^T \in \mathbb{R}^p$ are generated from different distributions to simulate different types of variables commonly found in real data. The error term ε_i is generated by three different distributions, specifically, by drawing independent identically distributed random numbers $\tilde{\varepsilon}_i$ with capacity n, and letting $\varepsilon_i = \tilde{\varepsilon}_i - F_{\tilde{\varepsilon}_i}^{-1}(\tau)$, where $\tilde{\varepsilon}_i$ obeys the distributions : (i) $t(4)$; (ii) $\chi^2(1)$; (iii) $Laplace(0, 1)$. Let the regression

coefficient $\boldsymbol{\beta} = (1, ..., 1)^T \in \mathbb{R}^p$, given the quantile $\tau \in (0, 1)$; then, the response variable y_i is modeled by

$$y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_p x_{ip} + \varepsilon_i.$$

For both low- and high-dimensional models, the right censoring variable is set as $c_i \sim const + E(pa)$, where $const$, pa are unknown parameters, which can take different values to make the censoring ratio of the response variable y_i up to the set 15%, 30%, or 45%. In the actual simulation, in order to obtain the value of $G(z_i)$, the parameters $const$, pa are estimated by using maximum likelihood estimation in the simulations of Sections 3.2 and 3.3. In Section 3.4, we discuss the smoothing estimation under the misspecification for the distribution of censored variables. KM estimation is also taken into consideration. Let the number of simulation repetitions be K, and for the parameter estimates $\widehat{\boldsymbol{\beta}}$ in the kth simulation, write

$$SE_k = \frac{1}{p}\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\|_2^2,$$

Then, we can use

$$DMSE = \frac{1}{K}\sum_{k=1}^{K} SE_k,$$

to evaluate the performance of classical parameter estimation for censored quantile regression models (CQ) and smoothing estimation methods (SCQ). In the actual simulation, we set $K = 500$.

3.2. Low-Dimensional Performance of Regression Smoothing Estimates for Censored Quantiles

In the low-dimensional numerical study of smoothing estimation, the number of covariates is set to be $p = 5$ and sample sizes are 100, 200, and 500. In order to assess the performance of smoothing method in the low-dimensional case, the generation of covariates is categorized into three cases.

Case 1: The p-dimensional covariates are generated from multivariate uniform distribution on $[10, 20]^p$, and the covariance matrix is a unit matrix;
Case 2: The p-dimensional covariate is generated from multivariate uniform distribution on $[10, 20]^p$, where $\boldsymbol{\Sigma} = (0.5^{|j-k|})_{1\le j,k\le p}$ is the covariance matrix;
Case 3: The p-dimensional covariates consist of a mixture of distributions, where the first two dimensional covariates are generated by a multivariate uniform distribution on $[10, 20]^2$, with a covariance matrix of $\boldsymbol{\Sigma}_1 = (0.5^{|j-k|})_{1\le j,k\le 2}$. The covariates in the posterior three dimensions are generated from $N(\boldsymbol{\mu}, \boldsymbol{\Sigma}_2)$ with mean $\boldsymbol{\mu} = (11, 12, 13)$ and covariance matrix $\boldsymbol{\Sigma}_2 = (\sigma_{jk})_{1\le j,k\le p}$, where

$$\sigma_{jk} = \begin{cases} 0.2^{|j-k|}, 1 \le j \ne k \le p, \\ 3 \quad , 1 \le j = k \le p. \end{cases}$$

Tables 1–3 show the simulation results when the covariates are generated according to the three scenarios, where CP denotes the censoring ratio of the response variable, n is the sample size, and columns 3–12 show the results of CQ and SCQ at different quantiles. From the estimation results, when the regression errors are generated by symmetric distributions, i.e., t and *Laplace* distributions, SCQ has higher accuracy than CQ, especially at the lower and higher quantiles. When the regression error term is generated by the χ^2 distribution, the estimation accuracy of CQ decreases as τ increases from a global perspective, and

the estimation accuracy of SCQ is much better than that of CQ at the higher quantiles, although CQ is better than SCQ at the lower quantiles. This may be because the density function of the χ^2 distribution is biased and the observations are excessively clustered in the lower quantiles, so that CQ decreases in estimation accuracy as the number of quantiles increases, while SCQ maintains better estimation accuracy. It can be seen from Tables 1–3 that SCQ is more stable than CQ, regardless of whether the error terms follow symmetric or asymmetric distribution. Specifically, the estimation accuracy of SCQ is almost the same in all quantiles, while that of CQ fluctuates with the change of τ, especially in the case of the asymmetric distribution of the error terms. Overall, the estimation accuracy of CQ depends greatly on the value of τ, the size of the censoring ratio, and the distribution of the error term, while the estimation effect of SCQ is minimally affected by these factors and shows good robustness.

Table 1. The DMSE of CQ and SCQ estimators for the low-dimensional model with covariate X generated from Case 1, where the values in columns 3–12 are DMSE $\times$ 10^4.

CP%	*n*	$\tau = 0.1$		$\tau = 0.3$		$\tau = 0.5$		$\tau = 0.7$		$\tau = 0.9$	
		CQ	**SCQ**	**CQ**	**SCQ**	**CQ**	**SCQ**	**CQ**	**SCQ**	**CQ**	**SCQ**
$\tilde{\varepsilon}_i \sim t(4)$											
15	100	397	116	143	112	112	109	137	116	387	114
	200	180	58	63	54	51	56	66	52	179	56
	500	71	24	28	23	21	22	27	21	73	21
30	100	489	151	190	134	144	133	170	131	518	135
	200	228	74	85	70	65	69	82	67	230	65
	500	95	34	30	27	25	27	32	26	91	25
45	100	687	200	224	179	181	178	220	178	628	173
	200	302	97	103	91	84	88	107	81	291	77
	500	118	45	38	39	34	35	39	31	118	32
$\tilde{\varepsilon}_i \sim \chi^2(1)$											
15	100	2	128	13	135	62	126	212	122	1103	123
	200	0	78	6	78	29	79	118	75	591	77
	500	0	39	2	40	13	39	48	39	236	37
30	100	2	164	17	155	77	164	270	155	1397	144
	200	0	97	7	98	37	97	136	95	718	90
	500	0	51	3	54	15	52	55	48	305	44
45	100	3	208	23	198	89	204	321	189	1746	194
	200	1	127	10	120	46	124	162	117	860	107
	500	0	68	4	69	20	62	73	59	366	55
$\tilde{\varepsilon}_i \sim Laplace(0,1)$											
15	100	603	235	246	230	213	235	258	223	564	234
	200	304	119	125	115	109	114	129	114	299	111
	500	119	52	50	49	43	46	52	45	109	46
30	100	743	322	343	302	271	291	336	288	762	270
	200	365	171	159	149	124	147	163	136	385	289
	500	140	74	62	66	53	60	59	54	138	54
45	100	983	441	410	374	370	364	435	361	964	330
	200	461	210	200	184	165	171	200	175	456	163
	500	195	101	76	80	65	81	76	71	187	64

Table 2. The DMSE of CQ and SCQ estimators for the low-dimensional model with covariate X generated from Case 2, where the values in columns 3–12 are DMSE $\times 10^4$.

CP%	*n*	$\tau = 0.1$		$\tau = 0.3$		$\tau = 0.5$		$\tau = 0.7$		$\tau = 0.9$	
		CQ	**SCQ**	**CQ**	**SCQ**	**CQ**	**SCQ**	**CQ**	**SCQ**	**CQ**	**SCQ**
						$\tilde{\varepsilon}_i \sim t(4)$					
15	100	661	173	241	168	180	179	245	183	667	173
	200	323	88	113	80	89	80	108	78	303	79
	500	128	33	46	32	36	32	47	34	119	33
30	100	840	230	266	219	226	205	288	208	847	209
	200	390	110	137	105	112	99	145	103	413	106
	500	160	47	53	42	44	42	55	42	151	40
45	100	1101	310	423	302	344	295	422	265	1210	275
	200	521	150	180	143	145	136	175	135	506	125
	500	186	62	69	53	55	52	74	50	210	52
						$\tilde{\varepsilon}_i \sim \chi^2(1)$					
15	100	3	190	21	195	97	199	352	194	1889	196
	200	1	118	10	113	50	114	195	110	996	115
	500	0	60	4	59	21	58	81	59	450	60
30	100	4	252	29	253	124	243	465	236	2219	236
	200	1	146	13	145	60	142	221	139	1181	136
	500	0	78	5	74	25	73	94	72	521	68
45	100	9	321	41	330	157	303	571	301	2653	307
	200	1	179	17	184	75	193	271	187	1578	179
	500	0	93	7	92	31	91	121	93	671	90
						$\tilde{\varepsilon}_i \sim Laplace(0,1)$					
15	100	1027	363	434	363	365	361	462	347	1028	356
	200	506	179	227	168	179	183	218	174	484	172
	500	195	75	86	67	76	68	85	70	204	71
30	100	1267	449	546	453	481	414	583	437	1366	443
	200	560	224	270	208	215	209	273	221	574	196
	500	241	93	112	85	94	86	100	86	244	83
45	100	1656	615	702	586	646	608	784	584	1642	564
	200	783	302	389	280	296	277	348	266	830	255
	500	305	122	133	115	114	116	136	105	313	101

Table 3. The DMSE of CQ and SCQ estimators for the low-dimensional model with covariate X generated from Case 3, where the values in columns 3–12 are DMSE $\times 10^4$.

CP%	*n*	$\tau = 0.1$		$\tau = 0.3$		$\tau = 0.5$		$\tau = 0.7$		$\tau = 0.9$	
		CQ	**SCQ**	**CQ**	**SCQ**	**CQ**	**SCQ**	**CQ**	**SCQ**	**CQ**	**SCQ**
						$\tilde{\varepsilon}_i \sim t(4)$					
15	100	863	259	315	255	257	253	302	252	841	249
	200	405	131	144	128	118	126	148	125	442	122
	500	166	52	56	50	43	49	54	48	164	46
30	100	1082	331	371	321	311	315	390	311	104	304
	200	490	168	172	159	141	153	181	149	528	144
	500	203	73	71	67	54	64	66	61	187	58
45	100	1443	467	518	442	421	431	522	425	1372	415
	200	653	226	233	209	187	202	247	196	712	186
	500	260	101	91	89	70	83	89	78	248	73

Table 3. *Cont.*

CP%	*n*	$\tau = 0.1$		$\tau = 0.3$		$\tau = 0.5$		$\tau = 0.7$		$\tau = 0.9$	
		CQ	**SCQ**	**CQ**	**SCQ**	**CQ**	**SCQ**	**CQ**	**SCQ**	**CQ**	**SCQ**
						$\tilde{\varepsilon}_i \sim \chi^2(1)$					
15	100	3	286	30	286	127	284	468	282	2529	277
	200	1	180	12	180	65	179	262	177	1290	172
	500	0	87	5	87	26	86	101	85	535	82
30	100	5	358	36	357	154	354	566	348	2992	335
	200	1	218	15	216	76	212	295	208	1551	199
	500	0	118	6	116	32	114	122	109	688	100
45	100	9	467	51	465	210	457	709	450	382	431
	200	2	281	21	279	97	274	374	265	1866	246
	500	0	154	8	152	41	148	161	139	907	124
						$\tilde{\varepsilon}_i \sim Laplace(0,1)$					
15	100	1372	553	619	545	537	539	615	535	1337	527
	200	638	259	276	252	229	247	269	244	649	241
	500	260	117	103	112	89	109	107	106	255	103
30	100	1757	718	780	693	655	678	766	664	1675	639
	200	780	358	334	337	295	325	336	315	819	304
	500	311	170	133	154	114	146	137	139	318	130
45	100	2274	985	1007	939	865	907	945	874	2145	822
	200	999	483	452	446	388	424	433	407	1064	386
	500	405	238	172	210	150	196	178	183	415	165

3.3. High-Dimensional Performance of Smoothing Estimators of Censored Quantile Regression

In high-dimensional large-sample numerical studies, the ratio of sample size to dimension is fixed at $n/p = 20$, the sample size is set at 1000–5000, and the step size is 500. In order to simulate the smoothing estimation of censored quantile regression with the change of dimension and sample size, the covariate generation is categorized into three cases.

Case 1: The p-dimensional covariate is generated from a multivariate uniform distribution on $[10, 20]^p$, with covariance matrix $\mathbf{\Sigma} = (0.5^{|j-k|})_{1 \le j,k \le p}$.
Case 2: The p-dimensional covariate is generated from a multivariate uniform distribution on $[10, 20]^p$, with covariance matrix $\mathbf{\Sigma'} = (\sigma_{jk})_{1 \le j,k \le p}$, where

$$\sigma_{jk} = \begin{cases} 0.2^{|j-k|}, 1 \le j \ne k \le p, \\ 3 \quad , 1 \le j = k \le p. \end{cases}$$

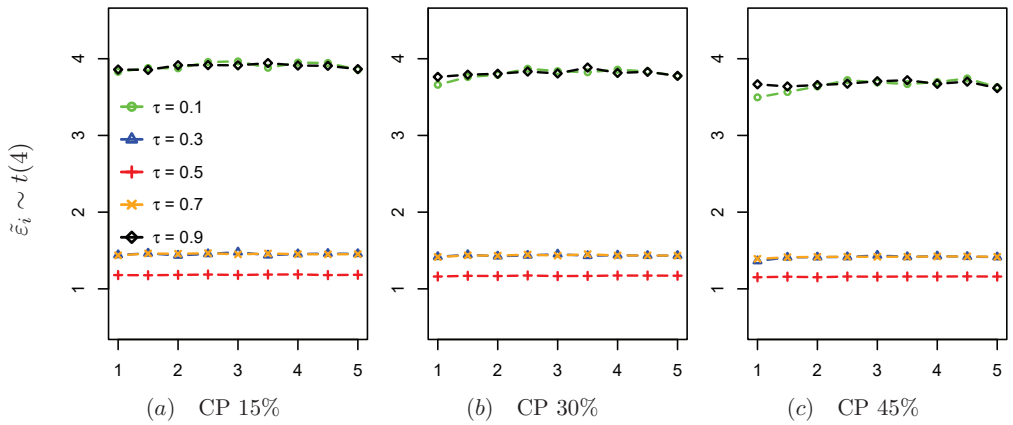
Case 3: The p-dimensional covariate consists of a mixture of distributions, where the first $[p/2]$-dimensional covariate is generated from a multivariate uniform distribution on $[10, 20]^{[p/2]}$, and the covariance matrix is $\mathbf{\Sigma}_1 = (0.5^{|j-k|})_{1 \le j,k \le [p/2]}$; the second $[p/2]+1$-dimensional covariate is generated by $N(\boldsymbol{\mu}, \mathbf{\Sigma}_2)$ with mean $\boldsymbol{\mu} = (11, ..., 30)$, where the jth component $\mu_j = 11 + \frac{20(j-1)}{[p/2]+1} (j = 1, \cdots, [p/2]+1)$, and covariance matrix $\mathbf{\Sigma}_2 = (\sigma_{jk})_{1 \le j,k \le [p/2]+1}$, where

$$\sigma_{jk} = \begin{cases} 0.2^{|j-k|}, 1 \le j \ne k \le [p/2]+1, \\ 3 \quad , 1 \le j = k \le [p/2]+1. \end{cases}$$

The ratio of DMSE between CQ and SCQ is firstly calculated, and the simulation results of the covariates under Case 1 are displayed in Figure 1. Since the results of the three covariate generation cases are very similar, we do not show the results of Case 2 and Case 3. The results in Figure 1 show that the ratio of DMSE estimated by CQ and SCQ is not significantly affected by changes in sample size and dimensionality. When the regression

error terms are generated by symmetric distributions, i.e., the *t*-distribution and *Laplace*-distribution, the DMSE ratios of the regression coefficients estimators between CQ and SCQ remain above one. This indicates that SCQ has a higher precision compared with CQ, especially in the case where the difference between the lower quantile of $\tau = 0.1$ and the upper quantile of $\tau = 0.9$ is more obvious, and the ratio of DMSE for regression coefficients' estimators between CQ and SCQ can reach three. When the error term is generated by the asymmetric χ^2 distribution, the ratio of DMSE between CQ and SCQ is less than one at the low quantile level, and CQ is superior to SCQ in this case. However, the ratio of DMSE between the two methods increases as τ increases, and the ratio reaches around 10 when $\tau = 0.9$. In order to clarify the reason for this phenomenon, we calculate the DMSE of two estimation methods when the error term is generated by the χ^2 distribution. As shown in Figure 2, CQ performs better when $\tau = 0.1$. As τ increases, the DMSE of CQ increases while its growth rate gradually increases. The CQ's DMSE is significantly higher than the SCQ's when $\tau = 0.9$, whereas the SCQ's DMSE stabilizes at each quantile point in a straight line. The observed phenomenon aligns with its lower-dimensional counterpart, underscoring that the estimation accuracy of the CQ depends on the magnitude of τ, the censoring ration of response variables and the distribution characteristics of the error term. In contrast, the estimation accuracy of the SCQ exhibits robustness against variations in these factors.

To assess the computational efficiency of the smoothing estimation, it is imperative to compare both the DMSE of CQ and SCQ estimations within the context of high-dimensional simulations, as well as the time expenditure associated with each estimation method. In Figure 3, the computational time ratios of CQ and SCQ are both greater than 1, which tends to increase with the augmentation of dimensionality and the sample size, and the computational time ratios of CQ and SCQ are more than 10 in all the cases when the sample size is $n = 5000$. The results of the covariate generation in Case 2 and Case 3, which are not shown here, are similar to those in Case 1. Combined with the previous study, it is clear that, when compared to CQ, SCQ significantly decreases calculation time and increases estimation accuracy in the majority of circumstances.



(*a*) CP 15% (*b*) CP 30% (*c*) CP 45%

Figure 1. *Cont.*

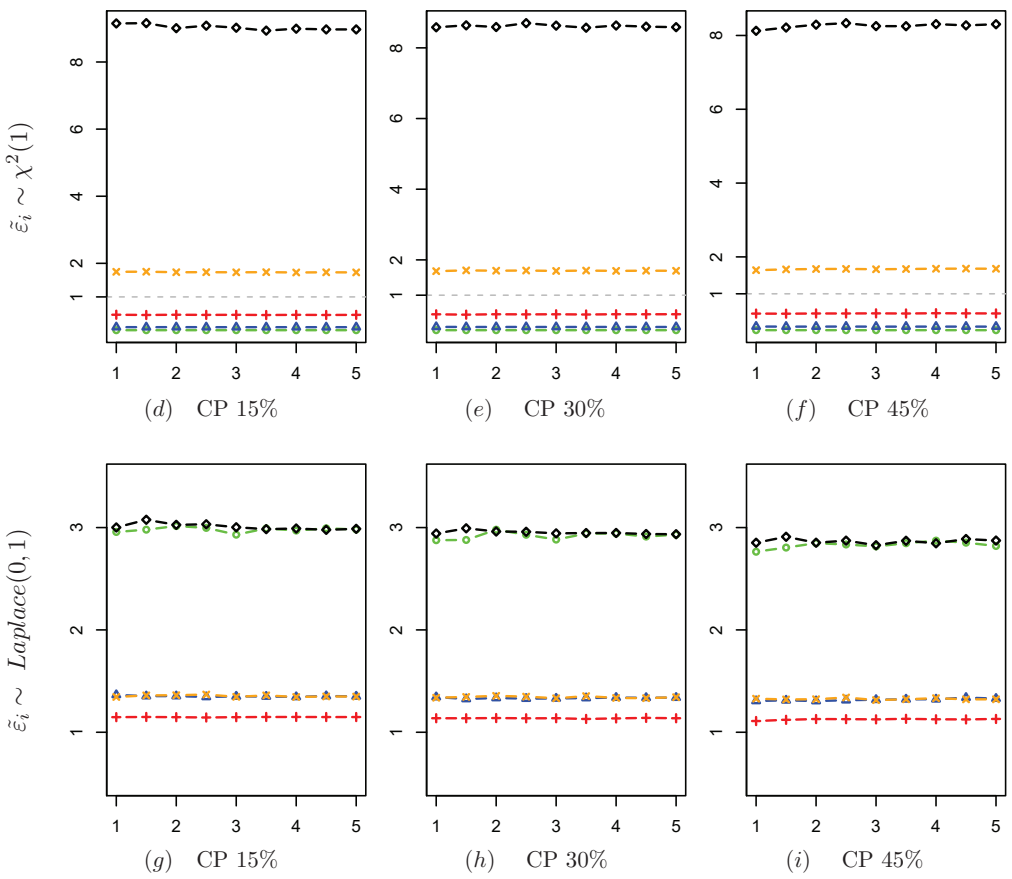


Figure 1. Estimation results for the high-dimensional model when covariate X is generated from Case 1. The horizontal coordinates of the plots indicate the sample size (in thousands), and the vertical coordinates indicate the ratio of DMSE for regression coefficients' estimators between CQ and SCQ.

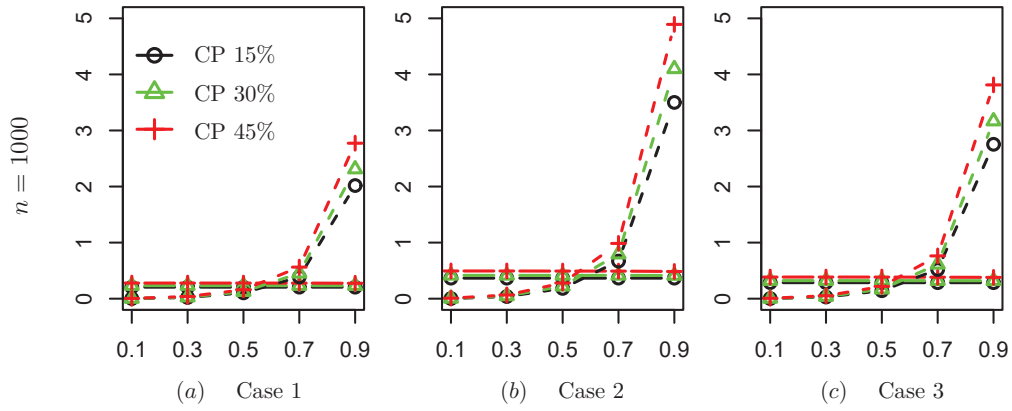


Figure 2. *Cont.*

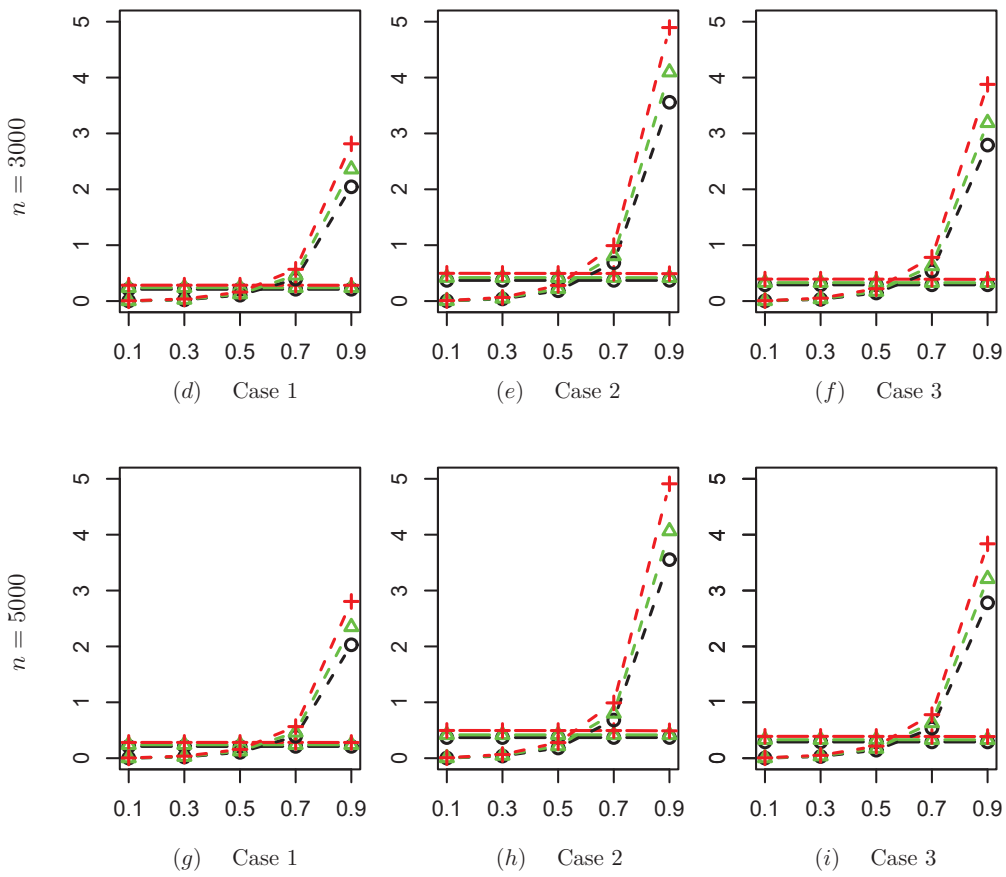


Figure 2. The DMSE of CQ and SCQ for three high-dimensional covariate generation cases when the error terms obey the χ^2 distribution, with the horizontal coordinates denoting the different quantile levels, and the vertical axes denoting the DMSEs of the CQ and SCQ estimations scaled up by a factor of 10, where the solid line denotes the SCQ, and the dashed line denotes the CQ.

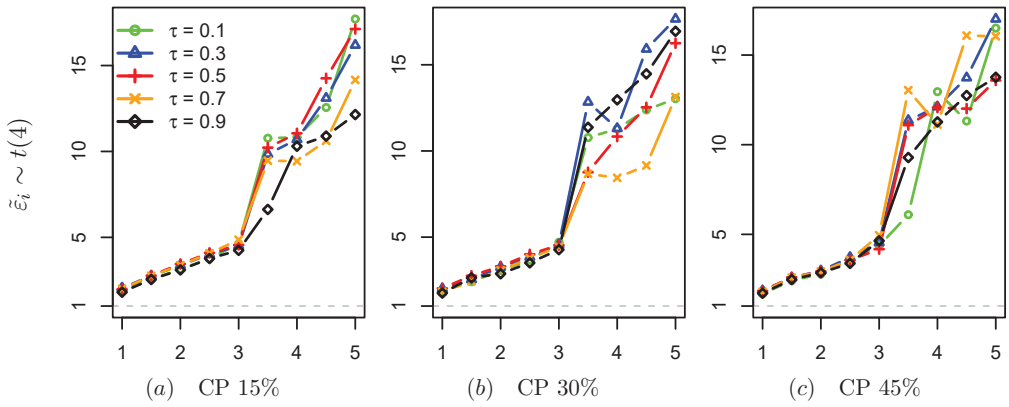


Figure 3. *Cont.*

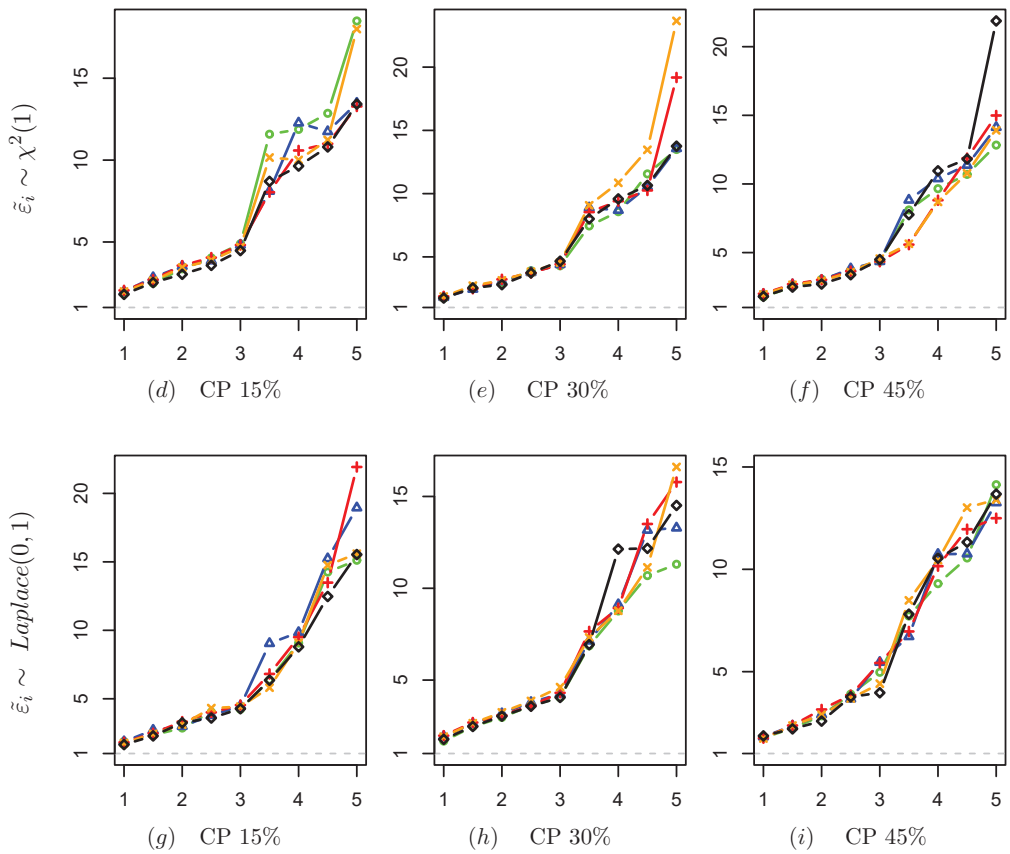


Figure 3. Simulation results under the high-dimensional model when the covariates are generated from Case 1, where the horizontal coordinate denotes the sample size (in thousands), and the vertical coordinate denotes the ratio of estimated running time between CQ and SCQ.

3.4. Robustness of Smoothing Estimates

In order to compare the effects under the misspecification for the distribution of censored variables and KM estimation of censored distributions on the smoothing estimation of parameters, we choose Case 1 of the covariate generation method in the low- and high-dimensional simulations for numerical research.

When the sample size is 200, the model coefficients are estimated using convolutional smoothing after assuming that the distribution of the censored variables is misclassified as Normal, Weibull, and Lognormal, and using the KM estimation of G with different conditions of τ quantiles, censoring proportions, and error terms. The simulation in Table 4 shows that the smoothing estimation using the misclassified distribution of the censored variables is more robust than the smoothing estimation using the KM estimation to estimate the distribution of the censored variables. Similar results are obtained for a sample size of 3000, as shown in Table 5.

The simulation shows that although there is a possibility of mis-setting in estimating the parameters of the distribution of the censored variables by using $\hat{G}(\cdot) = G(\cdot, \hat{\boldsymbol{\theta}}_n)$, this method is better and more robust than the smoothing of regression models after estimating the distribution of the censored variables by using the KM estimation. During the simulation process, it is also found in Tables 4 and 5 that when the censored variables are estimated in different distribution forms, the smoothing estimation errors of the model coefficients exhibit a minimal degree of variation, and the estimation accuracy is higher than that of using KM estimation to estimate the distribution of the censored variables and then carry out the smoothing estimation. Regarding computational efficiency, the smoothing estimation method, when applied in scenarios of distributional misclassification,

demonstrates a running time ratio to the SCQ that fluctuates within the range of 0.9 to 1.1. In contrast, the smoothing estimation method, once preceded by the KM estimation, incurs a significantly extended running time.

Table 4. Smoothing estimators of regression model parameters after misclassifying the distribution of censored variables as Normal, Weibull, and Lognormal, and after estimating G using KM estimation. The covariate X in the low-dimensional model is generated from Case 1, where the values in columns 2–13 are DMSE $\times 10^5$. The sample size of the simulation is fixed at 200.

τ	Normal			Weibull			Lognormal			KM Estimation		
	15%	30%	45%	15%	30%	45%	15%	30%	45%	15%	30%	45%
						$\tilde{\varepsilon}_i \sim t(4)$						
0.1	588	766	991	567	722	938	564	709	916	1366	1404	1448
0.3	554	682	861	552	677	858	552	677	858	1158	1192	1203
0.5	537	644	808	543	655	823	545	660	832	1123	1128	1157
0.7	524	618	774	536	639	800	539	648	815	1037	1051	1100
0.9	511	594	754	529	620	776	534	631	793	976	1000	1091
						$\tilde{\varepsilon}_i \sim \chi^2(1)$						
0.1	808	1006	1270	793	974	1227	791	966	1212	1786	1847	1982
0.3	803	995	1250	790	969	1215	789	962	1202	1788	1810	1902
0.5	792	969	1212	785	955	1191	785	953	1185	1716	1778	1856
0.7	772	922	1150	776	929	1151	778	933	1155	1521	1684	1820
0.9	740	858	1076	759	886	1088	765	898	1102	1305	1409	1472
						$\tilde{\varepsilon}_i \sim Laplace(0,1)$						
0.1	1213	1671	2138	1160	1560	2047	1151	1525	1999	1778	2023	2348
0.3	1141	1472	1835	1126	1451	1851	1125	1444	1851	1370	1595	1897
0.5	1102	1372	1695	1106	1392	1754	1110	1400	1777	1208	1434	1731
0.7	1071	1295	1593	1092	1344	1678	1099	1363	1717	1116	1326	1609
0.9	1034	1215	1513	1072	1285	1604	1083	1314	1655	1078	1304	1598

Table 5. Smoothing estimates of regression model parameters after misclassifying the distribution of censored variables as Normal, Weibull, and Lognormal, and after estimating G using KM estimation. The covariate X in the high-dimensional model is generated from Case 1, where the values in columns 2–13 are DMSE $\times 10^4$. The sample size of the simulation is fixed at 3000.

τ	Normal			Weibull			Lognormal			KM Estimation		
	15%	30%	45%	15%	30%	45%	15%	30%	45%	15%	30%	45%
						$\tilde{\varepsilon}_i \sim t(4)$						
0.1	1962	2272	2891	1952	2270	2945	2020	2274	2965	2634	3593	4324
0.3	1951	2260	2937	1940	2258	2925	2022	2263	2943	2570	3532	4343
0.5	1945	2253	2891	1935	2251	2913	2026	2255	2936	2489	3459	4373
0.7	1943	2249	2902	1933	2247	2900	2019	2253	2931	2394	3381	4273
0.9	1943	2239	2892	1932	2236	2876	2013	2239	2912	2369	3306	4172
						$\tilde{\varepsilon}_i \sim \chi^2(1)$						
0.1	2411	2427	2855	2391	2415	2844	2421	2445	2868	3605	3613	4105
0.3	2402	2439	2864	2392	2437	2844	2416	2448	2885	3390	3438	4035
0.5	2421	2435	2851	2391	2426	2842	2432	2447	2865	3308	3350	4062
0.7	2412	2446	2854	2392	2434	2843	2425	2465	2864	3307	3353	3955
0.9	2409	2442	2832	2390	2431	2825	2413	2444	2850	3262	3271	4009
						$\tilde{\varepsilon}_i \sim Laplace(0,1)$						
0.1	3857	4742	6153	3846	4739	6143	3861	4757	6165	3640	4442	5388
0.3	3840	4724	6096	3834	4711	6085	3853	4742	6101	3602	4349	5403
0.5	3838	4710	6051	3827	4692	6045	3841	4740	6067	3728	4252	5332
0.7	3835	4682	6018	3819	4675	6006	3851	4695	6024	3789	4169	5236
0.9	3834	4662	5963	3816	4654	5951	3849	4686	5975	3610	4097	5195

4. Discussion

In the smoothing estimation of censored quantile regression models, the distribution of the censored variables is usually unknown. The problem of the unknown distribution of the censored variables can be solved by estimating the density function of the censored variables through existing nonparametric methods, then determining the type of the censored distribution using the goodness-of-fit test, and accessing the unknown parameters in the distribution using the great likelihood method. In the numerical simulation , this paper also discusses using such a method to fit the unknown censored distribution $G(\cdot)$ to estimate $\hat{G}(\cdot) = G(\cdot, \hat{\boldsymbol{\theta}}_n)$, and then estimate the regression parameter $\boldsymbol{\beta}$.

We also discuss the parametric smoothing estimation method in the case of misspecification of the distribution of the censored variables, and compare it with the smoothing estimation method of the parameters after estimating the distribution of the censored variables $\hat{G}(\cdot)$ with KM estimation. The simulation results show that the smoothing estimation method is still more robust than the method of estimating the distribution of censored variables $\hat{G}(\cdot)$ with KM estimation even if there is a misspecification in the estimation of the distribution of the censored variables. Meanwhile, the smoothing estimation method is more robust than the classical censored quantile regression model.

Our research has certain limitations and there are some issues that need to be further explored. Firstly, we have performed some analysis and research on parameter smoothing estimation for quantile linear models; moreover, we can plan further research on parameter estimation and interval estimation for complex models such as generalized linear models. Secondly, our proof is based on the assumption that the form of censored distribution is known, and the theoretical proof by using a nonparametric model to estimate the censored distribution is still a challenging task. This requires understanding and knowledge in the theory of nonparametric statistics and probability limits.

5. Conclusions

In this paper, a convolutional smoothing estimation method for the censored quantile regression model is proposed to address the problem that the loss function is not derivable. Our method associates the convolutional smoothing estimation with the loss function of censored quantile regression, which is quadratically derivable, compared with the classical censored quantile regression estimation. Moreover, the smoothing estimation method for censored quantile regression models improves the estimation accuracy, computational speed, and robustness over the classical parameter estimation method. The contribution and significance of this paper can be summarized as follows:

1. The method establishes links between the convolutional smoothing method and the loss function of the censored quantile regression model, and the use of the non-negative kernel function ensures that the smoothed loss function is quadratically derivable and globally convex, which can be used to improve the computational speed by using the gradient-based iterative algorithm.
2. Theoretically, we characterize the bias of the smoothing estimation for censored quantile regression and establish the convergence rate, Bahadur–Kiefer representation, and Berry–Esseen upper bound of the smoothing estimation under high-dimensional and large-sample conditions.
3. The numerical simulation shows that the smoothing estimation method greatly reduces the computation time and improves the estimation accuracy in most cases, compared with the classical parameter estimation. In addition, the accuracy of CQ estimator is highly dependent on τ, censoring ratio(CP), and the distribution of error term, but the SCQ estimation is robust to these factors.

Author Contributions: Conceptualization, M.W.; methodology, M.W.; software, M.W.; validation, M.W., X.Z. and Q.G.; formal analysis, M.W.; investigation, M.W.; resources, M.W.; data curation, M.W.; writing—original draft preparation, M.W.; writing—review and editing, X.W., X.M. and J.W.; visualization, M.W.; supervision, X.Z. and Q.G.; project administration, X.Z. and Q.G.; funding acquisition, X.W., X.Z. and Q.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by National Social Science Fund (No.23BTJ065 and No.21BTJ030).

Data Availability Statement: The datasets used and analyzed of this study are available from the corresponding author(s) on reasonable request.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A

Suppose $\boldsymbol{x} = (x_1, ..., x_p)^T$, where $E(x_j) = 0,\ j = 1, 2, ..., p$, and $\boldsymbol{\Sigma} = E(\boldsymbol{x}\boldsymbol{x}^T)$ is positive definite. For a real number $r \geq 0$, define the set $\boldsymbol{\Theta}(r) = \{\boldsymbol{u} \in \mathbb{R}^p : \|\boldsymbol{u}\|_{\boldsymbol{\Sigma}} \leq r\}$, $partial\boldsymbol{\Theta}(r) = \{\boldsymbol{u} \in \mathbb{R}^p : \|\boldsymbol{u}\|_{\boldsymbol{\Sigma}} = r\}$.

In order to prove Theorem 1, two lemmas are given first.

Lemma A1. *Assuming conditions* ***A1, A2*** *hold,* $m_3 = \sup_{\boldsymbol{u}\in\mathbb{S}^{p-1}} \mathbb{E}|\langle \boldsymbol{u}, \boldsymbol{\omega}\ range\rangle|^3 < \infty$, *the window width satisfies* $0 < h < \frac{1}{L\{\kappa_1 + (m_3\kappa_2)^{1/2}\}} \underline{f}$. *Secondly,* $\boldsymbol{\beta}_h^*$ *is the unique solution to* $Q_h^*(\boldsymbol{\beta})$ *and satisfies*

$$\theta_h := \|\boldsymbol{\beta}_h^* - \boldsymbol{\beta}^*\|_{\boldsymbol{\Sigma}} < \frac{L\kappa_2 h^2}{\underline{f} - L\kappa_1 h}. \tag{A1}$$

Moreover, assuming that $f_{\varepsilon|\boldsymbol{x}}(\cdot)$ *is continuously derivable and is satisfied for a constant* > 0, $|f'_{\varepsilon|\boldsymbol{x}}(u) - f'_{\varepsilon|\boldsymbol{x}}(0)| \leq l|u|$ *holds almost everywhere under* $\boldsymbol{x}$. *Then, we have*

$$\|\boldsymbol{\Sigma}^{-1/2}\boldsymbol{J}(\boldsymbol{\beta}_h^* - \boldsymbol{\beta}^*) + \frac{1}{2}\kappa_2 h^2 \cdot \boldsymbol{\Sigma}^{-1/2}\mathbb{E}\{f'_{\varepsilon|\boldsymbol{x}}(0)\boldsymbol{x}\}\|_2 \tag{A2}$$

$$\leq (\frac{1}{6}L\kappa_3 h^3 + \frac{1}{2}Lm_3\theta_h^2 + L\kappa_1 h\theta_h)(1 + O(n^{-1/2})), \tag{A3}$$

where $\boldsymbol{J} = \mathbb{E}\{\frac{\delta}{1-G(z)} f_{\varepsilon|\boldsymbol{x}}(0)\boldsymbol{x}\boldsymbol{x}^T\}$.

Proof of Lemma A1. To prove (A1), we define $\boldsymbol{\theta}_h = \boldsymbol{\beta}_h^* - \boldsymbol{\beta}^* \in \mathbb{R}^p$ and $\theta_h = \|\boldsymbol{\theta}_h\|_{\boldsymbol{\Sigma}}$. Because of the convexity of the loss function $Q_h(\boldsymbol{\beta})$ and the fact that $\boldsymbol{\beta}_h^*$ is an optimal solution of $Q_h^*(\boldsymbol{\beta})$—i.e., there exists a first-order optimality condition $\nabla Q_h^*(\boldsymbol{\beta}_h^*) = 0$—so we have

$$0 \leq \langle \nabla Q_h^*(\boldsymbol{\beta}_h^*) - \nabla Q_h^*(\boldsymbol{\beta}^*), \boldsymbol{\beta}_h^* - \boldsymbol{\beta}^* \rangle \leq \|\boldsymbol{\Sigma}^{-1/2}\nabla Q_h^*(\boldsymbol{\beta}^*)\|_2 \|\boldsymbol{\theta}_h\|_{\boldsymbol{\Sigma}}. \tag{A4}$$

The last step of (A4) is given by the *hölder* inequality.

By using Taylor's formula, we obtain

$$\frac{1 - G(z_i, \boldsymbol{\theta}_n)}{1 - G(z_i, \hat{\boldsymbol{\theta}}_n)} = \frac{1 - G(z_i, \boldsymbol{\theta}_n)}{1 - G(z_i, \boldsymbol{\theta}_n) - \frac{\partial G(z_i, \boldsymbol{\theta}_n^*)}{\partial \boldsymbol{\theta}}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_n)},$$

where $\|\boldsymbol{\theta}_n^* - \boldsymbol{\theta}_n\|_2 \leq \|\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_n\|_2$. Combined with ***A3*** and the assumption that $\hat{\boldsymbol{\theta}}_n$ is the maximum likelihood estimation of $\boldsymbol{\theta}_n$, we have $\|\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_n\|_2 = O(\frac{1}{\sqrt{n}})$. When the number of parameters is limited, the estimated distribution of censored variable equals to its true

distribution and converges to probability one. For ease of illustration, the equals sign is used in the following explanations.

Notice that $\nabla Q_h^*(\boldsymbol{\beta}^*) = \mathbb{E}\{\frac{\delta[\mathcal{K}(-\varepsilon/h)-\tau]}{1-\hat{G}(z)}\boldsymbol{x}\} = \mathbb{E}\{\frac{\delta[\mathcal{K}(-\varepsilon/h)-\tau)}{1-\hat{G}(y)}\boldsymbol{x}\}$. Through expanding using Taylor's formula yields, we have

$$\begin{aligned}\mathbb{E}\{\frac{\delta}{1-\hat{G}(z)}[\mathcal{K}(-\varepsilon/h)-\tau]|\boldsymbol{x}\} &= \int_{-\infty}^{+\infty}\frac{\int_{t+\boldsymbol{\beta}^T\boldsymbol{x}}^{+\infty}\mathrm{d}G(s)}{1-\hat{G}(t+\boldsymbol{\beta}^T\boldsymbol{x})}\{\mathcal{K}(-t/h)-\tau\}\mathrm{d}F_{\varepsilon|\boldsymbol{x}}(t)\\ &= \int_{-\infty}^{+\infty}K(u)\int_0^{hu}\{f_{\varepsilon|\boldsymbol{x}}(t)-f_{\varepsilon|\boldsymbol{x}}(0)\}\mathrm{d}t\mathrm{d}u(1+O_p(n^{-1/2})).\end{aligned}$$

Combined with the above equation, we can obtain

$$\|\boldsymbol{\Sigma}^{-1/2}\nabla Q_h^*(\boldsymbol{\beta}^*)\|_2 = \sup_{\boldsymbol{u}\in\mathbb{S}^{p-1}}\mathbb{E}\{\frac{\delta[\mathcal{K}(-\varepsilon/h)-\tau]}{1-\hat{G}(z)}\}\langle\boldsymbol{u},\boldsymbol{\Sigma}^{-1/2}\boldsymbol{x}\rangle \le \frac{1}{2}L\kappa_2h^2(1+O(n^{-1/2})). \tag{A5}$$

For the left-hand side of Equation (A4), the mean value theorem for vector functions is given by

$$\nabla Q_h^*(\boldsymbol{\beta}_h^*) - \nabla Q_h^*(\boldsymbol{\beta}^*) = \int_0^1\nabla^2Q_h^*(\boldsymbol{\beta}^*+t\boldsymbol{\theta}_h)\mathrm{d}t\boldsymbol{\theta}_h\,, \tag{A6}$$

where $\nabla^2Q_h^*(\boldsymbol{\beta}) = \mathbb{E}\{\frac{\delta}{1-\hat{G}(z)}K_h(z-\boldsymbol{\beta}^T\boldsymbol{x})\boldsymbol{x}\boldsymbol{x}^T\}$. By $\boldsymbol{\theta} = \boldsymbol{\beta}-\boldsymbol{\beta}^*$ and Lipschitz condition, we obtain

$$\mathbb{E}\{\frac{\delta K_h(z-\boldsymbol{\beta}^T\boldsymbol{x})}{1-\hat{G}(z)}|\boldsymbol{x}\} = \int K(v)f_{\varepsilon|\boldsymbol{x}}(\boldsymbol{\theta}^T\boldsymbol{x}+hv)\mathrm{d}v = (f_{\varepsilon|\boldsymbol{x}}(0)+R_h^*(\boldsymbol{\theta}))(1+O(n^{-1/2})), \tag{A7}$$

and $R_h^*(\boldsymbol{\theta})$ satisfies $|R_h^*(\boldsymbol{\theta})| \le L(|\boldsymbol{\theta}^T\boldsymbol{x}|+\kappa_1h)$. Combining (A6) and (A7) and the assumption that $f_{\varepsilon|\boldsymbol{x}}(0) \ge \underline{f} > 0$,

$$\langle\nabla Q_h^*(\boldsymbol{\beta}_h^*)-\nabla Q_h^*(\boldsymbol{\beta}^*),\boldsymbol{\beta}_h^*-\boldsymbol{\beta}^*\rangle \ge (\underline{f}\theta_h^2-\frac{1}{2}Lm_3\cdot\theta_h^3-L\kappa_1h\cdot\theta_h^2)(1+O(n^{-1/2})). \tag{A8}$$

Combining (A4), (A5), and (A8), it can be found that $\theta_h \ge 0$ satisfies $0.5Lm_3\cdot\theta_h^2-(\underline{f}-L\kappa_1h)\theta_h+0.5L\kappa_2h^2 \ge 0$. Under the condition that $L\{\kappa_1+(m_3\kappa_2)^{1/2}\}h < \underline{f}$, the solution to the inequality yields

$$\theta_h \le \frac{L\kappa_2h^2}{\underline{f}-L\kappa_1h+\Delta_h^{1/2}}\,, \tag{A9}$$

$$\theta_h \ge \frac{\underline{f}-L\kappa_1h+\Delta_h^{1/2}}{Lm_3}, \tag{A10}$$

where $\Delta_h := (\underline{f}-L\kappa_1h)^2-L^2m_3\kappa_2h^2$, but it is necessary to exclude the (A10) equation.

Suppose θ_h satisfies (A10), $\theta_h > L(m_3\kappa_2)^{1/2}h/(Lm_3) = (\kappa_2/m_3)^{1/2}h =: r_0$. There must exist $\beta\in(0,1)$, such that $\widetilde{\boldsymbol{\beta}} := (1-\eta)\boldsymbol{\beta}^*+\beta\boldsymbol{\beta}_h^*$ satisfies $\|\widetilde{\boldsymbol{\beta}}-\boldsymbol{\beta}^*\|_\Sigma = \eta\theta_h = r_0$. By the convexity of $Q_h(\boldsymbol{\beta})$, it can be proved that

$$\begin{aligned}\langle\nabla Q_h^*(\widetilde{\boldsymbol{\beta}})-\nabla Q_h^*(\boldsymbol{\beta}^*),\widetilde{\boldsymbol{\beta}}-\boldsymbol{\beta}^*\rangle &\le \eta\cdot\langle\nabla Q_h^*(\boldsymbol{\beta}_h^*)-\nabla Q_h^*(\boldsymbol{\beta}^*),\boldsymbol{\beta}_h^*-\boldsymbol{\beta}^*\rangle\\ &= \langle-\nabla Q_h^*(\boldsymbol{\beta}^*),\widetilde{\boldsymbol{\beta}}-\boldsymbol{\beta}^*\rangle.\end{aligned}$$

Repeating the analysis of (A5) and (A8), the right of above the inequality is found, $\leq 0.5L\kappa_2 h^2 \cdot r_0$, as is the left, $\geq \underline{f} \cdot r_0^2 - 0.5Lm_3 \cdot r_0^3 - L\kappa_1 h \cdot r_0^2 = \{\underline{f} - L\kappa_1 h - 0.5L(m_3\kappa_2)^{1/2}h\}r_0^2$. Eliminating the common element r_0 on both sides, there exists

$$r_0 \leq \frac{0.5L\kappa_2 h^2}{\underline{f} - L\kappa_1 h - 0.5L(m_3\kappa_2)^{1/2}h} < \frac{0.5L\kappa_2 h^2}{0.5L(m_3\kappa_2)^{1/2}h} = (\kappa_2/m_3)^{1/2}h = r_0.$$

This leads to a contradiction. Therefore, h must be satisfied to fulfill (A9), the first inequality constraint.

In order to demonstrate (A3), the bias of the estimator is examined below. We define the variables

$$\boldsymbol{\Delta} = \boldsymbol{\Sigma}^{-1/2}\{\nabla Q_h^*(\boldsymbol{\beta}_h^*) - \nabla Q_h^*(\boldsymbol{\beta}^*) - \boldsymbol{J}(\boldsymbol{\beta}_h^* - \boldsymbol{\beta}^*)\},$$

and the matrix $\boldsymbol{H} = \boldsymbol{\Sigma}^{-1/2}\boldsymbol{J}\boldsymbol{\Sigma}^{-1/2} = \mathbb{E}\{\frac{\delta}{1-\hat{G}(z)} f_{\varepsilon|\boldsymbol{x}}(0)\boldsymbol{\omega}\boldsymbol{\omega}^T\}$, where $\boldsymbol{\omega} = \boldsymbol{\Sigma}^{-1/2}\boldsymbol{x}$. Similarly, according to the mean value theorem for vector functions it follows that

$$\boldsymbol{\Delta} = \left\{\int_0^1 \boldsymbol{\Sigma}^{-1/2}\nabla^2 Q_h^*(\boldsymbol{\beta}^* + t\boldsymbol{\theta}_h)\boldsymbol{\Sigma}^{-1/2}\mathrm{d}t - \boldsymbol{H}\right\}\boldsymbol{\Sigma}^{1/2}\boldsymbol{\theta}_h, \tag{A11}$$

where $\boldsymbol{\theta}_h = \boldsymbol{\beta}_h^* - \boldsymbol{\beta}^*$. Combining the Lipschitz continuity of $f_{\varepsilon|\boldsymbol{x}}(\cdot)$ and (A11),

$$\|\boldsymbol{\Delta}\|_2 \leq L(0.5m_3\|\boldsymbol{\theta}_h\|_{\boldsymbol{\Sigma}} + \kappa_1 h)\|\boldsymbol{\theta}_h\|_{\boldsymbol{\Sigma}}(1 + O(n^{-1/2})). \tag{A12}$$

Furthermore, expanding $f_{\varepsilon|\boldsymbol{x}}(\cdot)$ by a second-order Taylor series, we obtain

$$\|\boldsymbol{\Sigma}^{-1/2}\nabla Q_h^*(\boldsymbol{\beta}^*) - \frac{1}{2}\kappa_2 h^2 \cdot \boldsymbol{\Sigma}^{-1/2}\mathbb{E}\{f'_{\varepsilon|\boldsymbol{x}}(0)\boldsymbol{x}\}\|_2 \leq \frac{1}{6}l\kappa_3 h^3(1 + O(n^{-1/2})). \tag{A13}$$

Combining (A12) and (A13) proves that Lemma A1 holds. □

Lemma A1 discusses the connection between $\boldsymbol{\beta}_h^*$ and $\boldsymbol{\beta}^*$. Lemma A2 must be proved before discussing the relationship between $\hat{\boldsymbol{\beta}}_h$ and $\boldsymbol{\beta}^*$.

$$D_h(\boldsymbol{\theta}) = Q_h(\boldsymbol{\beta}^* + \boldsymbol{\theta}) - Q_h(\boldsymbol{\beta}^*), R_h(\boldsymbol{\theta}) = D_h(\boldsymbol{\theta}) - \langle \nabla Q_h(\boldsymbol{\beta}^*), \boldsymbol{\theta}\rangle,$$

$$D_h^*(\boldsymbol{\theta}) = Q_h^*(\boldsymbol{\beta}^* + \boldsymbol{\theta}) - Q_h^*(\boldsymbol{\beta}^*), R_h^*(\boldsymbol{\theta}) = D_h^*(\boldsymbol{\theta}) - \langle \nabla Q_h^*(\boldsymbol{\beta}^*), \boldsymbol{\theta}\rangle.$$

Lemma A2. *For any $u \geq 0$, given $r \geq 0$, the following equation holds:*

$$\mathbb{P}\left\{\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}(r)}\{D_h^*(\boldsymbol{\theta}) - D_h(\boldsymbol{\theta})\} \leq 3\bar{\tau}v_0^* r \cdot \left(\sqrt{\frac{P_d \cdot u}{n}} + \frac{u}{n}\right)\right\} \geq 1 - e^{4p-u}, \tag{A14}$$

where $P_d = \frac{\sum_{i=1}^n \delta_i}{n}$ is the uncensored proportion, $\bar{\tau} = max\{\tau, 1-\tau\}$, and $v_0^* = C_0 v_0$. Furthermore, given $r_u > r_l > 0$, for any $u \geq 0$, the probability inequality

$$\mathbb{P}\left\{D_h^*(\boldsymbol{\theta}) - D_h(\boldsymbol{\theta}) \leq 4.25\bar{\tau}v_0^*\|\boldsymbol{\theta}\|_{\boldsymbol{\Sigma}}\left(\sqrt{\frac{P_d \cdot u}{n}} + \frac{u}{n}\right)\right\} \geq 1 - \left\lceil e\, log(\frac{r_u}{r_l})\right\rceil e^{4p-u} \tag{A15}$$

holds, where $\boldsymbol{\theta}$ satisfies $r_l \leq \|\boldsymbol{\theta}\|_{\boldsymbol{\Sigma}} \leq r_u$.

Proof of Lemma A2. For each sample $\boldsymbol{s_i} = (\boldsymbol{x_i}, \varepsilon_i)$, define the loss function Difference $d_h(\boldsymbol{\theta}; \boldsymbol{s_i}) = \ell_h(\varepsilon_i - \langle \boldsymbol{x_i}, \boldsymbol{\theta} \rangle) - \ell_h(\varepsilon_i)$, so $D_h(\boldsymbol{\theta}) = (1/n) \sum_{i=1}^{n} d_h(\boldsymbol{\theta}; \boldsymbol{s_i})$. By the loss function $\ell_h(u)$ being Lipschitz continuous, $d_h(\boldsymbol{\theta}; \boldsymbol{s_i})$ is also $\bar{\tau} - Lipschitz$ continuous in $\langle \boldsymbol{x_i}, \boldsymbol{\theta} \rangle$, i.e., for any $\boldsymbol{s_i}$ and $\boldsymbol{\theta}, \boldsymbol{\theta}' \in \mathbb{R}^p$, $|d_h(\boldsymbol{\theta}; \boldsymbol{s_i}) - d_h(\boldsymbol{\theta'}; \boldsymbol{s_i})| \leq \frac{\delta_i}{1-\hat{G}(z_i)} \bar{\tau} |\langle \boldsymbol{x_i}, \boldsymbol{\theta} \rangle - \langle \boldsymbol{x_i}, \boldsymbol{\theta'} \rangle|$.

For any given $r > 0$ and some $\epsilon \in (0, 1)$, define $\triangle_\epsilon(r) = n(1-\epsilon) \sup_{e \in \boldsymbol{\Theta}(r)} \{D_h^*(\boldsymbol{\theta}) - D_h(\boldsymbol{\theta})\} / (2\bar{\tau} r)$, where $D_h^*(\boldsymbol{\theta}) = \mathbb{E} D_h(\boldsymbol{\theta})$. Utilizing Chebyshev's inequality, we obtain that

$$\mathbb{P}\{\triangle_\epsilon(r) \geq u\} \leq \exp\left[- \sup_{\lambda \geq 0} \{\lambda u - \log \mathbb{E} e^{\lambda \triangle_\epsilon(r)}\} \right]. \tag{A16}$$

Through Rademacher's symmetric control moment function $\mathbb{E} e^{\lambda \triangle_\epsilon(r)}$, there exists

$$\mathbb{E} e^{\lambda \triangle_\epsilon(r)} \leq \mathbb{E} \exp\left\{ 2\lambda(1-\epsilon) \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}(r)} \frac{1}{2\bar{\tau} r} \sum_{\delta_i = 1} \pi_i d_h(\boldsymbol{\theta}; \boldsymbol{s_i}) \right\},$$

where $\pi_1, ..., \pi_n$ are independent Rademacher random variables. If $\langle \boldsymbol{x_i}, \boldsymbol{\theta} \rangle = 0$, then $d_h(\boldsymbol{\theta}; \boldsymbol{s_i}) = 0$ considering that $d_h(\boldsymbol{\theta}; \boldsymbol{s_i})$ is also $\bar{\tau} - Lipschitz$ continuous in $\langle \boldsymbol{x_i}, \boldsymbol{\theta} \rangle$. By using the contraction inequality, we have

$$\begin{aligned} &\mathbb{E} \exp\left\{ 2\lambda(1-\epsilon) \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}(r)} \frac{1}{2\bar{\tau} r} \sum_{\delta_i = 1} \pi_i d_h(\boldsymbol{\theta}; \boldsymbol{s_i}) \right\} \\ \leq &\mathbb{E} \exp\left\{ \frac{\lambda}{r}(1-\epsilon) \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}(r)} \sum_{\delta_i = 1} \pi_i \frac{\delta_i \langle \boldsymbol{x_i}, \boldsymbol{\theta} \rangle}{1 - \hat{G}(y_i)} \right\}, \end{aligned}$$

where $\boldsymbol{\omega}_i = \boldsymbol{\Sigma}^{-1/2} \boldsymbol{x_i}$. For such a $\epsilon \in (0, 1)$, there exists an ϵ net $\{\boldsymbol{u}_1, ..., \boldsymbol{u}_{N_\epsilon}\}$ on $\mathbb{S}^{p-1}$, when $N_\epsilon \leq (1 + 2/\epsilon)^p$ such that $\| \sum_{\delta_i = 1} \frac{\pi_i \delta_i \boldsymbol{\omega}_i}{1 - \hat{G}(y_i)} \|_2 \leq (1-\epsilon)^{-1} \max_{1 \leq j \leq N_\epsilon} \sum_{\delta_i = 1} \frac{\delta_i \pi_i \boldsymbol{u}_j^T \boldsymbol{\omega}_i}{1 - \hat{G}(y_i)}$. This suggests that

$$\mathbb{E} \exp\left\{ \lambda(1-\epsilon) \left\| \sum_{\delta_i = 1} \frac{\pi_i \delta_i \boldsymbol{\omega}_i}{1 - \hat{G}(y_i)} \right\|_2 \right\} \leq \sum_{j=1}^{N_\epsilon} \mathbb{E} \exp\left\{ \lambda \sum_{\delta_i = 1} \frac{\pi_i \boldsymbol{u}_j^T \delta_i \boldsymbol{\omega}_i}{1 - \hat{G}(y_i)} \right\}.$$

Define $S_j = \sum_{\delta_i = 1} \frac{\pi_i \boldsymbol{u}_j^T \delta_i \boldsymbol{\omega}_i}{1 - \hat{G}(y_i)}$, and we notice that $\pi_i \in \{-1, 1\}$ is symmetric. From condition ***A3***, there exists a constant C_0 such that $\frac{1}{1-G(z)} \leq C_0$, then we have

$$\mathbb{P}(|\langle \boldsymbol{u}, \frac{\delta}{1 - G(z)} \boldsymbol{\omega} \rangle| \geq v_0^* t) \leq \mathbb{P}(C_0 |\boldsymbol{u}^T \boldsymbol{\omega}| \geq v_0^* t) = \mathbb{P}(|\boldsymbol{u}^T \boldsymbol{\omega}| \geq v_0 t) \leq e^{-t}.$$

Taking $v_0^* = C_0 v_0$, for any $k \geq 3$, $\mathbb{E} |\frac{\boldsymbol{u}_j^T \delta_i \boldsymbol{\omega}_i}{1 - \hat{G}(y_i)}|^k \leq (v_0^*)^k k \int_0^{+\infty} t^{k-1} e^{-t} \mathrm{d}t = (v_0^*)^k k! (1 + O(n^{-1/2}))$. Thus, for all $0 \leq c < 1/v_0$,

$$\begin{aligned} &\mathbb{E} e^{c \frac{\pi_i \boldsymbol{u}_j^T \delta_i \boldsymbol{\omega}_i}{1 - \hat{G}(y_i)}} = 1 + \frac{c^2}{2} \mathbb{E} (\frac{\pi_i \boldsymbol{u}_j^T \delta_i \boldsymbol{\omega}_i}{1 - \hat{G}(y_i)})^2 + \sum_{\ell=3}^{\infty} \frac{c^\ell}{\ell!} (\frac{\pi_i \boldsymbol{u}_j^T \delta_i \boldsymbol{\omega}_i}{1 - \hat{G}(y_i)})^\ell \\ \leq &\left(1 + \frac{c^2 v_0^{*2}}{2} + \sum_{\ell=2}^{\infty} \frac{c^{2\ell}}{(2\ell)!} v_0^{*2\ell} (2\ell)! \right) \leq \left(1 + \frac{c^2 v_0^{*2}}{2} + \sum_{\ell=2}^{\infty} (c v_0^*)^{2\ell} \right) (1 + O(n^{-1/2})). \end{aligned}$$

It follows that for every $0 < \lambda < 1/(\sqrt{2}v_0^*)$ and $j = 1, ..., N_\epsilon$,

$$\log \mathbb{E}e^{\lambda S_j} \leq \frac{(n_0 + O(\sqrt{n_0}))v_0^{*2}\lambda}{2(1-\sqrt{2}v_0^*\lambda)}, s.t. \log \mathbb{E}e^{\lambda S_j} \leq \frac{n_0 v_0^{*2}\lambda}{2(1-\sqrt{2}v_0^*\lambda)}$$

where n_0 is the n sample size that satisfies the condition $\delta_i = 1$. For any $u \geq 0$, notice that

$$\sup_{\lambda \geq 0}\{\lambda u - \log \mathbb{E}e^{\lambda \triangle_\epsilon(r)}\} \geq -\log N_\epsilon + \sup_{\lambda \in (0,(\sqrt{2}v_0^*)^{-1})}\left\{\lambda u - \frac{n_0 v_0^{*2}\lambda}{2(1-\sqrt{2}v_0^*\lambda)}\right\}.$$

Substituting the above inequality into (A16), it can be shown that when $\epsilon = 2/(e^4 - 1)$,

$$\mathbb{P}\left\{\sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}(r_0)}\{D_h^*(\boldsymbol{\theta}) - D_h(\boldsymbol{\theta})\} \leq \frac{2\sqrt{2}}{1-\epsilon}\bar{\tau}v_0^* r \cdot \left(\sqrt{\frac{P_d \cdot u}{n}} + \frac{u}{n}\right)\right\} \geq 1 - e^{4p-u}, \tag{A17}$$

where $P_d = n_0/n$. The following is a proof of Equation (A15), which by disproof yields $\boldsymbol{\theta} \in \boldsymbol{\Theta}(r_l, r_u) := \{\boldsymbol{v} \in \mathbb{R}^p . r_l \leq \|\boldsymbol{v}\|_{\boldsymbol{\Sigma}} \leq r_u\}$. For some $\gamma > 1$ and positive integers $k = 1, ..., N := \lceil \log(\frac{r_u}{r_l})/\log(\gamma) \rceil$, define the set $\boldsymbol{\Theta}_k = \{\boldsymbol{v} \in \mathbb{R}^p : \gamma^{k-1} r_l \leq \|\boldsymbol{v}\|_{\boldsymbol{\Sigma}} \leq \gamma^k r_l\},$. So $\boldsymbol{\Theta}(r_l, r_u) \subseteq \bigcup_{k=1}^N \boldsymbol{\Theta}_k$. Then, we repeat (A17) and let $r = \gamma^k r_l, k = 1, ..., N$,

$$\begin{aligned}
&\mathbb{P}\left\{\exists \boldsymbol{\theta} \in \boldsymbol{\Theta}(r_l, r_u) s.t. D_h^*(\boldsymbol{\theta}) - D_h(\boldsymbol{\theta}) > \frac{2\sqrt{2}\gamma}{1-\epsilon}\bar{\tau}v_0^*\|\boldsymbol{\theta}\|_{\boldsymbol{\Sigma}} \cdot \left(\sqrt{\frac{P_d \cdot u}{n}} + \frac{u}{n}\right)\right\} \\
&\leq \sum_{k=1}^N \mathbb{P}\left\{\sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}(\gamma^k r_l)} D_h^*(\boldsymbol{\theta}) - D_h(\boldsymbol{\theta}) > \frac{2\sqrt{2}}{1-\epsilon}\bar{\tau}v_0^*\gamma^k r_l \cdot \left(\sqrt{\frac{P_d \cdot u}{n}} + \frac{u}{n}\right)\right\} \\
&\leq \sum_{k=1}^N \exp\{p\log(1+2/\epsilon) - u\} \leq \lceil \log(\frac{r_u}{r_l})/\log(\gamma) \rceil \exp\{p\log(1+2/\epsilon) - u\}.
\end{aligned}$$

The required constraint can be obtained by taking $\epsilon = 2/(e^4 - 1)$ and $\gamma = e^{1/e}$. □

Proof of Theorem 1. According to the definition of $D_h(\boldsymbol{\theta})$ and (A5), it can be found that

$$\begin{aligned}
D_h(\boldsymbol{\theta}) &= \langle \nabla Q_h^*(\boldsymbol{\beta}^*), \boldsymbol{\theta} \rangle + R_h^*(\boldsymbol{\theta}) + \{D_h(\boldsymbol{\theta}) - D_h^*(\boldsymbol{\theta})\} \\
&\geq R_h^*(\boldsymbol{\theta}) - \|\boldsymbol{\Sigma}^{-1/2}\nabla Q_h^*(\boldsymbol{\beta}^*)\|_2 \|\boldsymbol{\theta}\|_{\boldsymbol{\Sigma}} - \{D_h^*(\boldsymbol{\theta}) - D_h(\boldsymbol{\theta})\}
\end{aligned} \tag{A18}$$

where $\{D_h^*(\boldsymbol{\theta}) - D_h(\boldsymbol{\theta})\}$ is the sampling error. From (A8), it follows again that for each $\boldsymbol{\theta} \in \mathbb{R}^p$, there are

$$R_h^*(\boldsymbol{\theta}) \geq \frac{1}{2}(\underline{f} - L\kappa_1 h - 0.5Lm_3 \cdot \|\boldsymbol{\theta}\|_{\boldsymbol{\Sigma}}) \cdot \|\boldsymbol{\theta}\|_{\boldsymbol{\Sigma}}^2 (1 + O(n^{-1/2})). \tag{A19}$$

Take $r_0 = (2\kappa_2/m_3)^{1/2}h$ as the radius of convergence. For any $\boldsymbol{\theta} \in \partial\boldsymbol{\Theta}(r_0)$, combining (A18) and (A19), there exists

$$R_h^*(\boldsymbol{\theta}) - 0.5L\kappa_2 h^2 \cdot \|\boldsymbol{\theta}\|_{\boldsymbol{\Sigma}}(1 + O(n^{-1/2})) \geq \frac{1}{2}(\underline{f} - L\kappa_1 h - L(2\kappa_2 m_3)^{1/2}h)r_0^2(1 + O(n^{-1/2})). \tag{A20}$$

To control the last term on the right-hand side of (A18), Lemma A2 gives the law of large numbers for $\{D_h(\boldsymbol{\theta}) - D_h^*(\boldsymbol{\theta}), \boldsymbol{\theta} \in \boldsymbol{\Theta}(r)\}$. First, applying (A14) and letting $r = r_0, u = 4p + t$, we obtain the following variation.

$$\mathbb{P}\left\{\sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}(r_0)}\{D_h^*(\boldsymbol{\theta}) - D_h(\boldsymbol{\theta})\} \leq 3\bar{\tau}v_0^* r_0 \cdot \left(\sqrt{\frac{P_d(4p+t)}{n}} + \frac{4p+t}{n}\right)\right\} \geq 1 - e^{-t}. \tag{A21}$$

Combining the inequalities (A18) and (A20), and the upper bound of (A21), it is sufficient that the window width h satisfies $\underline{f}^{-1}m_3^{1/2}v_0^*\sqrt{P_d(p+t)/n} \lesssim h \lesssim \underline{f}m_3^{-1/2}$. Then, for all $\boldsymbol{\theta} \in \partial\boldsymbol{\Theta}(r_0)$, the

$$\mathbb{P}\{D_h(\boldsymbol{\theta}) > 0\} \geq 1 - e^{-t}. \tag{A22}$$

Using the condition (A22), $D_h(0) = 0$ and $\widehat{\boldsymbol{\theta}} = argmin D_h(\boldsymbol{\theta})$, and the convexity of $D_h(\cdot)$ ensures that the $\|\widehat{\boldsymbol{\theta}}\|_{\boldsymbol{\Sigma}} \leq r_0$ holds with high probability. On the other hand, according to the optimality of $\widehat{\boldsymbol{\beta}}_h$, $\widehat{\boldsymbol{\theta}} := \widehat{\boldsymbol{\beta}}_h - \boldsymbol{\beta}^*$ satisfies $D_h(\widehat{\boldsymbol{\theta}}) \leq 0$.

Secondly, when $\|\widehat{\boldsymbol{\theta}}\|_{\boldsymbol{\Sigma}}$ has a lower bound, i.e., $\|\widehat{\boldsymbol{\theta}}\|_{\boldsymbol{\Sigma}}$ falls into an interval, the rate of convergence of $\widehat{\boldsymbol{\beta}}_h$ needs to be redetermined. Consider the cyclic set $\boldsymbol{\Theta}(r_l, r_0) = \{\boldsymbol{\theta} \in \mathbb{R}^p : r_l \leq \|\boldsymbol{\theta}\|_{\boldsymbol{\Sigma}} \leq r_0\}$ and $r_l = r_0 h$. If $\widehat{\boldsymbol{\theta}} \notin \boldsymbol{\Theta}(r_l, r_0)$, then there must exist $\widehat{\boldsymbol{\theta}} \in \boldsymbol{\Theta}(r_l)$, in which case the constraints can be reduced to the first case. Assume that $\widehat{\boldsymbol{\theta}} \in \boldsymbol{\Theta}(r_l, r_0)$. Using Equation (A15), let $(r_l, r_u) = (r_0 h, r_0)$, $u = P_d(log(elogh^{-1}) + 4p + t)$, and $r_1 = 4.25\bar{\tau}v_0^*\left(\sqrt{\frac{P_d(log(elogh^{-1})+4p+t)}{n}} + \frac{log(elogh^{-1})+4p+t}{n}\right)$, we get

$$\mathbb{P}\{D_h^*(\boldsymbol{\theta}) - D_h(\boldsymbol{\theta}) \leq \|\boldsymbol{\theta}\|_{\boldsymbol{\Sigma}} \cdot r_1\} \geq 1 - e^{-t}. \tag{A23}$$

The above inequality in $\boldsymbol{\theta}$ satisfies $\boldsymbol{\theta} \in \boldsymbol{\Theta}(r_l, r_0)$; thus, in $\boldsymbol{\theta}(r_l, r_0)$, it must contain $\widehat{\boldsymbol{\theta}}$. Applying this formula and combining it with (A18) and (A19) and the fact that $D_h(\widehat{\boldsymbol{\theta}}) \leq 0$, we obtain

$$\begin{aligned}(\underline{f} - L\kappa_1 h)\|\widehat{\boldsymbol{\theta}}\|_{\boldsymbol{\Sigma}}^2(1 + O(n^{-1/2})) &\leq (2r_1 + L\kappa_2 h)\|\widehat{\boldsymbol{\theta}}\|_{\boldsymbol{\Sigma}} + 0.5Lm_3\|\widehat{\boldsymbol{\theta}}\|_{\boldsymbol{\Sigma}}^3 \\ &\leq 2r_1\|\widehat{\boldsymbol{\theta}}\|_{\boldsymbol{\Sigma}} + 2L\kappa_2 h\|\widehat{\boldsymbol{\theta}}\|_{\boldsymbol{\Sigma}}(1 + O(n^{-1/2})).\end{aligned}$$

The conclusion of Theorem 1 can be proved by eliminating $\|\widehat{\boldsymbol{\theta}}\|_{\boldsymbol{\Sigma}}$ from both the leftmost and rightmost sides of the above equation. The proof of Theorem 1 differs from reference [10] in that $\|\widehat{\boldsymbol{\theta}}\|_{\boldsymbol{\Sigma}}$ will have $O(n^{-1/2})$ more convergence order, but since h is related to the sample size, it will only retain a higher convergence speed, so only the slower converging orders are retained. Thus, it looks like the result is the same as in the non-censoring case. □

Proof of Theorem 2. The procedure for the proof of Theorem 2 will retain the notation used in the proof of Theorem 1, for any $t \geq 0$ such that $r = r(n, p, t) \asymp \sqrt{(p+t)/n} + h^2 > 0$ as long as $\sqrt{(p+t)/n} \lesssim h \lesssim 1$ there exists $\mathbb{P}\{\widehat{\boldsymbol{\beta}}_h \in \boldsymbol{\beta}^* + \boldsymbol{\Theta}(r)\} \geq 1 - 2e^{-t}$. Define a vector-valued stochastic process,

$$\varphi(\boldsymbol{\theta}) = \boldsymbol{\Sigma}^{-1/2}\{\nabla Q_h^*(\boldsymbol{\beta}^* + \boldsymbol{\theta}) - \nabla Q_h^*(\boldsymbol{\beta}^*) - \boldsymbol{J}_h\boldsymbol{\theta}\}, \tag{A24}$$

where $\boldsymbol{J}_h = \nabla^2 Q_h^*(\boldsymbol{\beta}^*)$ is denoted as the overall Hessian array of $\boldsymbol{\beta}^*$. Since $\widehat{\boldsymbol{\beta}}_h$ falls in the domain of $\boldsymbol{\beta}^*$ with a high probability, $\widehat{\boldsymbol{\beta}}_h$ is able to constrain well the $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}(r)} \|\varphi(\boldsymbol{\theta})\|_2$. By means of the trigonometric inequality,

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}(r)} \|\varphi(\boldsymbol{\theta})\|_2 \leq \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}(r)} \|\mathbb{E}\varphi(\boldsymbol{\theta})\|_2 + \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}(r)} \|\varphi(\boldsymbol{\theta}) - \mathbb{E}\varphi(\boldsymbol{\theta})\|_2 := M + N. \tag{A25}$$

It is now necessary to give upper bounds for M and N, respectively. Consider first the upper bound for M, using the median theorem,

$$\mathbb{E}\varphi(\boldsymbol{\theta}) = \left\langle \int_0^1 \boldsymbol{\Sigma}^{-1/2}\nabla^2 Q_h^*(\boldsymbol{\beta}^* + t\boldsymbol{\theta})\boldsymbol{\Sigma}^{-1/2}\mathrm{d}t - \boldsymbol{H}_h, \boldsymbol{\Sigma}^{1/2}\boldsymbol{\theta} \right\rangle,$$

where $\boldsymbol{H}_h := \boldsymbol{\Sigma}^{-1/2}\boldsymbol{J}_h\boldsymbol{\Sigma}^{-1/2} = \mathbb{E}\{\frac{\delta}{1-\hat{G}(z)}K_h(\varepsilon)\boldsymbol{\omega}\boldsymbol{\omega}^T\}$. Expanding the second-order derivative of the above expectation second-order derivative, we write $\boldsymbol{v} = \boldsymbol{\Sigma}^{1/2}\boldsymbol{\theta}$. For all $\boldsymbol{\theta} \in \boldsymbol{\Theta}(r)$, we have

$$\boldsymbol{\Sigma}^{-1/2}\nabla^2 Q_h^*(\boldsymbol{\beta}^* + t\boldsymbol{\theta})\boldsymbol{\Sigma}^{-1/2} = \mathbb{E}\left\{\int_{-\infty}^{+\infty} K(u)f_{\varepsilon|\boldsymbol{x}}(t\langle\boldsymbol{\omega},\boldsymbol{v}\rangle - hu)\mathrm{d}u \cdot \boldsymbol{\omega}\boldsymbol{\omega}^T\right\}.$$

Utilizing the Lipschitz continuity of $f_{\varepsilon|\boldsymbol{x}}(\cdot)$,

$$\begin{aligned}
&\|\boldsymbol{\Sigma}^{-1/2}\left[\int_0^1 \nabla^2 Q_h^*(\boldsymbol{\beta}^* + t\boldsymbol{\theta})\mathrm{d}t\right]\boldsymbol{\Sigma}^{-1/2} - \boldsymbol{H}_h\|_2 \\
=&\left\|\mathbb{E}\int K(u)\{f_{\varepsilon|\boldsymbol{x}}(t\langle\boldsymbol{\omega},\boldsymbol{v}\rangle - hu) - f_{\varepsilon|\boldsymbol{x}}(-hu)\}\mathrm{d}u \cdot \boldsymbol{\omega}\boldsymbol{\omega}^T\right\|_2 \\
\leq& Lm_3rt(1 + O(n^{-1/2})).
\end{aligned}$$

The last inequality can be obtained from the Cauchy–Schwarz inequality. Therefore,

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}(r)} \|\mathbb{E}\varphi(\boldsymbol{\theta})\|_2 \leq 0.5Lm_3r^2(1 + O(n^{-1/2})). \tag{A26}$$

The proof procedure for the upper bound of N can be found in [10,29], i.e., the upper bound of the zero-mean stochastic process $\varphi(\boldsymbol{\theta}) - \mathbb{E}\varphi(\boldsymbol{\theta})$ in the sense of l_2-paradigm numbers. □

Proof of Theorem 3. Let $\boldsymbol{a} \in \mathbb{R}^p$ be a p-dimensional real vector. Given $h = h_n > 0$, define $S_n = \frac{1}{\sqrt{n}}\sum_{i=1}^n \gamma_i\eta_i$ and the centered variable $S_n^0 = S_n - \mathbb{E}S_n$, where $\eta_i = \frac{\delta_i}{1-\hat{G}(z_i)}\{\tau - \mathcal{K}(-\varepsilon_i/h)\}, \gamma_i = \langle\boldsymbol{J}_h^{-1}\boldsymbol{a}, \boldsymbol{x_i}\rangle$. According to the Lipschitz continuity for $f_{\varepsilon_i|\boldsymbol{x_i}}(\cdot)$ and the Fundamental Theorem of Calculusm we obtain $|\mathbb{E}(\eta_i|\boldsymbol{x_i})| \leq 0.5L\kappa_2h^2(1 + O(n^{-1/2}))$, $\kappa_2(h^2(1 + O(n^{-1/2}))$, and then $\mathbb{E}(\gamma_i\eta_i) \leq 0.5L\kappa_2\|\boldsymbol{J}_h^{-1}\boldsymbol{a}\|_{\boldsymbol{\Sigma}} \cdot h^2$.

The proof procedure of Theorem 3 can be found in [10]. □

Appendix B

Estimation results for the high-dimensional model when covariate X is generated from Case 2 (Figure A1) and Case 3 (Figure A2). Simulation results under the high-dimensional model when the covariates are generated from Case 2 (Figure A3) and Case 3 (Figure A4).

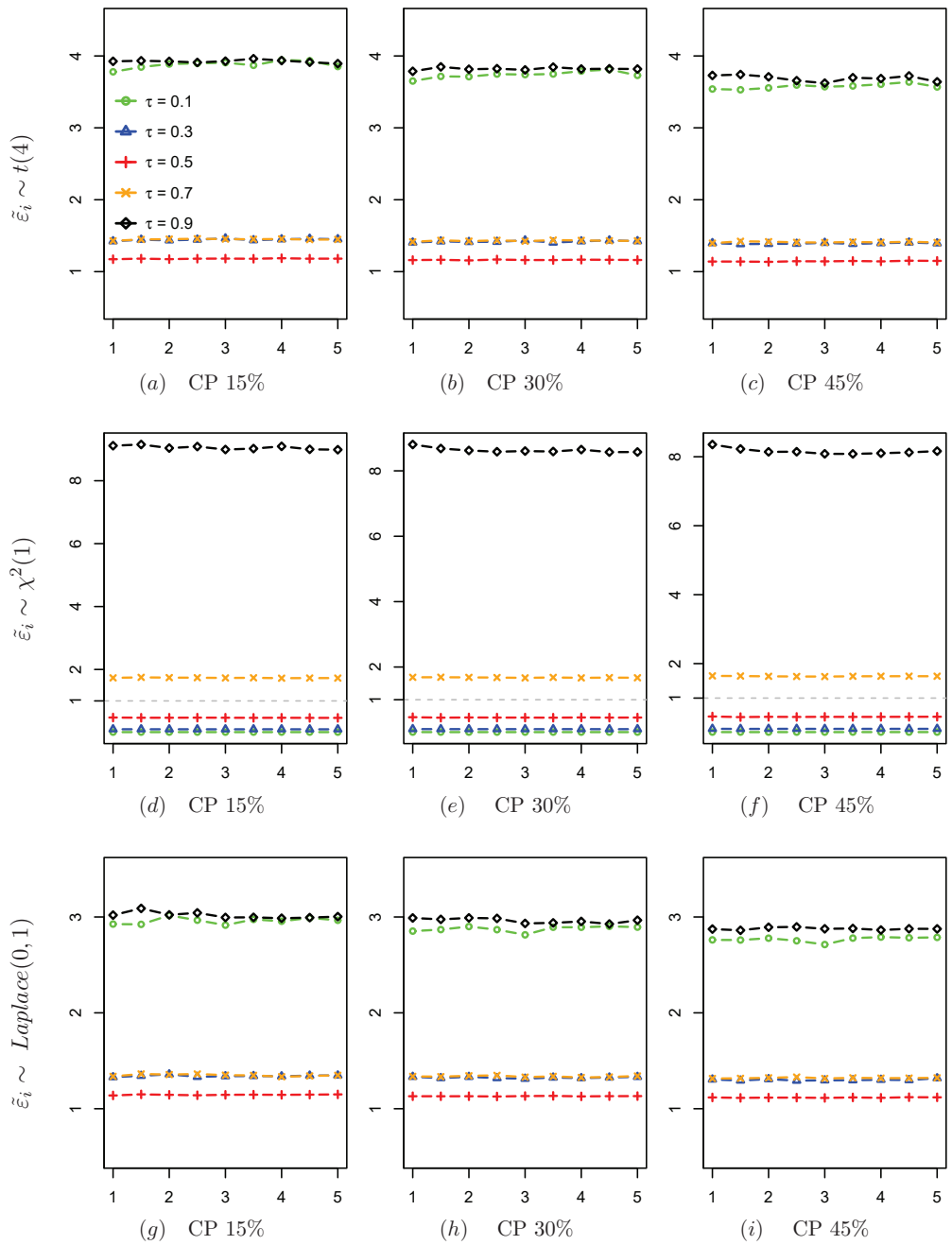


Figure A1. Estimation results for the high-dimensional model when covariate X is generated from Case 2. The horizontal coordinates of the plots indicate the sample size (in thousands), and the vertical coordinates indicate the ratio of DMSE for regression coefficients' estimators between CQ and SCQ.

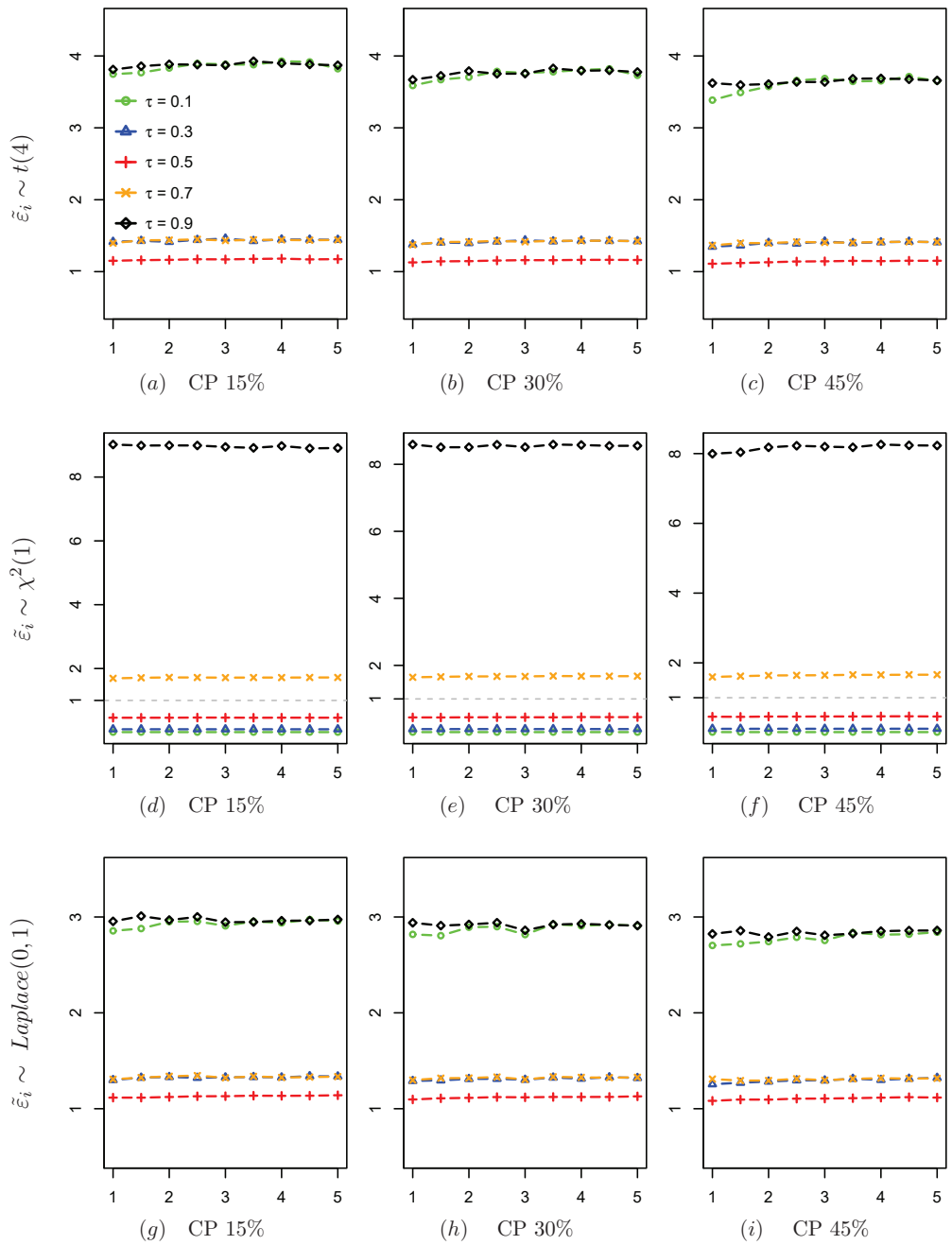


Figure A2. Estimation results for the high-dimensional model when covariate X is generated from Case 3. The horizontal coordinates of the plots indicate the sample size (in thousands), and the vertical coordinates indicate the ratio of DMSE for regression coefficients' estimators between CQ and SCQ.

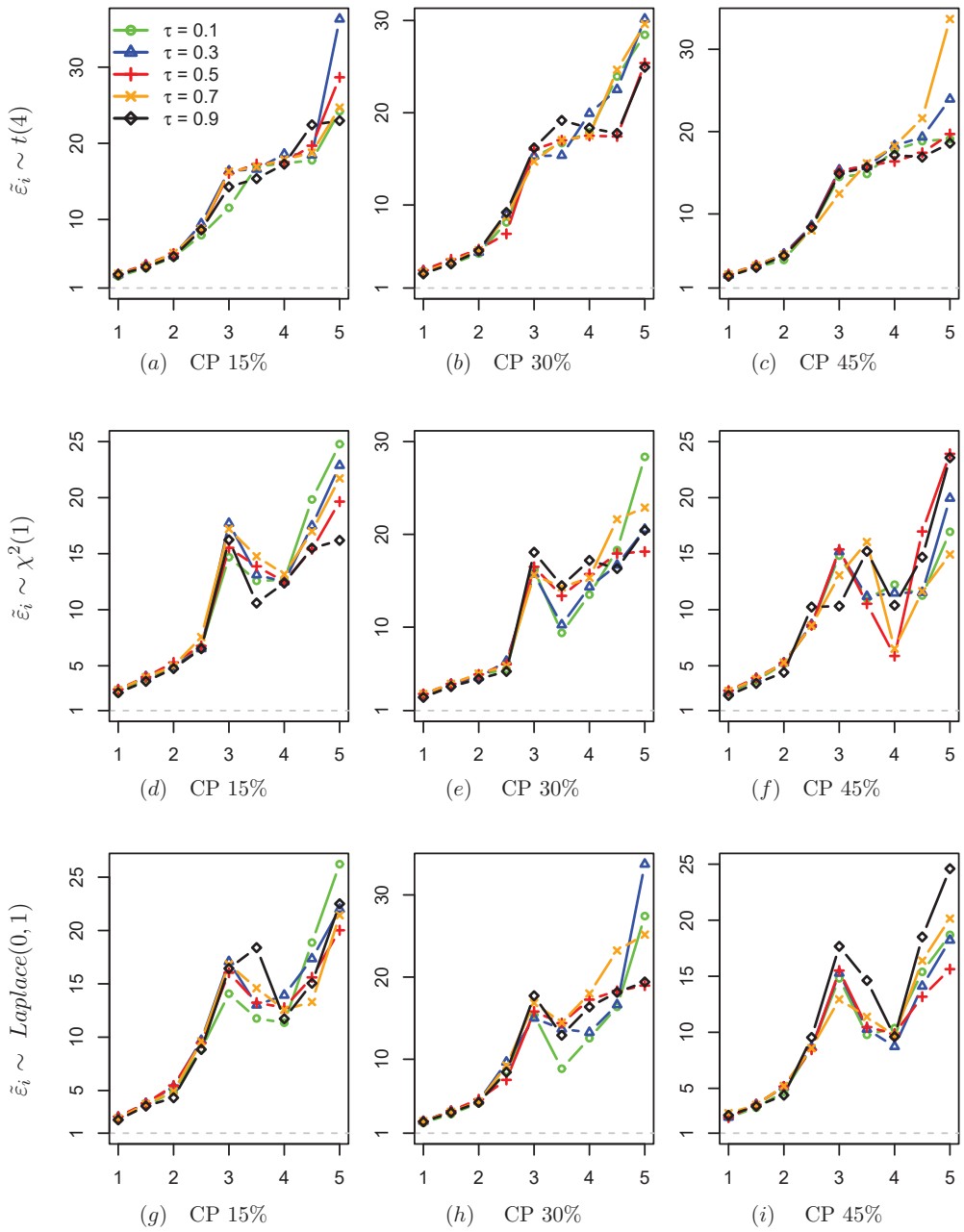


Figure A3. Simulation results under the high-dimensional model when the covariates are generated from Case 2, where the horizontal coordinate denotes the sample size (in thousands), and the vertical coordinate denotes the ratio of estimated running time between CQ and SCQ.

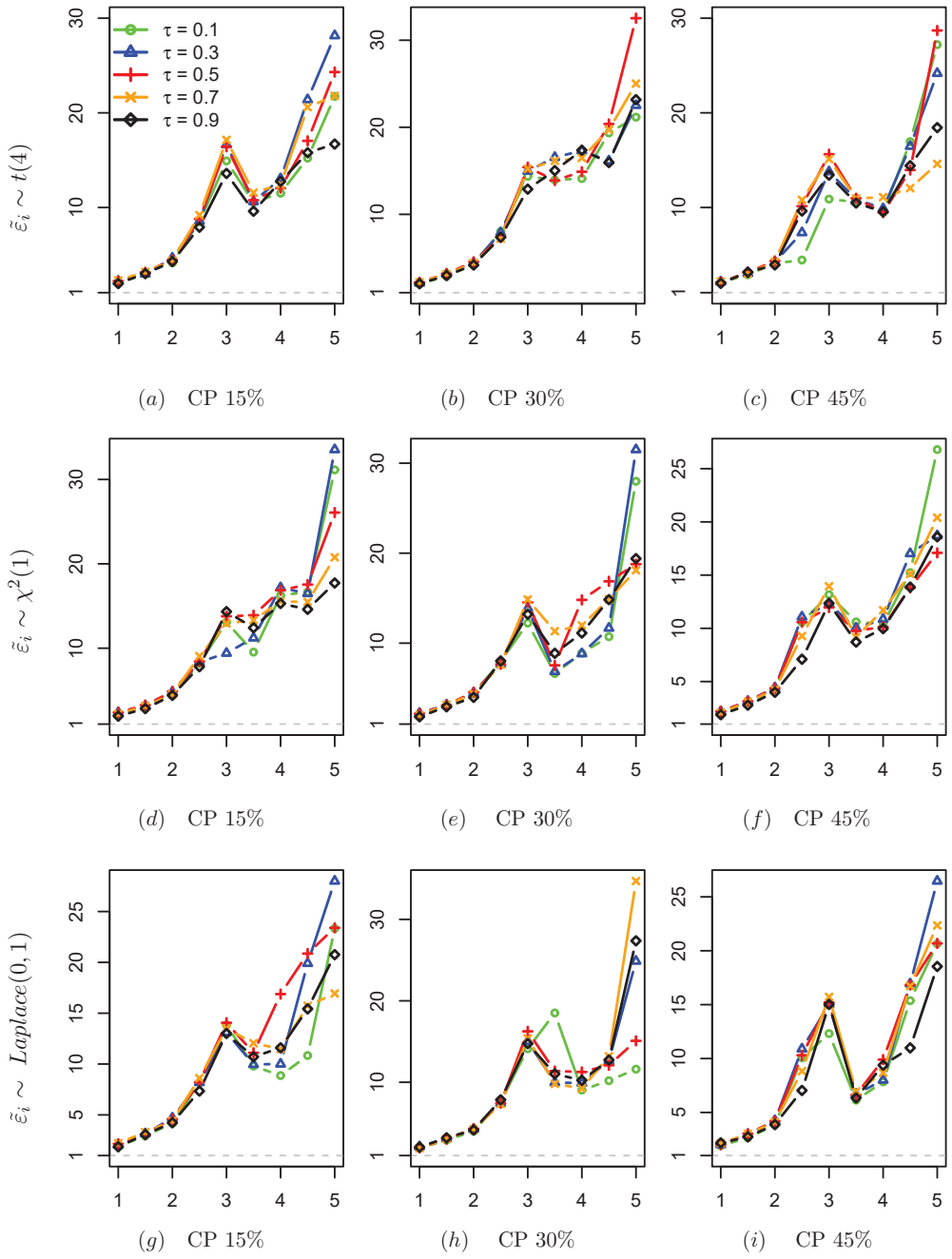


Figure A4. Simulation results under the high-dimensional model when the covariates are generated from Case 3, where the horizontal coordinate denotes the sample size (in thousands), and the vertical coordinate denotes the ratio of estimated running time between CQ and SCQ.

References

1. Koenker, R.; Gilbert, B., Jr. Regression quantiles. *Econometrica* **1978**, *46*, 33–50. [CrossRef]
2. Koenker, R. *Quantile Regression*; Cambridge University Press: Cambridge, UK, 2005.
3. Koenker, R.; Chernozhukov, V.; He, X.M.; Peng, L. *Handbook of Quantile Regression*; Chapman & Hall/CRC: Boca Raton, FL, USA, 2017.
4. Horowitz, J.L. Bootstrap methods for median regression models. *Econometrica* **1998**, *66*, 1327–1351. [CrossRef]
5. de Castro, L.; Galvao, A.F.; Kaplan, D.M.; Liu, X. Smoothed GMM for quantile models. *J. Econom.* **2019**, *213*, 121–144. [CrossRef]
6. Chen, X.; Liu, W.; Zhang, Y. Quantile regression under memory constraint. *Ann. Stat.* **2019**, *47*, 3244–3273. [CrossRef]
7. Galvao, A.F.; Kato, K. Smoothed quantile regression for panel data. *J. Econom.* **2016**, *193*, 92–112. [CrossRef]
8. Whang, Y.J. Smoothed empirical likelihood methods for quantile regression models. *Econ. Theory* **2006**, *22*, 173–205. [CrossRef]

9. Fernandes, M.; Guerre, E.; Horta, E. Smoothing quantile regressions. *J. Bus. Econom Stat.* **2021**, *39*, 338–357. [CrossRef]
10. He, X.; Pan, X.; Tan, K.M.; Zhou, W.X. Smoothed quantile regression with large-scale inference. *J. Econom.* **2023**, *232*, 367–388. [CrossRef] [PubMed]
11. Bayarassou, N.; Hamrani, F.; Ould, S.E. Nonparametric relative error estimation of the regression function for left truncated and right censored time series data. *J. Nonparametr. Stat.* **2023**, *36*, 706–729. [CrossRef]
12. Ying, Z.; Jung, S.H.; Wei, L.J. Survival analysis with median regression models. *J. Am. Stat. Assoc.* **1995**, *90*, 178–184. [CrossRef]
13. Honoré, B.; Khan, S.; Powell, J.L. Quantile regression under random censoring. *J. Econom.* **2002**, *109*, 67–105. [CrossRef]
14. Portnoy, S. Censored regression quantiles. *J. Am. Stat. Assoc.* **2003**, *98*, 1001–1012. [CrossRef]
15. Peng, L. Self-consistent estimation of censored quantile regression. *J. Multivar. Anal.* **2012**, *105*, 368–379. [CrossRef]
16. Yuan, X.; Zhang, X.; Guo, W.; Hu, Q. An adapted loss function for composite quantile regression with censored data. *Comput. Stat.* **2024**, *39*, 1371–1401. [CrossRef]
17. Gao, Q.; Zhou, X.; Feng, Y.; Du, X.; Liu, X. An empirical likelihood method for quantile regres sion models with censored data. *Metrika* **2021**, *84*, 75–96. [CrossRef]
18. Hao, R.; Weng, C.; Liu, X.; Yang, X. Data augmentation based estimation for the censored quantile regression neural network model. *Expert Syst. Appl.* **2023**, *214*, 119097. [CrossRef]
19. Yang, X.; Narisetty, N.N.; He, X. A new approach to censored quantile regression estimation. *J. Comput. Graph. Stat.* **2018**, *27*, 417–425. [CrossRef]
20. Peng, L.; Huang, Y. Survival analysis with quantile regression models. *J. Am. Stat. Assoc.* **2008**, *103*, 637–649. [CrossRef]
21. Xu, G.; Sit, T.; Wang, L.; Huang, C.Y. SEstimation and inference of quantile regression for survival data under biased sampling. *J. Am. Stat. Assoc.* **2017**, *112*, 1571–1586. [CrossRef] [PubMed]
22. Cai, Z.; Sit, T. On interquantile smoothness of censored quantile regression with induced smoothing. *Biometrics* **2023**, *79*, 3549–3563. [CrossRef] [PubMed]
23. Kim, K.H.; Caplan, D.J.; Kang, S. Smoothed quantile regression for censored residual life. *Comput Stat.* **2023**, *38*, 1001–1022. [CrossRef]
24. Wu, Y.; Ma, Y.; Yin, G. Smoothed and corrected score approach to censored quantile regression with measurement errors. *J. Am. Stat. Assoc.* **2015**, *110*, 1670–1683. [CrossRef]
25. He, X.; Pan, X.; Tan, K.M.; Zhou, W.X. Scalable estimation and inference for censored quantile regression process. *Ann. Stat.* **2022**, *50*, 2899–2924. [CrossRef]
26. Fei, Z.; Zheng, Q.; Hong, H.G.; Li, Y. Inference for high-dimensional censored quantile regression. *J. Am. Stat. Assoc.* **2023**, *118*, 898–912. [CrossRef] [PubMed]
27. De Backer, M.; Ghouch, A.E.; Van, K.I. An adapted loss function for cen sored quantile regression. *J. Am. Stat. Assoc.* **2019**, *114*, 1126–1137. [CrossRef]
28. Leng, C.; Tong, X. Censored quantile regression via Box-Cox transformation under conditional independence. *Stat. Sin.* **2014**, *24*, 221–249. [CrossRef]
29. Spokoiny, V. Bernstein–von Mises theorem for growing parameter dimension. *arXiv* **2013**, arXiv:1302.3430. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.