

# Single-index composite quantile regression for massive data

Rong Jiang<sup>a</sup>, Keming Yu<sup>b,\*</sup>

<sup>a</sup> Department of Statistics, Donghua University, Shanghai, 201620, China

<sup>b</sup> Brunel University London, UK



## ARTICLE INFO

### Article history:

Received 26 August 2019

Received in revised form 4 August 2020

Accepted 9 August 2020

Available online 19 August 2020

### AMS 2010 subject classifications:

60G08

62G20

### Keywords:

Composite quantile regression

Massive data

Single-index model

## ABSTRACT

Composite quantile regression (CQR) is becoming increasingly popular due to its robustness from quantile regression. Recently, the CQR method has been studied extensively with single-index models. However, the numerical inference of CQR methods for single-index models must involve iteration. In this study, we propose a non-iterative CQR (NICQR) estimation algorithm and derive the asymptotic distribution of the proposed estimator. Moreover, we extend the NICQR method to the analysis of massive datasets via a divide-and-conquer strategy. The proposed approach significantly reduces the computing time and the required primary memory. Simulation studies and two real data applications are conducted to illustrate the finite sample performance of the proposed methods.

© 2020 Elsevier Inc. All rights reserved.

## 1. Introduction

Single-index models provide an efficient way of coping with high-dimensional nonparametric estimation problems and avoid the “curse of dimensionality” by assuming that the response is only related to a single linear combination of the covariates. Because of its usefulness in several areas such as discrete choice analysis in econometrics and dose–response models in biometrics (Härdle et al. [5]), we restrict our attention to the single-index model in the following form:

$$Y = g_0(\mathbf{X}^\top \boldsymbol{\gamma}_0) + \varepsilon, \quad (1)$$

where  $Y$  is the univariate response and  $\mathbf{X}$  is a vector of the  $p$ -dimensional covariates. The function  $g_0(\cdot)$  is an unspecified, nonparametric smoothing function;  $\boldsymbol{\gamma}_0$  is the unknown index vector coefficient; and for the sake of identifiability, following Lin and Kulasekera [16], we assume that  $\|\boldsymbol{\gamma}_0\| = 1$  and that the first component of  $\boldsymbol{\gamma}_0$  is positive, where  $\|\cdot\|$  denotes the Euclidean norm and the error term  $\varepsilon$  is assumed to be independent of  $\mathbf{X}$  with  $E[\varepsilon] = 0$ .

To estimate the parameters in model (1), Xia and Härdle [30] proposed the minimum average variance estimation (MAVE) method, which was originally introduced by Xia et al. [31] for dimension reduction. Wu et al. [29] studied quantile regression (QR). Jiang et al. [9] studied CQR estimation for single-index model. These estimators need to be solved via an iterative procedure. That is, iteratively estimating both the nonparametric component and the parametric component usually involves high computational complexity. For this problem, the non-iterative procedure is studied; for example, Wang et al. [28] proposed a two-stage procedure and Liang et al. [15] employed a profile least squares approach. Christou and Akritas [3] developed a Nadaraya–Watson QR method for single-index models.

Existing non-iterative estimation procedures for single-index model were built on either least squares (LS) or QR methods. However, the LS method is sensitive to outliers and does not perform well when the error distribution is heavily

\* Corresponding author.

E-mail address: [keming.yu@brunel.ac.uk](mailto:keming.yu@brunel.ac.uk) (K. Yu).

skewed. The QR method is an obvious alternative to the LS. However, the relative efficiency of the QR can be arbitrarily small when compared with the LS. In contrast to the above methods, the CQR method was first proposed by Zou and Yuan [33] for estimating the regression coefficients in the classical linear regression model. The loss function of CQR is  $\sum_{k=1}^K \rho_{\tau_k}(\cdot)$ , where  $\rho_{\tau_k}(r) = \tau_k r - rI(r < 0)$  is the QR loss functions with  $k \in \{1, \dots, K\}$  and  $0 < \tau_1 < \dots < \tau_K < 1$ . It is easy to see that the CQR method is a sum of different quantile regressions. Zou and Yuan [33] showed that the CQR estimator shares robustness from QR and the relative efficiency of the CQR estimator compared with the LS estimator is greater than 70% regardless of the error distribution.

In this study, we propose a non-iterative method based on the CQR method for estimating the parametric component of model (1) to avoid such computational complexity. The proposed method is computationally more attractive while being as efficient as the iterative CQR method proposed by Jiang et al. [9]. Therefore, the proposed procedure is a valuable method to analyze massive datasets. There are two major challenges in analyzing massive datasets whose sizes usually exceed the capacity of a single computer: (i) the data can be too big to hold in a computer's memory and (ii) the computing task can take too long to obtain the results. As a solution to the memory and storage limitation problems, the divide-and-conquer method [2,17,24] could be an effective approach to ease the statistical analysis of massive datasets. Divide-and-conquer involves (i) dividing data into subsets, (ii) performing statistical analysis independently on each subset, and (iii) combining the results.

Combining the results in this way has long been studied in the statistical literature under the topic of meta. The classical meta-analysis method is based on the inverse variance weighted average of separate point estimates, each from one data batch. Lin and Zeng [18] showed that such a meta-estimator asymptotically achieves the same asymptotic variance as that based on the entire dataset. Xie et al. [32] developed a robust meta-analysis-type approach through the confidence distribution approach. Liu et al. [19] proposed combining the confidence density function in the same way as combining likelihood functions for inference. An advantage of the confidence distribution approach is rooted in the fact that it provides a unified framework for combining the distributions of estimators, thus allowing statistical inference with the combined estimator to be established in a straightforward and mathematically rigorous fashion.

Current references about massive datasets are based on the linear model (Lee et al. [13]), generalized linear model (Chen and Xie [2], Lin and Xi [17]), and nonparametric model (Kong and Xia [11]). There are no references about single-index models with massive datasets. Therefore, the goal of this study is to introduce a divide-and-conquer approach for single-index models with massive datasets, using the approach of combining the confidence density functions derived from the summary statistics of each subset's analysis.

Overall, this study offers a novel approach and makes the following key contributions:

(i) Our estimation procedure directly targets the model parameter  $\gamma_0$  in model (1) and no iteration is needed for the numerical computation. The method is robust under heavy-tailed noise distributions and is valid even when the first two moments of the noise distribution do not exist.

(ii) In terms of the limited existing work on single-index models for massive datasets, we develop a divide-and-conquer NICQR (DC-NICQR) method for single-index models with massive datasets. The proposed approach significantly reduces the required primary memory and the resulting estimates are as efficient as if the entire dataset was analyzed simultaneously.

The remainder of the paper is organized as follows. In Section 2, we introduce the NICQR procedure for model (1). We consider the NICQR method for massive datasets in Section 3. Both the simulation examples and the applications of two real datasets are given in Section 4 to illustrate the proposed procedures. Final remarks are given in Section 5. All the conditions and their discussions as well as technical proofs are presented in Section 6.

## 2. NICQR method for single-index models

In this section, we propose an NICQR estimation algorithm for single-index models.

### 2.1. NICQR method

Theoretically, the true parameter vector  $\gamma_0$  in model (1) solves the following minimization problem:

$$\gamma_0 = \arg \min_{\gamma} \sum_{k=1}^K E [\rho_{\tau_k} \{Y - Q_{\tau_k}(Y|\mathbf{X}^\top \gamma)\}], \quad (2)$$

where  $Q_{\tau_k}(Y|\mathbf{X}^\top \gamma) = c_k + g(\mathbf{X}^\top \gamma)$ , and  $c_k = F^{-1}(\tau_k)$ ,  $F(\cdot)$  is the cumulative distribution function of the model error  $\varepsilon$ .

For model (1),  $Q_{\tau_k}(Y|\mathbf{x}^\top \gamma) = \inf\{y : P(Y \leq y | \mathbf{X}^\top \gamma = \mathbf{x}^\top \gamma) \geq \tau_k\}$ . Since in (2),  $Q_{\tau_k}(Y|\mathbf{X}^\top \gamma)$ ,  $k \in \{1, \dots, K\}$ , are unknown, (2) should be minimized by solving one simple problem that  $Q_{\tau_k}(Y|\mathbf{X}^\top \gamma)$  must be replaced with an estimator for each  $k$ . However, there are no closed-form expressions for the estimator of  $Q_{\tau_k}(Y|\mathbf{X}^\top \gamma)$ . Thus, this often leads to iterative algorithms for estimating  $\gamma_0$ , which raises the computational complexity. To overcome this difficulty, we define, for any given  $\gamma \in R^p$ , the function  $H_{\tau_k}(t | \gamma) : R \rightarrow R$  as  $H_{\tau_k}(t | \gamma) = E \{Q_{\tau_k}(Y|\mathbf{X}) | \mathbf{X}^\top \gamma = t\}$ ,  $k \in \{1, \dots, K\}$ , where

$Q_{\tau_k}(Y|\mathbf{x}) = \inf\{y : P(Y \leq y | \mathbf{X} = \mathbf{x}) \geq \tau_k\}$ . Hence, under single-index models, this specifies that  $Q_{\tau_k}(Y|\mathbf{X}^\top \boldsymbol{\gamma}_0) = H_{\tau_k}(\mathbf{X}^\top \boldsymbol{\gamma}_0 | \boldsymbol{\gamma}_0)$ ,  $k \in \{1, \dots, K\}$ ; thus  $\boldsymbol{\gamma}_0$  also satisfies

$$\boldsymbol{\gamma}_0 = \arg \min_{\boldsymbol{\gamma}} \sum_{k=1}^K E[\rho_{\tau_k}\{Y - H_{\tau_k}(\mathbf{X}^\top \boldsymbol{\gamma} | \boldsymbol{\gamma})\}]. \quad (3)$$

Let  $\{Y_i, \mathbf{X}_i\}_{i=1}^n$  be an independent and identically distributed (i.i.d.) sample from  $\{Y, \mathbf{X}\}$ . Thus, the right term of (3) can be approximated by

$$\frac{1}{n} \sum_{i=1}^n \sum_{k=1}^K \rho_{\tau_k}\{Y_i - H_{\tau_k}(\mathbf{X}_i^\top \boldsymbol{\gamma} | \boldsymbol{\gamma})\}.$$

Thus, a non-iterative estimation process based on the Nadaraya–Watson estimator method can be constructed. First, we obtain the Nadaraya–Watson estimator of  $H_{\tau_k}(\cdot | \boldsymbol{\gamma})$  for each  $k$ :

$$\hat{H}_{\tau_k}(t | \boldsymbol{\gamma}) = \frac{\sum_{i=1}^n \hat{Q}_{\tau_k}(Y|\mathbf{X}_i) \tilde{K}_{h_k}(\mathbf{X}_i^\top \boldsymbol{\gamma} - t)}{\sum_{i=1}^n \tilde{K}_{h_k}(\mathbf{X}_i^\top \boldsymbol{\gamma} - t)}, \quad (4)$$

where  $\tilde{K}_{h_k}(\cdot) = \tilde{K}(\cdot/h_k)$ ,  $\tilde{K}(\cdot)$  is a univariate kernel function, and  $h_k$  is the bandwidth. Thus, we estimate  $\boldsymbol{\gamma}_0$  by solving the following minimization problem:

$$\hat{\boldsymbol{\gamma}} = \arg \min_{\boldsymbol{\gamma}} \frac{1}{n} \sum_{i=1}^n \sum_{k=1}^K \rho_{\tau_k}\{Y_i - \hat{H}_{\tau_k}(\mathbf{X}_i^\top \boldsymbol{\gamma} | \boldsymbol{\gamma})\}. \quad (5)$$

After obtaining the estimator  $\hat{\boldsymbol{\gamma}}$  of  $\boldsymbol{\gamma}_0$  in model (1) by (5), we can estimate  $g_0(\cdot)$  in model (1). We used the weighted local CQR (WLCQR) proposed by Jiang et al. [7], which is valid without a symmetric error condition. For any given point  $u$ , the final estimate of  $g_0(\cdot)$  is

$$\hat{g}(u | \hat{\boldsymbol{\gamma}}) = \sum_{k=1}^K v_k \hat{a}_k, \quad (6)$$

where the weight vector  $\mathbf{v} = (v_1, \dots, v_K)^\top$  satisfies conditions  $\sum_{k=1}^K v_k = 1$ ,  $\sum_{k=1}^K v_k c_k = 0$ , and

$$(\hat{a}_1, \dots, \hat{a}_K, \hat{b}) = \arg \min_{(a_1, \dots, a_K, b)} \frac{1}{n} \sum_{i=1}^n \sum_{k=1}^K \rho_{\tau_k}\{Y_i - a_k - b(\mathbf{X}_i^\top \hat{\boldsymbol{\gamma}} - u)\} \tilde{K}_h(\mathbf{X}_i^\top \hat{\boldsymbol{\gamma}} - u). \quad (7)$$

**Remark 1.** In general, given  $K$ , one can use the equally spaced quantiles at  $\tau_k = k/(K+1)$  for  $k \in \{1, \dots, K\}$ ; see Zou and Yuan [33]. Moreover, we can use redefined BIC (Tian et al. [26]) to select  $K$  as follows:

$$BIC(K) = \frac{2}{n} RS_K + \frac{\log(n)}{n} (2K + p - 1), \quad K \in \{1, \dots, K_{\max}\},$$

where  $K_{\max}$  is a possible upper bound and  $RS_K = \sum_{i=1}^n \sum_{k=1}^K \rho_{\tau_k}\{Y_i - \hat{H}_{\tau_k}(\mathbf{X}_i^\top \hat{\boldsymbol{\gamma}} | \hat{\boldsymbol{\gamma}})\}$  is the residual sum of the estimated model. The resulting optimal value of  $K$  is the smallest redefined BIC value.

**Remark 2.**  $\hat{Q}_{\tau_k}(Y|\mathbf{X}_i)$ ,  $k \in \{1, \dots, K\}$ , in (4) can be obtained by the D-vine copula proposed by Kraus and Czado [12], as follows:

$$\hat{Q}_{\tau_k}(Y|\mathbf{x}) = \hat{F}_Y^{-1}\left(\hat{C}_{F_Y|F_1, \dots, F_p}^{-1}\left(\tau_k | \hat{F}_1(x_1), \dots, \hat{F}_p(x_p)\right)\right),$$

where  $\hat{C}_{F_Y|F_1, \dots, F_p}^{-1}$  is the estimator of the conditional Copula quantile function  $C_{F_Y|F_1, \dots, F_p}^{-1}$ ,  $\hat{F}_Y(y) = (n+1)^{-1} \sum_{i=1}^n I(Y_i \leq y)$ , and  $\hat{F}_j(x_j) = (n+1)^{-1} \sum_{i=1}^n I(X_{ij} \leq x_j)$ ,  $j \in \{1, \dots, p\}$ . The detailed estimation process can be found in Section 3.2 of Kraus and Czado [12].

## 2.2. Asymptotic properties

Let  $f(\cdot)$  be the density function of the model error and denote by  $f_{\mathbf{U}_0}(\cdot)$  the marginal density function of  $\mathbf{U}_0 = \mathbf{X}^\top \boldsymbol{\gamma}_0$ . We choose the kernel  $\tilde{K}(\cdot)$  as a symmetric density function and write  $\mu_j = \int u^j \tilde{K}(u) du$ ,  $v_j = \int u^j \tilde{K}^2(u) du$ ,  $R_1 = \left\{ \sum_{k=1}^K f(c_k) \right\}^{-2} \sum_{k=1}^K \sum_{k'=1}^K \tau_{kk'}$ ,  $R_2(\mathbf{v}) = \sum_{k=1}^K \sum_{k'=1}^K (v_k v_{k'} \tau_{kk'}) / \{f(c_k) f(c_{k'})\}$ ,  $\tau_{kk'} = \tau_k \wedge \tau_{k'} - \tau_k \tau_{k'}$ , and  $h_{\max} = \max_{1 \leq k \leq K} \{h_k\}$ ,  $h_{\min} = \min_{1 \leq k \leq K} \{h_k\}$ .

**Theorem 1.** Suppose that Conditions C1–C4 given in the Appendix hold.  $n \rightarrow \infty$ ,  $nh_{\max}^4 \rightarrow 0$  and  $nh_{\min} \rightarrow \infty$ ; then,

$$\sqrt{n}(\hat{\gamma} - \gamma_0) \xrightarrow{L} \mathcal{N}(0, \mathbf{S}^+ \mathbf{R}_1), \quad (8)$$

where  $\xrightarrow{L}$  stands for convergence in the distribution,

$$\mathbf{S} = \mathbb{E} \left[ \{g'_0(\mathbf{X}^\top \gamma_0)\}^2 \{\mathbf{X} - \mathbb{E}(\mathbf{X} | \mathbf{X}^\top \gamma_0)\} \{\mathbf{X} - \mathbb{E}(\mathbf{X} | \mathbf{X}^\top \gamma_0)\}^\top \right],$$

and  $\mathbf{S}^+$  is the Moore–Penrose inverse of symmetric matrix  $\mathbf{S}$ , since  $\mathbf{S}$  is not full ranked (see Ma and He [20] and Tang et al. [25]).

**Theorem 2.** Under the same conditions as in Theorem 1, if  $n \rightarrow \infty$ ,  $h \rightarrow 0$  and  $nh \rightarrow \infty$ , then for an interior point  $u$  of the support of  $f_{u_0}(\cdot)$ ,

$$\sqrt{nh} \left\{ \hat{g}(u | \hat{\gamma}) - g_0(u) - \frac{1}{2} g''_0(u) \mu_2 h^2 \right\} \xrightarrow{L} \mathcal{N} \left( 0, \frac{v_0 R_2(\mathbf{v})}{f_{u_0}(u)} \right).$$

The bias of  $\hat{g}(u | \hat{\gamma})$  is independent of the choice of the weight vector  $\mathbf{v}$ , and only the variance term depends on the weight vector  $\mathbf{v}$ . Then, the optimal weights correspond to the minimum asymptotic variance of  $\hat{g}(u | \hat{\gamma})$ . Thus,

$$\mathbf{v}_{opt} = \arg \min_{\mathbf{v}} R_2(\mathbf{v}) = \frac{(\mathbf{c}^\top \mathbf{A}^{-1} \mathbf{c}) \mathbf{A}^{-1} \mathbf{1} - (\mathbf{1}^\top \mathbf{A}^{-1} \mathbf{c}) \mathbf{A}^{-1} \mathbf{c}}{(\mathbf{c}^\top \mathbf{A}^{-1} \mathbf{c})(\mathbf{1}^\top \mathbf{A}^{-1} \mathbf{1}) - (\mathbf{1}^\top \mathbf{A}^{-1} \mathbf{c})^2}, \quad (9)$$

where  $\mathbf{c}$  is a  $K$ -dimensional column vector with the  $k$ th element  $c_k$ ,  $\mathbf{1}$  is a  $K$ -dimensional column vector with all elements 1, and  $\mathbf{A}$  is a  $K \times K$  matrix with  $(k, k')$ -element  $\tau_{kk'} / (f(c_k) f(c_{k'}))$ ,  $k, k' \in \{1, \dots, K\}$ . Thus, with these optimal weights, the asymptotic variance of  $\hat{g}(u | \hat{\gamma})$  is  $f_{u_0}^{-1}(u) v_0 R_2(\mathbf{v}_{opt})$ .

**Remark 3.** For Theorem 1, the proposed estimator achieves the same efficiency as the iterative CQR estimator proposed by Jiang et al. [9]. The results of Theorem 2 are thus the same as those of Theorem 5 in Jiang et al. [7]. Thus, the selection of the optimal bandwidth  $h$  can be found in Jiang et al. [7].

**Remark 4.** From (9), we can see that the optimal weight vector  $\mathbf{v}_{opt}$  is complicated and involves the density of the errors  $f(c_k)$ ,  $k \in \{1, \dots, K\}$ . In practice, the error density  $f(c_k)$  is generally unknown. Typical nonparametric density estimation methods such as kernel smoothing based on the estimated residual  $\hat{\epsilon}$  can provide a consistent estimation  $\hat{f}(\hat{c}_k)$  of  $f(c_k)$ ,  $k \in \{1, \dots, K\}$ . The details can be found in Section 2.3 of Jiang et al. [7].

### 2.3. Asymptotic relative efficiency (ARE)

In this section, we first investigate the ARE of the NICQR method relative to the MAVE method proposed by Xia et al. [31]. The asymptotic variance of the MAVE method is  $\mathbf{S}^+$  under the homoscedastic model. Therefore, the ARE of the CQR method with respect to the MAVE method is

$$ARE_{\gamma_0}(\text{NICQR}, \text{MAVE}) = R_1^{-1}.$$

Note that  $ARE_{\gamma_0}(\text{NICQR}, \text{MAVE})$  is the same as the result obtained by Zou and Yuan [33]. Thus,

$$ARE_{\gamma_0}(\text{NICQR}, \text{MAVE}) \geq 70\%.$$

Next, we consider the ARE of the WLCQR method relative to the mean regression by adopting the MAVE method and the local CQR (LCQR) method proposed by Jiang et al. [9]. From a similar deduction to that of Jiang et al. [8], we find that the asymptotic efficiency of the WLCQR estimation of  $g_0(\cdot)$  relative to the MAVE and LCQR estimations for the case of symmetric errors is

$$ARE_{g_0}(\text{WLCQR}, \text{MAVE}) = R_2(\mathbf{v}_{opt})^{-4/5}, \quad ARE_{g_0}(\text{WLCQR}, \text{LCQR}) = \left[ \frac{R_2(\mathbf{v}_{opt})}{R_3} \right]^{-4/5},$$

where  $R_3 = \sum_{k=1}^K \sum_{k'=1}^K [\tau_{kk'} / \{f(c_k) f(c_{k'})\}] / K^2$ . Jiang et al. [8] showed that when the error distribution is symmetric, we have

$$\lim_{K \rightarrow \infty} \inf ARE_{g_0}(\text{WLCQR}, \text{MAVE}) \geq 1, \quad \lim_{K \rightarrow \infty} \inf ARE_{g_0}(\text{WLCQR}, \text{LCQR}) \geq 1.$$

**Remark 5.** To appreciate how much efficiency  $ARE_{\gamma_0}(\text{NICQR}, \text{MAVE})$ ,  $ARE_{g_0}(\text{WLCQR}, \text{MAVE})$ , and  $ARE_{g_0}(\text{WLCQR}, \text{LCQR})$  is gained in practice, Fig. 2 in Zou and Yuan [33] reports  $ARE_{\gamma_0}(\text{NICQR}, \text{MAVE})$  with various error distributions for various  $K$  and Table 2 in Jiang et al. [8] reports  $ARE_{g_0}(\text{WLCQR}, \text{MAVE})$  and  $ARE_{g_0}(\text{WLCQR}, \text{LCQR})$  with various error distributions for various  $K$ .

## 2.4. Algorithm

To obtain the estimator  $\hat{\gamma}$  by minimizing (5), we use the procedure introduced in Wang and Wu [27], which consists of using a local linear approximation of  $\hat{H}_{\tau_k}(\mathbf{X}_i^\top \boldsymbol{\gamma} \mid \boldsymbol{\gamma})$  around an initial value  $\tilde{\boldsymbol{\gamma}}$  of  $\boldsymbol{\gamma}_0$ . This yields

$$\hat{H}_{\tau_k}(\mathbf{X}_i^\top \boldsymbol{\gamma} \mid \boldsymbol{\gamma}) \approx \hat{H}_{\tau_k}(\mathbf{X}_i^\top \tilde{\boldsymbol{\gamma}} \mid \tilde{\boldsymbol{\gamma}}) + \hat{H}'_{\tau_k}(\mathbf{X}_i^\top \tilde{\boldsymbol{\gamma}} \mid \tilde{\boldsymbol{\gamma}})(\boldsymbol{\gamma} - \tilde{\boldsymbol{\gamma}}),$$

where  $\hat{H}'_{\tau_k}(\mathbf{X}_i^\top \tilde{\boldsymbol{\gamma}} \mid \tilde{\boldsymbol{\gamma}}) = \partial \hat{H}_{\tau_k}(\mathbf{X}_i^\top \boldsymbol{\gamma} \mid \boldsymbol{\gamma}) / \partial \boldsymbol{\gamma} |_{\boldsymbol{\gamma}=\tilde{\boldsymbol{\gamma}}}$ . Then, the proposed estimator can be obtained from

$$\hat{\boldsymbol{\gamma}} = \arg \min_{\boldsymbol{\gamma}} \frac{1}{n} \sum_{i=1}^n \sum_{k=1}^K \rho_{\tau_k} \left\{ Y_i - \hat{H}_{\tau_k}(\mathbf{X}_i^\top \tilde{\boldsymbol{\gamma}} \mid \tilde{\boldsymbol{\gamma}}) - \hat{H}'_{\tau_k}(\mathbf{X}_i^\top \tilde{\boldsymbol{\gamma}} \mid \tilde{\boldsymbol{\gamma}})(\boldsymbol{\gamma} - \tilde{\boldsymbol{\gamma}}) \right\}. \quad (10)$$

The steps of the NICQR procedure are summarized as follows.

**Step 1 (Initialization step).** Obtain an initial estimate  $\tilde{\boldsymbol{\gamma}}$  and  $\tilde{g}(\cdot)$  by using the MAVE method, which can be obtained from the R package MAVE.

**Step 2.**  $\hat{Q}_{\tau_k}(\mathbf{Y} \mid \mathbf{X}_i)$  in (4) is obtained by the D-vine copula proposed by Kraus and Czado [12], which can be implemented by using the R package `vinereg`. Thus, we can obtain  $\hat{H}_{\tau_k}(\mathbf{X}_i^\top \tilde{\boldsymbol{\gamma}} \mid \tilde{\boldsymbol{\gamma}})$  and  $\hat{H}'_{\tau_k}(\mathbf{X}_i^\top \tilde{\boldsymbol{\gamma}} \mid \tilde{\boldsymbol{\gamma}})$ .

**Step 3 (Estimation of  $\boldsymbol{\gamma}_0$ ).** Update  $\hat{\boldsymbol{\gamma}}$  by minimizing the objective function in (10), which can be achieved by using the `cqr.cd` in the R package `cqrReg`.

**Step 4.** Obtain the estimate  $\hat{\mathbf{v}}$  of  $\mathbf{v}_{opt}$  in (9). The estimation of  $\hat{\mathbf{f}} = (\hat{f}(\hat{c}_1), \dots, \hat{f}(\hat{c}_K))^\top$  is as follows: compute  $\hat{\varepsilon}_i = Y_i - \tilde{g}(\mathbf{X}_i^\top \tilde{\boldsymbol{\gamma}})$  from the results in Step 1 and then use the kernel density estimation  $\hat{f}(\cdot) = (n\tilde{h})^{-1} \sum_{i=1}^n \tilde{K}_{\tilde{h}}(\hat{\varepsilon}_i - \cdot)$  to estimate  $f(\cdot)$ . The  $\hat{c}_k$  of  $c_k$  is the sample  $\tau_k$ -quantile of  $\{\hat{\varepsilon}_i\}_{i=1}^n$ , where the bandwidth  $\tilde{h}$  is chosen from  $\tilde{h} = 0.9 \times \min \{ \text{std}(\hat{\varepsilon}_1, \dots, \hat{\varepsilon}_n), \text{IQR}(\hat{\varepsilon}_1, \dots, \hat{\varepsilon}_n)/1.34 \} \times n^{-1/5}$ ,  $\text{std}$  and  $\text{IQR}$  denote the sample standard deviation and sample interquantile, respectively (Jiang et al. [8]).

**Step 5 (Estimation of  $g_0(\cdot)$ ).** From the estimate  $\hat{\boldsymbol{\gamma}}$  in Step 3, for a given point  $u$ , estimate  $g_0(u)$  from (6) with  $\hat{\mathbf{v}}$  in Step 4, which can be obtained by using a modified `cqr.cd` in the R package `cqrReg`.

## 3. NICQR method for massive datasets

In this section, we propose a DC-NICQR estimation algorithm for massive datasets analysis.

### 3.1. DC-NICQR method

It is infeasible to solve the optimization problem in (5) and (7) when sample size  $n$  is too large. To solve the above problem, we consider a divide-and-conquer method that divides the dataset into several blocks, each containable in the computer's memory. Without loss of generality, the entire dataset is partitioned into  $M$  subsets and the  $m$ th subset contains  $n_m$  observations:  $(\mathbf{X}_{m,i}, Y_{m,i})$ ,  $i \in \{1, \dots, n_m\}$ , and  $n = \sum_{m=1}^M n_m$ . Here we assume  $M$  is fixed. Based on the asymptotic normality in (8), we form the asymptotic confidence density of  $\boldsymbol{\gamma}_0$  as

$$\hat{h}_n(\boldsymbol{\gamma}_0) \propto \exp \left\{ -\frac{n}{2} R_1^{-1} (\boldsymbol{\gamma}_0 - \hat{\boldsymbol{\gamma}})^\top \mathbf{S} (\boldsymbol{\gamma}_0 - \hat{\boldsymbol{\gamma}}) \right\}.$$

Moreover, a data-driven version of the asymptotic confidence density is given by

$$\hat{h}_n(\boldsymbol{\gamma}_0) \propto \exp \left\{ -\frac{n}{2} \hat{R}_1^{-1} (\boldsymbol{\gamma}_0 - \hat{\boldsymbol{\gamma}})^\top \hat{\mathbf{S}} (\boldsymbol{\gamma}_0 - \hat{\boldsymbol{\gamma}}) \right\},$$

where  $\hat{R}_1 = \left\{ \sum_{k=1}^K \hat{f}(\hat{c}_k) \right\}^{-2} \sum_{k=1}^K \sum_{k'=1}^K \tau_{kk'}$  is the estimation of  $R_1$  and  $\hat{\mathbf{S}} = n^{-1} \sum_{i=1}^n \left\{ \hat{H}'(\mathbf{X}_i^\top \hat{\boldsymbol{\gamma}} \mid \hat{\boldsymbol{\gamma}}) \right\}^\top \hat{H}'(\mathbf{X}_i^\top \hat{\boldsymbol{\gamma}} \mid \hat{\boldsymbol{\gamma}})$  is the estimation of  $\mathbf{S}$ . Note that  $\hat{H}_{\tau_k}(\mathbf{X}_i^\top \boldsymbol{\gamma} \mid \boldsymbol{\gamma})$  is the estimator of  $H_{\tau_k}(\mathbf{X}^\top \boldsymbol{\gamma}_0 \mid \boldsymbol{\gamma}_0) = Q_{\tau_k}(\mathbf{Y} \mid \mathbf{X}^\top \boldsymbol{\gamma}_0) = c_k + g(\mathbf{X}^\top \boldsymbol{\gamma}_0)$ . Thus,  $\hat{H}'_{\tau_k}(\mathbf{X}^\top \boldsymbol{\gamma} \mid \boldsymbol{\gamma})$  is the estimator of  $g'_0(\mathbf{X}^\top \boldsymbol{\gamma}_0) \{ \mathbf{X} - \mathbf{E}(\mathbf{X} \mid \mathbf{X}^\top \boldsymbol{\gamma}_0) \}^\top$ , which is independent of  $\tau_k$ . A similar result can be found in Section 4.1 of Christou and Akritas [3]. Therefore, we rewrite  $\hat{H}'_{\tau_k}(\mathbf{X}^\top \boldsymbol{\gamma} \mid \boldsymbol{\gamma})$  as  $\hat{H}'(\mathbf{X}^\top \boldsymbol{\gamma} \mid \boldsymbol{\gamma})$ . In the simulation, we take  $\hat{H}'(\mathbf{X}^\top \boldsymbol{\gamma} \mid \boldsymbol{\gamma})$  with  $\hat{H}'_{1/2}(\mathbf{X}^\top \boldsymbol{\gamma} \mid \boldsymbol{\gamma})$ .

It is also infeasible to obtain  $\hat{h}_n(\boldsymbol{\gamma}_0)$  when sample size  $n$  is too large. Therefore, by again considering the divide-and-conquer method, for each sub-dataset  $(\mathbf{X}_m, Y_m)$ , we first apply (8) to construct the asymptotic confidence density  $\hat{h}_{n_m}(\boldsymbol{\gamma}_0)$ ,  $m \in \{1, \dots, M\}$ . Then, we can combine the  $M$  confidence densities to derive a combined estimator of  $\boldsymbol{\gamma}_0$ . The combined estimator is denoted by  $\hat{\boldsymbol{\gamma}}^{DC}$  according to the following equation:

$$\hat{\boldsymbol{\gamma}}^{DC} = \arg \max_{\boldsymbol{\gamma}} \ln \prod_{m=1}^M \hat{h}_{n_m}(\boldsymbol{\gamma}) = \arg \min_{\boldsymbol{\gamma}} \sum_{m=1}^M \frac{n_m}{2} \hat{R}_{1,m}^{-1} (\boldsymbol{\gamma} - \hat{\boldsymbol{\gamma}}_m)^\top \hat{\mathbf{S}}_m (\boldsymbol{\gamma} - \hat{\boldsymbol{\gamma}}_m), \quad (11)$$

where  $\hat{\gamma}_m$  is the estimation of  $\gamma_0$  by using the methodology to solve Eq. (5) for the  $m$ th subset, and

$$\hat{R}_{1,m} = \left\{ \sum_{k=1}^K \hat{f}_m(\hat{c}_{k,m}) \right\}^{-2} \sum_{k=1}^K \sum_{k'=1}^K \tau_{kk'}, \quad \hat{S}_m = \frac{1}{n_m} \sum_{i=1}^{n_m} \left\{ \hat{H}'(\mathbf{X}_{m,i}^\top \hat{\gamma}_m \mid \hat{\gamma}_m) \right\}^\top \hat{H}'(\mathbf{X}_{m,i}^\top \hat{\gamma}_m \mid \hat{\gamma}_m).$$

We can then use the full data to estimate  $f_m(c_{k,m})$  because of the i.i.d. error setting (the details can be found in (13) in Section 3.3). Thus, we use  $\hat{R}_{1,m} = \hat{R}_1$  for  $m \in \{1, \dots, M\}$ . By some simple algebra, the solution to the optimization problem in (11) can be expressed as a form of the weighted average of  $\hat{\gamma}_m$ ,  $m \in \{1, \dots, M\}$ , as

$$\hat{\gamma}^{DC} = \left\{ \sum_{m=1}^M n_m \hat{R}_1^{-1} \hat{S}_m \right\}^{-1} \sum_{m=1}^M n_m \hat{R}_1^{-1} \hat{S}_m \hat{\gamma}_m = \left\{ \sum_{m=1}^M n_m \hat{S}_m \right\}^+ \sum_{m=1}^M n_m \hat{S}_m \hat{\gamma}_m.$$

In summary, we show that the DC-NICQR method can be obtained by using the following three key steps:

**Step 1.** Without loss of generality, the entire dataset is partitioned into  $M$  subsets; the  $m$ th subset contains  $n_m$  observations:  $(\mathbf{X}_{m,i}, Y_{m,i})$ ,  $i \in \{1, \dots, n_m\}$ , and  $n = \sum_{m=1}^M n_m$ .

**Step 2.** For each subset, obtain the estimators  $\hat{\gamma}_m$ ,  $m \in \{1, \dots, M\}$ , using the methodology to solve Eq. (5). The aggregated estimator for  $\gamma_0$ , as a weighted average of  $\hat{\gamma}_m$ ,  $m \in \{1, \dots, M\}$ , is

$$\hat{\gamma}^{DC} = \left\{ \sum_{m=1}^M n_m \hat{S}_m \right\}^+ \sum_{m=1}^M n_m \hat{S}_m \hat{\gamma}_m.$$

**Step 3.** After obtaining the estimation  $\hat{\gamma}^{DC}$  of  $\gamma_0$  in model (1), we can estimate  $g_0(\cdot)$  in model (1). For any given point  $u$ , we find the estimators  $\hat{g}_m(u)$  for each subset, using the methodology in (6). Then, the final estimate of  $g_0(u \mid \hat{\gamma}^{DC})$ , as a weighted average of  $\hat{g}_m(u \mid \hat{\gamma}^{DC})$ ,  $m \in \{1, \dots, M\}$ , is

$$\hat{g}^{DC}(u \mid \hat{\gamma}^{DC}) = \frac{1}{n} \sum_{m=1}^M n_m \hat{g}_m(u \mid \hat{\gamma}^{DC}).$$

### 3.2. Asymptotic normality of the resulting estimator

To reveal the advantages of the proposed divide-and-conquer methods, we now establish the asymptotic normalities of  $\hat{\gamma}^{DC}$  and  $\hat{g}(\cdot)^{DC}$ .

**Theorem 3.** Assume that the conditions of Theorem 1 are satisfied, then

$$\sqrt{n}(\hat{\gamma}^{DC} - \gamma_0) \xrightarrow{L} \mathcal{N}(0, \mathbf{S}^+ R_1).$$

**Theorem 4.** Assume that the conditions of Theorem 2 are satisfied with the optimal weight  $\mathbf{v}_{opt}$  in (9), then

$$\sqrt{nh} \left\{ \hat{g}^{DC}(u \mid \hat{\gamma}^{DC}) - g_0(u) - \frac{1}{2} g_0''(u) \mu_2 h^2 \right\} \xrightarrow{L} \mathcal{N} \left( 0, \frac{\nu_0 R_2(\mathbf{v}_{opt})}{f_{U_0}(u)} \right).$$

**Remark 6.** The limiting distributions of  $\hat{\gamma}^{DC}$  and  $\hat{g}(\cdot)^{DC}$  in Theorems 3 and 4 are those of  $\hat{\gamma}$  and  $\hat{g}(\cdot)$  in Theorems 1 and 2, where all the data are analyzed. Thus, the DC-NICQR estimators are asymptotically equivalent to the corresponding estimator using the full datasets.

**Remark 7.** In Theorems 3 and 4, the resulting estimate is robust to the choice of block size  $M$  and subset size  $n_m$ , where  $m \in \{1, \dots, M\}$ . Thus,  $M$  and  $n_m$  are chosen so that the estimation of  $\gamma_0$  can be easily handled within each block.

**Remark 8.** In Step 3, we use the same bandwidth  $h$  for each subset by the method in Remark 3.

### 3.3. Estimation of the optimal weights $\mathbf{v}_{opt}$ for massive datasets

As mentioned in Remark 4, the optimal weight vector  $\mathbf{v}_{opt}$  involves the density of the errors  $f(c_k)$ ,  $k \in \{1, \dots, K\}$ . We can use the kernel density estimation  $n^{-1} \sum_{i=1}^n \tilde{K}_h(\hat{\epsilon}_i - \cdot)$  to estimate  $f(\cdot)$ . The estimator  $\hat{c}_k$  of  $c_k$  is the sample  $\tau_k$ -quantile of  $\{\hat{\epsilon}_i\}_{i=1}^n$ , where  $\hat{\epsilon}_i = Y_i - \hat{g}^{DC}(\mathbf{X}_i^\top \hat{\gamma}^{DC})$ . Therefore, we can obtain the estimation of  $f(c_k)$  from  $\hat{f}(\hat{c}_k)$  for  $k \in \{1, \dots, K\}$ . However, when the available computer memory is much smaller than  $n$ , sorting  $\{\hat{\epsilon}_i\}_{i=1}^n$  becomes impossible. To overcome this difficulty, Li et al. [14] proposed an approach to estimate the population parameters in a massive dataset. Their



method reduces the required primary memory, and the resulting estimate is as efficient as if the entire dataset was analyzed simultaneously. By following the method of Li et al. [14], we can estimate  $c_k$ ,  $k \in \{1, \dots, K\}$ , as follows:

$$\hat{c}_k = \frac{1}{n} \sum_{m=1}^M n_m \hat{c}_{k,m}, \quad (12)$$

where  $\hat{c}_{k,m}$  is the sample  $\tau_k$ -quantile of  $\{\hat{\varepsilon}_{m,i}\}_{i=1}^{n_m}$  and  $\hat{\varepsilon}_{m,i} = Y_{m,i} - \hat{g}_m(\mathbf{X}_{m,i}^\top \hat{\mathbf{y}}^{DC})$ . Then, from (12), the weighted combined estimator of  $f(c_k)$ ,  $k \in \{1, \dots, K\}$ , is given by

$$\hat{f}^{DC}(\hat{c}_k) = \frac{1}{n} \sum_{m=1}^M n_m \hat{f}_m(\hat{c}_k), \quad (13)$$

where  $\hat{f}_m(\hat{c}_k) = \frac{1}{n_m} \sum_{i=1}^{n_m} \tilde{K}_{\tilde{h}_m}(\hat{\varepsilon}_{m,i} - \hat{c}_k)$ ,  $m \in \{1, \dots, M\}$ , are the kernel density estimations within each subset. Then, from (13), we can obtain

$$\hat{\mathbf{v}} = \frac{(\hat{\mathbf{c}}^\top \hat{\mathbf{A}}^{-1} \hat{\mathbf{c}}) \hat{\mathbf{A}}^{-1} \mathbf{1} - (\mathbf{1}^\top \hat{\mathbf{A}}^{-1} \hat{\mathbf{c}}) \hat{\mathbf{A}}^{-1} \hat{\mathbf{c}}}{(\hat{\mathbf{c}}^\top \hat{\mathbf{A}}^{-1} \hat{\mathbf{c}})(\mathbf{1}^\top \hat{\mathbf{A}}^{-1} \mathbf{1}) - (\mathbf{1}^\top \hat{\mathbf{A}}^{-1} \hat{\mathbf{c}})^2},$$

where  $\hat{\mathbf{c}}$  is a  $K$ -dimensional column vector with the  $k$ th element  $\hat{c}_k$  and  $\hat{\mathbf{A}}$  is a  $K \times K$  matrix with the  $(k, k')$  element  $\tau_{kk'} / (\hat{f}^{DC}(\hat{c}_k) \hat{f}^{DC}(\hat{c}_{k'}))$ ,  $k, k' \in \{1, \dots, K\}$ .

**Remark 9.** The bandwidth selection is taken by  $\tilde{h}_m = (n_m/n)^{1/5} h_m^*$  as selected by Li et al. [14], where  $h_m^*$  is selected as  $0.9 \times 1.06 \times \sigma_m \times n_m^{-1/5}$  and  $\hat{\sigma}_m = \text{std}(\hat{\varepsilon}_{m,1}, \dots, \hat{\varepsilon}_{m,n_m})$  to estimate  $\sigma_m$ .

#### 4. Numerical studies

In this section, we first use Monte Carlo simulation studies to assess the finite sample performance of the proposed procedures and then demonstrate the application of the proposed methods with two real data analyses. Tian et al. [26] proposed redefined BIC to select the number of composite quantiles  $K$ . However, the performance of the CQR method with different  $K$  values is similar in their simulation. Moreover, from Tables 1 and 2 in Jiang et al. [8], we see that  $K = 9$  is a good choice for single-index model. Therefore, we choose  $K = 9$  as a compromise between the estimation and computational efficiency of the CQR method and let the equally spaced quantile levels be  $\tau_k = k/10$ ,  $k \in \{1, \dots, 9\}$ . All programs are written in R and our computer has a 2.4 GHz Pentium processor and 4G memory.

##### 4.1. Example for the NICQR method

In this section, we include five competitors in our comparison:

- (1) MAVE (see Xia and Härdle [30]);
- (2) QR with  $\tau = 0.5$  (QR<sub>0.5</sub>) (see Wu et al. [29]);
- (3) CQR with  $K = 9$  (CQR<sub>9</sub>) (see Jiang et al. [9]);
- (4) Non-iterative LS estimation (NILSE) (see Wang and Wu [27]); and
- (5) Non-iterative QR with  $\tau = 0.5$  (NIQR<sub>0.5</sub>) (see Christou and Akritas [3]),

where methods (1)–(3) need to be solved via an iterative procedure and (4) and (5) are non-iterative estimation algorithms.

##### 4.1.1. Simulation Example 1

We conduct a small simulation study with  $n = 200$  and the data are generated from the following “sine-bump” model:

$$\mathbf{Y} = \sin\{\pi(\mathbf{X}^\top \boldsymbol{\gamma}_0 - A)/(B - A)\} + 0.2\boldsymbol{\varepsilon}, \quad (14)$$

where  $\mathbf{X}$  is uniformly distributed on  $[0, 1]^3$ ,  $\boldsymbol{\gamma}_0 = (1, 1, 1)^\top / \sqrt{3}$ , and  $A = \sqrt{3}/2 - 1.645/\sqrt{12}$  and  $B = \sqrt{3}/2 + 1.645/\sqrt{12}$  are taken to ensure that the design is relatively thick in the tail. In our simulation, we consider three error distributions for  $\boldsymbol{\varepsilon}$ : a standard normal distribution ( $\mathcal{N}(0, 1)$ ), uniformly distributed on  $[-2, 2]$  ( $\mathcal{U}(-2, 2)$ ), and a Chi-square distribution with three degrees of freedom ( $\chi^2(3)$ ). All the simulations are run for 500 replicates.

Table 1 depicts the mean squared errors ( $MSE = \sqrt{(\hat{\mathbf{y}} - \boldsymbol{\gamma}_0)^\top (\hat{\mathbf{y}} - \boldsymbol{\gamma}_0)}$ ), and absolute bias ( $AB = |\hat{\mathbf{y}} - \boldsymbol{\gamma}_0|$ ) of the estimate  $\hat{\mathbf{y}}$  to assess the accuracy of the estimation methods. From Table 1, the following conclusions can be drawn:

- (i) All the estimators are close to the true value because the absolute bias is very small.
- (ii) The NICQR<sub>9</sub> estimator performs better than NIQR<sub>9</sub> and is close to CQR<sub>9</sub> for different error distributions. Moreover, NICQR<sub>9</sub> is consistently superior to the other four methods except CQR<sub>9</sub> when the error is the Chi-square distribution.
- (iii)  $t$  in Table 1 and Fig. 1 are the average computing time in seconds used to estimate the index parameter. From  $t$ , we see that the operation time of NICQR<sub>9</sub> is faster than that of CQR<sub>9</sub> and with the increase of  $n$ , the gap is more obvious. Moreover, for the uniform and Chi-square error distributions, the operation time of NICQR<sub>9</sub> is faster than that of QR<sub>0.5</sub>.

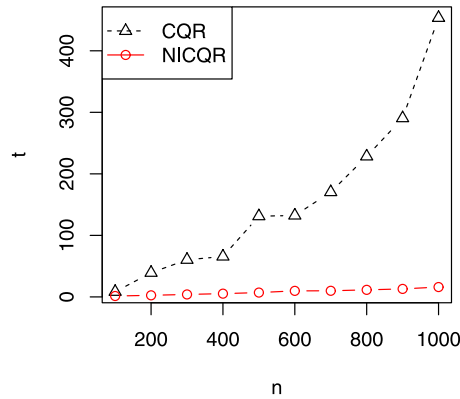


Fig. 1. The average computing time (in seconds) of estimated  $\gamma_0$  in model (14) based on 500 estimates by  $\text{CQR}_9$  and  $\text{NICQR}_9$  for different  $n$ .

Table 1

This table deals with comparison of some estimation methods (MAVE;  $\text{QR}_{0.5}$ ;  $\text{CQR}_9$ ; NILSE;  $\text{NIQR}_{0.5}$ ;  $\text{NICQR}_9$ ) based on the mean values and standard errors (in parenthesis) of absolute bias (AB), mean squared errors (MSE) and computing time ( $t$ ) based on 500 estimates for the parameter in model (14) under three different error distributions ( $\mathcal{N}(0, 1)$ ;  $\mathcal{N}(-2, 2)$ ;  $\chi^2(3)$ ).

Error	Methods	AB of $\gamma_1$	AB of $\gamma_2$	AB of $\gamma_3$	MSE	$t$
$\mathcal{N}(0, 1)$	MAVE	0.0188 (0.0137)	0.0183 (0.0144)	0.0171 (0.0126)	0.0349 (0.0178)	0.2831
	$\text{QR}_{0.5}$	0.0213 (0.0160)	0.0195 (0.0152)	0.0194 (0.0144)	0.0386 (0.0204)	1.2297
	$\text{CQR}_9$	0.0269 (0.0155)	0.0192 (0.0229)	0.0229 (0.0154)	0.0441 (0.0174)	47.4721
	NILSE	0.0218 (0.0165)	0.0207 (0.0166)	0.0203 (0.0147)	0.0399 (0.0220)	0.3156
	$\text{NIQR}_{0.5}$	0.0333 (0.0259)	0.0322 (0.0261)	0.0322 (0.0239)	0.0629 (0.0337)	0.5427
	$\text{NICQR}_9$	0.0309 (0.0251)	0.0319 (0.0233)	0.0287 (0.0235)	0.0586 (0.0329)	2.8645
$\mathcal{N}(-2, 2)$	MAVE	0.0245 (0.0180)	0.0250 (0.0198)	0.0218 (0.0167)	0.0459 (0.0242)	0.2932
	$\text{QR}_{0.5}$	0.0328 (0.0243)	0.0317 (0.0254)	0.0306 (0.0230)	0.0609 (0.0328)	3.5831
	$\text{CQR}_9$	0.0347 (0.0248)	0.0357 (0.0249)	0.0230 (0.0172)	0.0600 (0.0296)	44.6668
	NILSE	0.0245 (0.0188)	0.0238 (0.0201)	0.0219 (0.0167)	0.0450 (0.0258)	0.3279
	$\text{NIQR}_{0.5}$	0.0486 (0.0352)	0.0443 (0.0361)	0.0479 (0.0367)	0.0906 (0.0478)	0.5757
	$\text{NICQR}_9$	0.0316 (0.0217)	0.0285 (0.0223)	0.0305 (0.0244)	0.0581 (0.0304)	2.9971
$\chi^2(3)$	MAVE	0.0361 (0.0276)	0.0374 (0.0286)	0.0347 (0.0300)	0.0697 (0.0391)	0.3695
	$\text{QR}_{0.5}$	0.0448 (0.0355)	0.0475 (0.0363)	0.0440 (0.0344)	0.0874 (0.0481)	3.8438
	$\text{CQR}_9$	0.0294 (0.0249)	0.0344 (0.0166)	0.0295 (0.0203)	0.0605 (0.0207)	52.3415
	NILSE	0.0510 (0.0381)	0.0569 (0.0435)	0.0491 (0.0379)	0.0999 (0.0553)	0.4121
	$\text{NIQR}_{0.5}$	0.0620 (0.0471)	0.0656 (0.0480)	0.0600 (0.0443)	0.1198 (0.0621)	0.7341
	$\text{NICQR}_9$	0.0298 (0.0252)	0.0306 (0.0236)	0.0297 (0.0242)	0.0618 (0.0335)	3.4304

The performance of  $\hat{g}(\cdot)$  is assessed by taking the average squared error:

$$\text{ASE} = \frac{1}{n_{\text{grid}}} \sum_{i=1}^{n_{\text{grid}}} \{\hat{g}(u_i) - g(u_i)\}^2,$$

where  $u_i, i \in \{1, \dots, n_{\text{grid}}\}$  are the grid points of the support of  $\mathbf{X}^\top \gamma_0$ . Here,  $n_{\text{grid}} = 200$  is used. The results in Table 2 suggest the following findings:

- (i) When the error follows  $\mathcal{N}(0, 1)$ , MAVE is the best of the six estimators and the other methods perform nearly as well as MAVE;
- (ii) For a non-normal and symmetric error  $\mathcal{U}[-2, 2]$ ,  $\text{QR}_{0.5}$  and  $\text{NIQR}_9$  perform worse than the other four methods.  $\text{CQR}_9$  and  $\text{NICQR}_9$  perform nearly as well as MAVE and NILSE;
- (iii) For the asymmetric error  $\chi^2(3)$ ,  $\text{NICQR}_9$  performs best.

#### 4.1.2. Simulation Example 2

It is necessary to investigate the effect of heteroscedastic errors. Consider the following model:

$$\mathbf{Y} = \sin(2\mathbf{X}^\top \gamma_0) + \exp\{(\mathbf{X}^\top \gamma_0)^2\} + 0.5 \cos\{2\pi(\mathbf{X}^\top \gamma_0)\} \varepsilon, \quad (15)$$

where the index parameter  $\gamma_0 = (2, 2, 1)^\top/3$ , and  $\mathbf{X}$  is uniformly distributed on  $[0, 1]^3$ , and the residual  $\varepsilon$  follows a  $t$  distribution with three degrees of freedom. Other settings are defined the same as those in Simulation example 1. In this section, we add a method  $\text{WCQR}_9$  proposed by Jiang et al. [8] for comparison. The simulation results are summarized in Table 3. From Table 3, we can see that all the estimators are close to the true value because the absolute bias are very



**Table 2**

This table deals with comparison of some estimation methods (MAVE;  $QR_{0.5}$ ;  $CQR_9$ ; NILSE;  $NIQR_{0.5}$ ;  $NICQR_9$ ) based on the mean values and standard errors (in parenthesis) of average squared error (ASE) based on 500 estimates for the nonparametric function in model (14) under three different error distributions ( $\mathcal{N}(0, 1)$ ;  $\mathcal{N}(-2, 2)$ ;  $\chi^2(3)$ ).

Methods	$\mathcal{N}(0, 1)$	$\mathcal{N}(-2, 2)$	$\chi^2(3)$
MAVE	0.0027 (0.0010)	0.0039 (0.0012)	0.3723 (0.0302)
$QR_{0.5}$	0.0035 (0.0019)	0.0084 (0.0030)	0.2756 (0.0355)
$CQR_9$	0.0033 (0.0016)	0.0046 (0.0013)	0.3297 (0.0272)
NILSE	0.0034 (0.0015)	0.0042 (0.0016)	0.3975 (0.0563)
$NIQR_{0.5}$	0.0036 (0.0023)	0.0091 (0.0038)	0.2872 (0.0456)
$NICQR_9$	0.0037 (0.0027)	0.0049 (0.0020)	0.2725 (0.0327)

**Table 3**

This table deals with comparison of some estimation methods (MAVE;  $QR_{0.5}$ ;  $CQR_9$ ;  $WCQR_9$ ; NILSE;  $NIQR_{0.5}$ ;  $NICQR_9$ ) based on the mean values and standard errors (in parenthesis) of absolute bias (AB), mean squared errors (MSE) and computing time ( $t$ ) based on 500 estimates for the parameter in model (15).

Methods	AB of $\gamma_1$	AB of $\gamma_2$	AB of $\gamma_3$	MSE	$t$
MAVE	0.0109 (0.0098)	0.0122 (0.0088)	0.0155 (0.0114)	0.0251 (0.0134)	0.2390
$QR_{0.5}$	0.0116 (0.0109)	0.0113 (0.0104)	0.0136 (0.0103)	0.0236 (0.0149)	0.8455
$CQR_9$	0.0097 (0.0081)	0.0076 (0.0053)	0.0089 (0.0081)	0.0171 (0.0109)	17.7260
$WCQR_9$	0.0064 (0.0070)	0.0062 (0.0030)	0.0079 (0.0082)	0.0131 (0.0095)	23.2382
NILSE	0.0134 (0.0098)	0.0136 (0.0102)	0.0152 (0.0126)	0.0273 (0.0162)	0.2646
$NIQR_{0.5}$	0.0120 (0.0105)	0.0126 (0.0098)	0.0137 (0.0130)	0.0246 (0.0144)	0.4396
$NICQR_9$	0.0089 (0.0073)	0.0094 (0.0068)	0.0111 (0.0079)	0.0189 (0.0097)	1.4211

**Table 4**

This table presents the single-index coefficient estimates, and deals with comparison of some estimation methods (MAVE;  $QR_{0.5}$ ;  $CQR_9$ ; NILSE;  $NIQR_{0.5}$ ;  $NICQR_9$ ) based on the mean squared errors (MSE) and computing time ( $t$ ) for Boston housing data.

Method	$\hat{\gamma}_1$	$\hat{\gamma}_2$	$\hat{\gamma}_3$	MSE	$t$
MAVE	0.3572	-0.9277	0.1087	21.2104	0.7660
$QR_{0.5}$	0.3001	-0.9513	-0.0688	19.4245	8.2829
$CQR_9$	0.3501	-0.9312	-0.1018	18.6905	162.2150
NILSE	0.3383	-0.9296	-0.1462	21.6905	0.6552
$NIQR_{0.5}$	0.5193	-0.8470	-0.1131	19.7036	1.7464
$NICQR_9$	0.4966	-0.8612	-0.1078	19.3236	8.2151

small. The  $NICQR_9$  estimator performs better than MAVE,  $QR_{0.5}$ , NILSE and  $NIQR_9$  because of smaller MSE. From  $t$ , we see that the operation time of  $NICQR_9$  is faster than that of  $CQR_9$  and  $WCQR_9$ .

#### 4.1.3. Data Example 1: Boston housing data

As an illustration, we now apply the proposed methodology to Boston housing data. These data contain 506 observations on 14 variables, and the dependent variable of interest is *medv*. Thirteen other statistical measurements on the 506 census tracts in suburban Boston from the 1970 census are also included.

Many regression studies have used this dataset and found potential relationships among *medv* and RM, LSTAT and DIS (see Wu et al. [29]). In this study, we focus on the following three covariates:

RM: average number of rooms per dwelling;

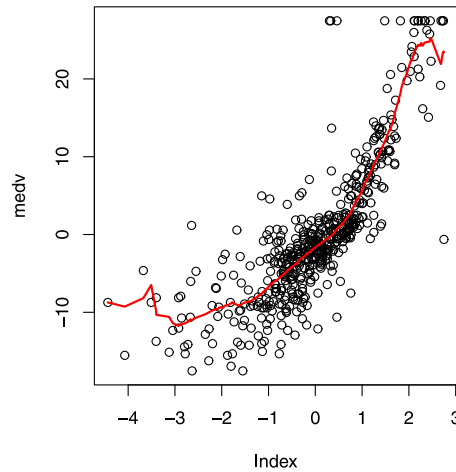
LSTAT: lower status of the population (percent);

DIS: weighted distances to five Boston employment centers.

All three covariates are standardized to have zero mean and unit variance. The dependent variable is centered around zero. In this study, the following single-index model is used to fit the data

$$medv = g\{\gamma_1 RM + \gamma_2 LSTAT + \gamma_3 DIS\} + \varepsilon.$$

The mean squared error (MSE) for fitting is used to assess the relative success of the six estimation methods, where  $MSE = \frac{1}{506} \sum_{i=1}^{506} (y_i - \hat{y}_i)^2$ , and  $\hat{y}_i$  is the fitted value of  $y_i$  (*medv*). Table 4 summarizes the estimated coefficients for the above model, showing that DIS has the smallest effect on house prices among the three covariates and LSTAT is the most important covariate. Table 4 also presents the MSE and  $t$  (computing time) for the estimation method fitting. We find that our methodology ( $NICQR_9$ ) fits the Boston housing dataset well and that the computing time of  $NICQR_9$  is faster than that of  $CQR_9$ . Fig. 2 shows the estimated *medv* along with the observations, where  $Index = \hat{\gamma}_1 RM + \hat{\gamma}_2 LSTAT + \hat{\gamma}_3 DIS$ , illustrating that  $NICQR_9$  is close to the true value.



**Fig. 2.** Estimated single index composite quantile regression for Boston housing data. The dots are the observations  $medv$  and the curve is the estimated  $medv$  by the proposed NICQR<sub>9</sub> method. The abscissa is  $Index = \hat{\gamma}_1 RM + \hat{\gamma}_2 LSTAT + \hat{\gamma}_3 DIS$ .

**Table 5**

This table presents the simulation results, the absolute bias (AB), mean squared errors (MSE) and computing time ( $t_\gamma$ ) based on 100 estimates by the proposed DC-NICQR<sub>9</sub> method for the parameter in model (16), and the absolute error (AE), computing time ( $t_g$ ) based on 100 estimates by the proposed DC-NICQR<sub>9</sub> method for the nonparametric function in model (16), under different sample sizes  $n$  and number of subsets  $M$ .

$n$	$M$	AB of $\gamma_1$	AB of $\gamma_2$	MSE	$t_\gamma$	AE	$t_g$
2,000	1	0.0185 (0.0205)	0.0094 (0.0109)	0.0208 (0.0232)	20.18	0.0262 (0.0181)	3.45
	5	0.0221 (0.0115)	0.0164 (0.0088)	0.0249 (0.0186)	15.98	0.0242 (0.0159)	2.12
	10	0.0300 (0.0140)	0.0154 (0.0076)	0.0337 (0.0159)	13.90	0.0296 (0.0147)	1.90
5,000	1	0.0112 (0.0061)	0.0057 (0.0031)	0.0136 (0.0068)	89.22	0.0126 (0.0146)	27.76
	5	0.0131 (0.0067)	0.0086 (0.0045)	0.0147 (0.0097)	41.31	0.0153 (0.0149)	5.91
	10	0.0121 (0.0062)	0.0076 (0.0040)	0.0139 (0.0086)	34.38	0.0124 (0.0111)	4.40
10,000	1	0.0118 (0.0085)	0.0060 (0.0044)	0.0133 (0.0096)	356.51	0.0123 (0.0137)	159.10
	5	0.0110 (0.0086)	0.0056 (0.0043)	0.0123 (0.0095)	100.23	0.0194 (0.0150)	17.64
	10	0.0130 (0.0077)	0.0071 (0.0040)	0.0145 (0.0087)	77.47	0.0223 (0.0176)	11.14
100,000	10	0.0045 (0.0030)	0.0022 (0.0015)	0.0050 (0.0033)	3977.09	0.0205 (0.0149)	1792.62

## 4.2. Example for massive datasets

In this section, we investigate the performance of our DC-NICQR<sub>9</sub> method compared with the oracle full data NICQR<sub>9</sub> method. We evaluate the methods from two perspectives: (i) accuracy in estimating the oracle full data NICQR<sub>9</sub> statistics and (ii) computational efficiency in terms of run time.

### 4.2.1. Simulation Example 3

We conduct a simulation study and the data are generated from the following model:

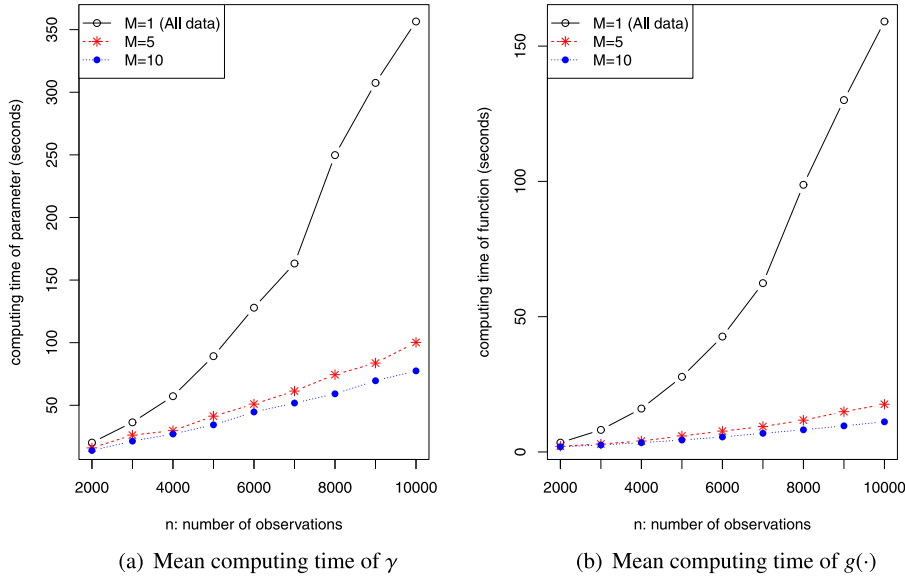
$$y = \cos(\mathbf{X}^\top \boldsymbol{\gamma}_0) + \exp\left\{-\left(\mathbf{X}^\top \boldsymbol{\gamma}_0\right)^2\right\} + 0.2\varepsilon, \quad (16)$$

where  $\boldsymbol{\gamma}_0 = (1, 2)^\top / \sqrt{5}$ , the covariate vector  $\mathbf{X}$  is generated as a multivariate normal with mean zero and covariance matrix  $\text{Var}(\mathbf{X}) = (\sigma_{ij})_{p \times p}$  with  $\sigma_{ij} = 0.5^{|i-j|}$ , and the residual  $\varepsilon$  follows a  $t$  distribution with three degrees of freedom. All the simulations are run for 100 replicates.

Table 5 shows the simulation results. Table 5 presents the AB, MSE, computing time ( $t_\gamma$ ) of the estimate  $\hat{\gamma}$ , the absolute error (AE =  $|g(\cdot) - \hat{g}(\cdot)|$ ) and computing time ( $t_g$ ) of the estimate  $\hat{g}(\cdot)$  for one point of the support of  $\mathbf{X}^\top \boldsymbol{\gamma}_0$ . According to Table 5, the performance of the DC-NICQR<sub>9</sub> estimators is similar to that of the estimators analyzing the full dataset ( $M = 1$ ). Further, the computing time of  $\gamma$  and  $g(\cdot)$  decreases drastically as  $M$  increases, as seen in Table 5 and Fig. 3. With  $n = 100,000$ , we cannot perform the DC-NICQR<sub>9</sub> method on the full dataset because of computer memory limitations. (The same situation also occurred in Chen and Xie [2], who considered the CQR<sub>9</sub> method for the linear model). However, we can obtain estimators by using the divide-and-conquer procedure with  $M = 10$  (see Table 5).

### 4.2.2. Data Example 2: Airline on-time data

Here, airline on-time performance data from the 2009 ASA Data Expo (<http://stat-computing.org/dataexpo/2009/the-data.html>) are used as a case study. These data are publicly available and were used as a demonstration of massive datasets



**Fig. 3.** The average computing time (in seconds) of estimated  $\gamma_0$  and  $g_0(\cdot)$  in model (16) based on 100 estimates by the proposed DC-NICQR<sub>9</sub> method under different sample sizes  $n$  and number of subsets  $M$ .

**Table 6**

This table presents the coefficient estimates, and deals with comparison of two estimation methods (LS; DC-NICQR<sub>9</sub>) based on the mean squared errors (MSE) for the airline on-time data.

Method	HD	DIS	NF	WF	MSE
LS	0.0044	-0.0505	-0.0004	-0.0451	0.0836
DC-NICQR <sub>9</sub>	0.0608	-0.8030	0.0377	-0.5917	0.0831

by Schifano et al. [24]. This dataset consists of flight arrival and departure details for all commercial flights within the United States from October 1987 to April 2008. About 12 million flights were recorded with 29 variables. Because of the computing limit, we only consider the 2008 data (the number of samples is 1,011,963). The first 1,000,000 data points are used for the estimation and the remaining 11,963 data are used for the prediction.

Schifano et al. [24] developed a linear model that fit the data as follows:

$$AD = \gamma_1 HD + \gamma_2 DIS + \gamma_3 NF + \gamma_4 WF + \varepsilon, \quad (17)$$

where  $AD$  is the arrival delay ( $ArrDelay$ ), which is a continuous variable found by modeling  $\log(ArrDelay - \min(ArrDelay) + 1)$ ,  $HD$  is the departure hour (range 0 to 24),  $DIS$  is the distance (in 1000 miles),  $NF$  is the dummy variable for a night flight (1 if departure between 8 p.m. and 5 a.m., 0 otherwise), and  $WF$  is the dummy variable for a weekend flight (1 if departure occurred during the weekend, 0 otherwise).

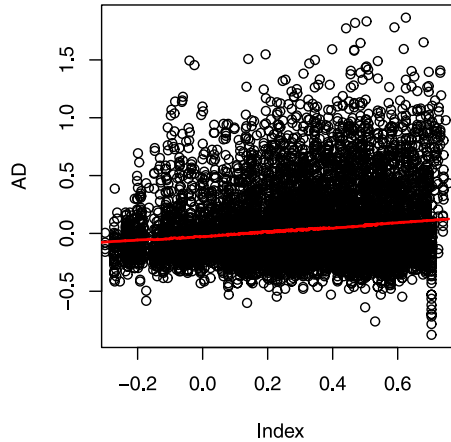
In this study, the following single-index model is used to fit the data:

$$AD = g\{\gamma_1 HD + \gamma_2 DIS + \gamma_3 NF + \gamma_4 WF\} + \varepsilon. \quad (18)$$

For comparison purposes, we use the least squares method proposed by Jiang et al. [6] to estimate  $(\gamma_1, \gamma_2, \gamma_3, \gamma_4)^T$  in model (17), and use the DC-NICQR<sub>9</sub> method proposed in Section 3 to estimate  $(\gamma_1, \gamma_2, \gamma_3, \gamma_4)^T$  in model (18). The number of blocks is 500 for these two methods. Furthermore, we evaluate the performance of these estimators based on their out-of-sample prediction. We present the MSE of the predictions. Table 6 presents the estimated coefficients and MSEs of the two methods. Fig. 4 illustrates the estimated  $AD$  along with the data, showing that the DC-NICQR<sub>9</sub> method performs well with smaller MSE.

## 5. Conclusions

The proposed NICQR method is a non-iterative estimation algorithm that allows us to analyze massive datasets more quickly. Specifically, we used the divide-and-conquer algorithm for massive datasets, which divides the dataset into several blocks, each within the computer's memory. For each block, we applied the NICQR method and constructed the asymptotic confidence density function. We obtained the final estimator by the maximization of the combination of the asymptotic confidence density functions.



**Fig. 4.** Estimated AD in the airline on-time data by model (18). The dots are the observations AD and the curve is the estimated AD by the proposed DC-NICQR<sub>9</sub> method. The abscissa is  $Index = \hat{\gamma}_1 HD + \hat{\gamma}_2 DIS + \hat{\gamma}_3 NF + \hat{\gamma}_4 WF$ .

The methods in this study were designed for the analysis of massive datasets. An extension to the case in which dimension  $p$  and sample size  $n$  are both extremely large is an interesting consideration (see Chen et al. [1]).

## 6. Proofs of theorems

To establish the asymptotic properties of the proposed estimators, the following technical conditions are imposed.

- C1. The kernel  $\tilde{K}(\cdot)$  is a symmetric density function with a bounded support, satisfying a Lipschitz condition.
- C2. The density function of  $\mathbf{U} = \mathbf{X}^\top \boldsymbol{\gamma}$  is positive and uniformly continuous for  $\boldsymbol{\gamma}$  in a neighborhood of  $\boldsymbol{\gamma}_0$ . Further the density of  $\mathbf{X}^\top \boldsymbol{\gamma}_0$  is continuous and bounded away from 0 and  $\infty$  on its support.
- C3. The function  $g_0(\cdot)$  has a continuous and bounded second derivative.
- C4. The model error  $\varepsilon$  has a positive density  $f(\cdot)$ .

**Remark 10.** Conditions C1–C4 are standard conditions, which are commonly used in single-index regression model, see Christou and Akritas [3] and Jiang et al. [7].

**Lemma 1.** Let  $(X_1, Y_1), \dots, (X_n, Y_n)$  be independent and identically distributed random vectors, where  $Y$  is scalar random variable. Further assume that  $E|Y|^s < \infty$ , and  $\sup_x \int |y|^s f(x, y) dy < \infty$ , where  $f(\cdot, \cdot)$  denotes the joint density of  $(X, Y)$ . Let  $\tilde{K}$  be a bounded positive function with a bounded support, satisfying a Lipschitz condition. Given that  $n^{2\varepsilon-1}h \rightarrow \infty$  for some  $\varepsilon < 1 - s^{-1}$ , then

$$\sup_x \left| \frac{1}{n} \sum_{i=1}^n [\tilde{K}_h(X_i - x)Y_i - E\{\tilde{K}_h(X_i - x)Y_i\}] \right| = O_p \left( \left[ \frac{\log(1/h)}{nh} \right]^{1/2} \right).$$

**Proof.** This follows immediately from the result obtained by Mack and Silverman [21].

**Lemma 2.** Suppose that conditions C1–C2 hold, and  $nh_{\max}^4 \rightarrow 0$ , then for any  $\boldsymbol{\gamma}$

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n \left\{ \hat{H}_{\tau_k}(\mathbf{X}_i^\top \boldsymbol{\gamma} \mid \boldsymbol{\gamma}) - H_{\tau_k}(\mathbf{X}_i^\top \boldsymbol{\gamma} \mid \boldsymbol{\gamma}) \right\} = o_p(1).$$

**Proof.** Write

$$\begin{aligned} & \frac{1}{\sqrt{n}} \sum_{i=1}^n \left\{ \hat{H}_{\tau_k}(\mathbf{X}_i^\top \boldsymbol{\gamma} \mid \boldsymbol{\gamma}) - H_{\tau_k}(\mathbf{X}_i^\top \boldsymbol{\gamma} \mid \boldsymbol{\gamma}) \right\} \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \sum_{j=1}^n \frac{\hat{Q}_{\tau_k}(\mathbf{Y}|\mathbf{X}_j) \tilde{K}_{h_k} \{(\mathbf{X}_i - \mathbf{X}_j)^\top \boldsymbol{\gamma}\}}{\sum_{j=1}^n \tilde{K}_{h_k} \{(\mathbf{X}_i - \mathbf{X}_j)^\top \boldsymbol{\gamma}\}} - \frac{1}{\sqrt{n}} \sum_{i=1}^n H_{\tau_k}(\mathbf{X}_i^\top \boldsymbol{\gamma} \mid \boldsymbol{\gamma}) \\ &= \frac{1}{n^{3/2}h_k} \sum_{i=1}^n \sum_{j=1}^n \frac{\hat{Q}_{\tau_k}(\mathbf{Y}|\mathbf{X}_j) \tilde{K}_{h_k} \{(\mathbf{X}_i - \mathbf{X}_j)^\top \boldsymbol{\gamma}\}}{\hat{f}_{\boldsymbol{\gamma}}(\mathbf{X}_i^\top \boldsymbol{\gamma})} - \frac{1}{\sqrt{n}} \sum_{i=1}^n H_{\tau_k}(\mathbf{X}_i^\top \boldsymbol{\gamma} \mid \boldsymbol{\gamma}) \end{aligned}$$

$$\begin{aligned}
&= \frac{1}{n^{3/2}h_k} \sum_{i=1}^n \sum_{j=1}^n \frac{\hat{Q}_{\tau_k}(\mathbf{Y}|\mathbf{X}_j) \tilde{K}_{h_k} \{(\mathbf{X}_i - \mathbf{X}_j)^\top \boldsymbol{\gamma}\}}{\hat{f}_{\boldsymbol{\gamma}}(\mathbf{X}_i^\top \boldsymbol{\gamma})} - \frac{1}{n^{3/2}h_k} \sum_{i=1}^n \sum_{j=1}^n \frac{Q_{\tau_k}(\mathbf{Y}|\mathbf{X}_j) \tilde{K}_{h_k} \{(\mathbf{X}_i - \mathbf{X}_j)^\top \boldsymbol{\gamma}\}}{f_{\boldsymbol{\gamma}}(\mathbf{X}_i^\top \boldsymbol{\gamma})} \\
&\quad + \left[ \frac{1}{n^{3/2}h_k} \sum_{i=1}^n \sum_{j=1}^n \frac{Q_{\tau_k}(\mathbf{Y}|\mathbf{X}_j) \tilde{K}_{h_k} \{(\mathbf{X}_i - \mathbf{X}_j)^\top \boldsymbol{\gamma}\}}{f_{\boldsymbol{\gamma}}(\mathbf{X}_i^\top \boldsymbol{\gamma})} - \frac{1}{\sqrt{n}} \sum_{i=1}^n H_{\tau_k}(\mathbf{X}_i^\top \boldsymbol{\gamma} | \boldsymbol{\gamma}) \right] \equiv T_1 + T_2,
\end{aligned}$$

where  $\hat{f}_{\boldsymbol{\gamma}}(t) = (nh_k)^{-1} \sum_{i=1}^n \tilde{K}_{h_k}(\mathbf{X}_i^\top \boldsymbol{\gamma} - t)$  and

$$\begin{aligned}
T_1 &= \frac{1}{n^{3/2}h_k} \sum_{i=1}^n \sum_{j=1}^n \frac{\hat{Q}_{\tau_k}(\mathbf{Y}|\mathbf{X}_j) \tilde{K}_{h_k} \{(\mathbf{X}_i - \mathbf{X}_j)^\top \boldsymbol{\gamma}\}}{\hat{f}_{\boldsymbol{\gamma}}(\mathbf{X}_i^\top \boldsymbol{\gamma})} - \frac{1}{n^{3/2}h_k} \sum_{i=1}^n \sum_{j=1}^n \frac{Q_{\tau_k}(\mathbf{Y}|\mathbf{X}_j) \tilde{K}_{h_k} \{(\mathbf{X}_i - \mathbf{X}_j)^\top \boldsymbol{\gamma}\}}{f_{\boldsymbol{\gamma}}(\mathbf{X}_i^\top \boldsymbol{\gamma})}, \\
T_2 &= \frac{1}{n^{3/2}h_k} \sum_{i=1}^n \sum_{j=1}^n \frac{Q_{\tau_k}(\mathbf{Y}|\mathbf{X}_j) \tilde{K}_{h_k} \{(\mathbf{X}_i - \mathbf{X}_j)^\top \boldsymbol{\gamma}\}}{f_{\boldsymbol{\gamma}}(\mathbf{X}_i^\top \boldsymbol{\gamma})} - \frac{1}{\sqrt{n}} \sum_{i=1}^n H_{\tau_k}(\mathbf{X}_i^\top \boldsymbol{\gamma} | \boldsymbol{\gamma}).
\end{aligned}$$

In the first step it will be shown that  $T_1 = o_p(1)$ ,

$$\begin{aligned}
T_1 &= \frac{1}{n^{3/2}h_k} \sum_{i=1}^n \sum_{j=1}^n \frac{\hat{Q}_{\tau_k}(\mathbf{Y}|\mathbf{X}_j) \tilde{K}_{h_k} \{(\mathbf{X}_i - \mathbf{X}_j)^\top \boldsymbol{\gamma}\}}{\hat{f}_{\boldsymbol{\gamma}}(\mathbf{X}_i^\top \boldsymbol{\gamma})} - \frac{1}{n^{3/2}h_k} \sum_{i=1}^n \sum_{j=1}^n \frac{Q_{\tau_k}(\mathbf{Y}|\mathbf{X}_j) \tilde{K}_{h_k} \{(\mathbf{X}_i - \mathbf{X}_j)^\top \boldsymbol{\gamma}\}}{\hat{f}_{\boldsymbol{\gamma}}(\mathbf{X}_i^\top \boldsymbol{\gamma})} \\
&\quad + \frac{1}{n^{3/2}h_k} \sum_{i=1}^n \sum_{j=1}^n \frac{Q_{\tau_k}(\mathbf{Y}|\mathbf{X}_j) \tilde{K}_{h_k} \{(\mathbf{X}_i - \mathbf{X}_j)^\top \boldsymbol{\gamma}\}}{\hat{f}_{\boldsymbol{\gamma}}(\mathbf{X}_i^\top \boldsymbol{\gamma})} - \frac{1}{n^{3/2}h_k} \sum_{i=1}^n \sum_{j=1}^n \frac{Q_{\tau_k}(\mathbf{Y}|\mathbf{X}_j) \tilde{K}_{h_k} \{(\mathbf{X}_i - \mathbf{X}_j)^\top \boldsymbol{\gamma}\}}{f_{\boldsymbol{\gamma}}(\mathbf{X}_i^\top \boldsymbol{\gamma})} \\
&= \frac{1}{n^{3/2}h_k} \sum_{i=1}^n \sum_{j=1}^n \left\{ \hat{Q}_{\tau_k}(\mathbf{Y}|\mathbf{X}_j) - Q_{\tau_k}(\mathbf{Y}|\mathbf{X}_j) \right\} \frac{\tilde{K}_{h_k} \{(\mathbf{X}_i - \mathbf{X}_j)^\top \boldsymbol{\gamma}\}}{\hat{f}_{\boldsymbol{\gamma}}(\mathbf{X}_i^\top \boldsymbol{\gamma})} \\
&\quad + \frac{1}{n^{3/2}h_k} \sum_{i=1}^n \sum_{j=1}^n Q_{\tau_k}(\mathbf{Y}|\mathbf{X}_j) \tilde{K}_{h_k} \{(\mathbf{X}_i - \mathbf{X}_j)^\top \boldsymbol{\gamma}\} \left\{ \frac{1}{\hat{f}_{\boldsymbol{\gamma}}(\mathbf{X}_i^\top \boldsymbol{\gamma})} - \frac{1}{f_{\boldsymbol{\gamma}}(\mathbf{X}_i^\top \boldsymbol{\gamma})} \right\} \equiv T_{11} + T_{12},
\end{aligned}$$

where

$$\begin{aligned}
T_{11} &= \frac{1}{n^{3/2}h_k} \sum_{i=1}^n \sum_{j=1}^n \left\{ \hat{Q}_{\tau_k}(\mathbf{Y}|\mathbf{X}_j) - Q_{\tau_k}(\mathbf{Y}|\mathbf{X}_j) \right\} \frac{\tilde{K}_{h_k} \{(\mathbf{X}_i - \mathbf{X}_j)^\top \boldsymbol{\gamma}\}}{\hat{f}_{\boldsymbol{\gamma}}(\mathbf{X}_i^\top \boldsymbol{\gamma})}, \\
T_{12} &= \frac{1}{n^{3/2}h_k} \sum_{i=1}^n \sum_{j=1}^n Q_{\tau_k}(\mathbf{Y}|\mathbf{X}_j) \tilde{K}_{h_k} \{(\mathbf{X}_i - \mathbf{X}_j)^\top \boldsymbol{\gamma}\} \left\{ \frac{1}{\hat{f}_{\boldsymbol{\gamma}}(\mathbf{X}_i^\top \boldsymbol{\gamma})} - \frac{1}{f_{\boldsymbol{\gamma}}(\mathbf{X}_i^\top \boldsymbol{\gamma})} \right\}.
\end{aligned}$$

The  $T_{11} = o_p(1)$  because of  $\sup_{\mathbf{x} \in \mathbf{X}} |\hat{Q}_{\tau_k}(\mathbf{Y}|\mathbf{x}) - Q_{\tau_k}(\mathbf{Y}|\mathbf{x})| = O_p(n^{-1/2})$ , see Rémillard et al. [23]. Under the condition C2 and  $nh_{\max}^4 = o(1)$ ,  $T_{12} = o_p(1)$ . Thus,  $T_1 = o_p(1)$ . Next, to show that  $T_2 = o_p(1)$ .

$$\begin{aligned}
T_2 &= \frac{1}{n^{3/2}h_k} \sum_{i=1}^n \sum_{j=1}^n \frac{Q_{\tau_k}(\mathbf{Y}|\mathbf{X}_j) \tilde{K}_{h_k} \{(\mathbf{X}_i - \mathbf{X}_j)^\top \boldsymbol{\gamma}\}}{f_{\boldsymbol{\gamma}}(\mathbf{X}_i^\top \boldsymbol{\gamma})} - \frac{1}{n^{3/2}h_k} \sum_{i=1}^n \sum_{j=1}^n \frac{H_{\tau_k}(\mathbf{X}_i^\top \boldsymbol{\gamma} | \boldsymbol{\gamma}) \tilde{K}_{h_k} \{(\mathbf{X}_i - \mathbf{X}_j)^\top \boldsymbol{\gamma}\}}{f_{\boldsymbol{\gamma}}(\mathbf{X}_i^\top \boldsymbol{\gamma})} \\
&\quad + \frac{1}{n^{3/2}h_k} \sum_{i=1}^n \sum_{j=1}^n \frac{H_{\tau_k}(\mathbf{X}_i^\top \boldsymbol{\gamma} | \boldsymbol{\gamma}) \tilde{K}_{h_k} \{(\mathbf{X}_i - \mathbf{X}_j)^\top \boldsymbol{\gamma}\}}{f_{\boldsymbol{\gamma}}(\mathbf{X}_i^\top \boldsymbol{\gamma})} - \frac{1}{\sqrt{n}} \sum_{i=1}^n H_{\tau_k}(\mathbf{X}_i^\top \boldsymbol{\gamma} | \boldsymbol{\gamma}) \\
&= \frac{1}{n^{3/2}h_k} \sum_{i=1}^n \sum_{j=1}^n \left\{ Q_{\tau_k}(\mathbf{Y}|\mathbf{X}_j) - H_{\tau_k}(\mathbf{X}_i^\top \boldsymbol{\gamma} | \boldsymbol{\gamma}) \right\} \frac{\tilde{K}_{h_k} \{(\mathbf{X}_i - \mathbf{X}_j)^\top \boldsymbol{\gamma}\}}{f_{\boldsymbol{\gamma}}(\mathbf{X}_i^\top \boldsymbol{\gamma})} \\
&\quad + \frac{1}{n^{3/2}h_k} \sum_{i=1}^n \sum_{j=1}^n H_{\tau_k}(\mathbf{X}_i^\top \boldsymbol{\gamma} | \boldsymbol{\gamma}) \tilde{K}_{h_k} \{(\mathbf{X}_i - \mathbf{X}_j)^\top \boldsymbol{\gamma}\} \left\{ \frac{1}{f_{\boldsymbol{\gamma}}(\mathbf{X}_i^\top \boldsymbol{\gamma})} - \frac{1}{\hat{f}_{\boldsymbol{\gamma}}(\mathbf{X}_i^\top \boldsymbol{\gamma})} \right\} \\
&= \frac{1}{n^{3/2}h_k} \sum_{i=1}^n \sum_{j=1}^n \left\{ Q_{\tau_k}(\mathbf{Y}|\mathbf{X}_j) - H_{\tau_k}(\mathbf{X}_i^\top \boldsymbol{\gamma} | \boldsymbol{\gamma}) \right\} \frac{\tilde{K}_{h_k} \{(\mathbf{X}_i - \mathbf{X}_j)^\top \boldsymbol{\gamma}\}}{f_{\boldsymbol{\gamma}}(\mathbf{X}_i^\top \boldsymbol{\gamma})} + o_p(1).
\end{aligned}$$

By the U-statistics techniques and condition  $nh_{\max}^4 = o(1)$ , we can prove that  $T_2 = o_p(1)$ .

**Proof of Theorem 1.** Set  $\hat{\boldsymbol{y}}^* = \sqrt{n}(\hat{\boldsymbol{y}} - \boldsymbol{y}_0)$  and  $\boldsymbol{y}^* = \sqrt{n}(\boldsymbol{y} - \boldsymbol{y}_0)$ . Then,  $\hat{\boldsymbol{y}}^*$  is also the minimizer of

$$L_n(\boldsymbol{y}^*) = \sum_{i=1}^n \sum_{k=1}^K \left\{ \rho_{\tau_k} \left( Y_{i,\tau_k}^* - \tilde{H}_{\tau_k}(\mathbf{X}_i \mid \boldsymbol{y}^*/\sqrt{n} + \boldsymbol{y}_0) \right) - \rho_{\tau_k}(Y_{i,\tau_k}^*) \right\},$$

where  $Y_{i,\tau_k}^* = Y_i - \hat{H}_{\tau_k}(\mathbf{X}_i^\top \boldsymbol{y}_0 \mid \boldsymbol{y}_0) + O_p(n^{-1})$ , and  $\tilde{H}_{\tau_k}(\mathbf{X}_i \mid \boldsymbol{y}^*/\sqrt{n} + \boldsymbol{y}_0) = \hat{H}_{\tau_k}(\mathbf{X}_i^\top \boldsymbol{y} \mid \boldsymbol{y}) - \hat{H}_{\tau_k}(\mathbf{X}_i^\top \boldsymbol{y}_0 \mid \boldsymbol{y}_0)$ . Write  $L_n(\boldsymbol{y}^*)$  as

$$L_n(\boldsymbol{y}^*) = E \{ L_n(\boldsymbol{y}^*) \mid \mathbf{X} \} - \sum_{i=1}^n \sum_{k=1}^K [\rho'_{\tau_k}(Y_{i,\tau_k}^*) - E \{ \rho'_{\tau_k}(Y_{i,\tau_k}^*) \mid \mathbf{X} \}] \tilde{H}_{\tau_k}(\mathbf{X}_i \mid \boldsymbol{y}^*/\sqrt{n} + \boldsymbol{y}_0) + R_n(\boldsymbol{y}^*),$$

where  $R_n(\boldsymbol{y}^*)$  is the remainder term, and to save space, we obtain  $R_n(\boldsymbol{y}^*) = o_p(1)$  by similar idea of proof in Fan et al. [4]. Note that

$$\begin{aligned} E \{ L_n(\boldsymbol{y}^*) \mid \mathbf{X} \} &= \sum_{i=1}^n \sum_{k=1}^K E \left\{ \rho_{\tau_k} \left( Y_{i,\tau_k}^* - \tilde{H}_{\tau_k}(\mathbf{X}_i \mid \boldsymbol{y}^*/\sqrt{n} + \boldsymbol{y}_0) \right) - \rho_{\tau_k}(Y_{i,\tau_k}^*) \mid \mathbf{X} \right\} \\ &= - \sum_{i=1}^n \sum_{k=1}^K E \{ \rho'_{\tau_k}(Y_{i,\tau_k}^*) \mid \mathbf{X} \} \tilde{H}_{\tau_k}(\mathbf{X}_i \mid \boldsymbol{y}^*/\sqrt{n} + \boldsymbol{y}_0) \\ &\quad + \frac{1}{2} \sum_{i=1}^n \sum_{k=1}^K E \{ \rho''_{\tau_k}(Y_{i,\tau_k}^*) \mid \mathbf{X} \} \tilde{H}_{\tau_k}^2(\mathbf{X}_i \mid \boldsymbol{y}^*/\sqrt{n} + \boldsymbol{y}_0) + o_p(1), \end{aligned}$$

and  $E \{ \rho''_{\tau_k}(Y_{i,\tau_k}^*) \mid \mathbf{X} \} = f_\varepsilon(c_k) + o_p(n^{-1/2})$ . Thus, we can obtain

$$\begin{aligned} L_n(\boldsymbol{y}^*) &= - \sum_{k=1}^K \sum_{i=1}^n \rho'_{\tau_k}(Y_{i,\tau_k}^*) \tilde{H}_{\tau_k}(\mathbf{X}_i \mid \boldsymbol{y}^*/\sqrt{n} + \boldsymbol{y}_0) + \frac{1}{2} \sum_{k=1}^K f_\varepsilon(c_k) \sum_{i=1}^n \tilde{H}_{\tau_k}^2(\mathbf{X}_i \mid \boldsymbol{y}^*/\sqrt{n} + \boldsymbol{y}_0) + o_p(1) \\ &= - \sum_{k=1}^K \sum_{i=1}^n \rho'_{\tau_k}(Y_{i,\tau_k}^*) \left\{ \hat{H}_{\tau_k}(\mathbf{X}_i^\top (\boldsymbol{y}^*/\sqrt{n} + \boldsymbol{y}_0) \mid \boldsymbol{y}^*/\sqrt{n} + \boldsymbol{y}_0) - \hat{H}_{\tau_k}(\mathbf{X}_i^\top \boldsymbol{y}_0 \mid \boldsymbol{y}_0) \right\} \\ &\quad + \frac{1}{2} \sum_{k=1}^K f_\varepsilon(c_k) \sum_{i=1}^n \left\{ \hat{H}_{\tau_k}(\mathbf{X}_i^\top (\boldsymbol{y}^*/\sqrt{n} + \boldsymbol{y}_0) \mid \boldsymbol{y}^*/\sqrt{n} + \boldsymbol{y}_0) - \hat{H}_{\tau_k}(\mathbf{X}_i^\top \boldsymbol{y}_0 \mid \boldsymbol{y}_0) \right\}^2 + o_p(1). \end{aligned}$$

By Lemma 2, we can obtain

$$\begin{aligned} &\sum_{i=1}^n \left\{ \hat{H}_{\tau_k}(\mathbf{X}_i^\top (\boldsymbol{y}^*/\sqrt{n} + \boldsymbol{y}_0) \mid \boldsymbol{y}^*/\sqrt{n} + \boldsymbol{y}_0) - \hat{H}_{\tau_k}(\mathbf{X}_i^\top \boldsymbol{y}_0 \mid \boldsymbol{y}_0) \right\} \\ &= \sum_{i=1}^n \left\{ H_{\tau_k}(\mathbf{X}_i^\top (\boldsymbol{y}^*/\sqrt{n} + \boldsymbol{y}_0) \mid \boldsymbol{y}^*/\sqrt{n} + \boldsymbol{y}_0) - H_{\tau_k}(\mathbf{X}_i^\top \boldsymbol{y}_0 \mid \boldsymbol{y}_0) \right\} + o_p(n^{-1/2}), \end{aligned}$$

and

$$\begin{aligned} &H_{\tau_k}(\mathbf{X}_i^\top (\boldsymbol{y}^*/\sqrt{n} + \boldsymbol{y}_0) \mid \boldsymbol{y}^*/\sqrt{n} + \boldsymbol{y}_0) - H_{\tau_k}(\mathbf{X}_i^\top \boldsymbol{y}_0 \mid \boldsymbol{y}_0) \\ &= \frac{\boldsymbol{y}^*}{\sqrt{n}} \frac{\partial H_{\tau_k}(\mathbf{X}_i^\top \boldsymbol{y} \mid \boldsymbol{y})}{\partial \boldsymbol{y}} \bigg|_{\boldsymbol{y}_0} + O_p(n^{-1}) = \frac{1}{\sqrt{n}} g'(\mathbf{X}_i^\top \boldsymbol{y}_0) (\mathbf{X}_i - E[\mathbf{X} \mid \mathbf{X}^\top \boldsymbol{y}_0])^\top \boldsymbol{y}^* + O_p(n^{-1}). \end{aligned}$$

Thus,

$$L_n(\boldsymbol{y}^*) = -W_n \boldsymbol{y}^* + \frac{1}{2} \{\boldsymbol{y}^*\}^\top \left\{ \sum_{k=1}^K f_\varepsilon(c_k) \right\} S_n \boldsymbol{y}^* + o_p(1),$$

where

$$\begin{aligned} W_n &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \sum_{k=1}^K \rho'_{\tau_k}(Y_{i,\tau_k}^*) g'(\mathbf{X}_i^\top \boldsymbol{y}_0) \{ \mathbf{X}_i - E[\mathbf{X} \mid \mathbf{X}^\top \boldsymbol{y}_0] \}^\top, \\ S_n &= \frac{1}{n} \sum_{i=1}^n \{ g'(\mathbf{X}_i^\top \boldsymbol{y}_0) \}^2 \{ \mathbf{X}_i - E[\mathbf{X} \mid \mathbf{X}^\top \boldsymbol{y}_0] \} \{ \mathbf{X}_i - E[\mathbf{X} \mid \mathbf{X}^\top \boldsymbol{y}_0] \}^\top. \end{aligned}$$

It is easy to show that  $S_n = \mathbf{S} + o_p(1)$ , thus

$$L_n(\boldsymbol{\gamma}^*) = -W_n \boldsymbol{\gamma}^* + \frac{1}{2} \{\boldsymbol{\gamma}^*\}^\top \left\{ \sum_{k=1}^K f_\varepsilon(c_k) \right\} \mathbf{S} \boldsymbol{\gamma}^* + o_p(1).$$

It follows by a convexity lemma (see Pollard [22]) that the quadratic approximation to  $L_n(\boldsymbol{\gamma}^*)$  holds uniformly for  $\boldsymbol{\gamma}^*$  in any compact set. Thus, it follows that

$$\hat{\boldsymbol{\gamma}}^* = - \left\{ \sum_{k=1}^K f_\varepsilon(c_k) \right\}^{-1} \mathbf{S}^\top W_n + o_p(1).$$

By the Cramér–Wald theorem and the Central Limit Theorem for  $W_n$  holds, then  $\text{Var}(W_n) \rightarrow \sum_{k=1}^K \sum_{k'=1}^K \tau_{kk'} \mathbf{S}$ . This completes the proof.

**Proof of Theorem 2.** Note that

$$\sqrt{nh} \{ \hat{g}(u | \hat{\gamma}) - g_0(u) \} = \sqrt{nh} \{ \hat{g}(u | \hat{\gamma}) - \hat{g}(u | \gamma_0) \} + \sqrt{nh} \{ \hat{g}(u | \gamma_0) - g_0(u) \},$$

where  $\hat{g}(u | \gamma_0)$  is a local linear estimator of  $g_0(u)$  when the index coefficient  $\gamma_0$  is known. For given  $u$ , for notational simplicity, we write  $\sum_{k=1}^K v_k \hat{a}_{k,\hat{\gamma}} = \hat{g}(u | \hat{\gamma})$ ,  $\hat{b}_{\hat{\gamma}} = \hat{g}'(u | \hat{\gamma})$ ,  $\sum_{k=1}^K v_k \hat{a}_{k,\gamma_0} = \hat{g}(u | \gamma_0)$ , and  $\hat{b}_{\gamma_0} = \hat{g}'(u | \gamma_0)$ , which are the solutions of the following minimization problems, respectively,

$$\begin{aligned} \min_{(a_1, \dots, a_K, b)} \sum_{k=1}^K \sum_{i=1}^n \rho_{\tau_k} \{ Y_i - a_k - b(X_i^\top \hat{\gamma} - u) \} \tilde{K}_h(X_i^\top \hat{\gamma} - u), \\ \min_{(a_1, \dots, a_K, b)} \sum_{k=1}^K \sum_{i=1}^n \rho_{\tau_k} \{ Y_i - a_k - b(X_i^\top \gamma_0 - u) \} \tilde{K}_h(X_i^\top \gamma_0 - u). \end{aligned}$$

Denote

$$\begin{aligned} \theta^* &= \sqrt{nh} [a_{1,\hat{\gamma}} - g_0(u) - c_1, \dots, a_{K,\hat{\gamma}} - g_0(u) - c_K, h\{\hat{b}_{\hat{\gamma}} - g'_0(u)\}]^\top, \\ \bar{\theta}^* &= \sqrt{nh} [\hat{a}_{1,\hat{\gamma}} - g_0(u) - c_1, \dots, \hat{a}_{K,\hat{\gamma}} - g_0(u) - c_K, h\{\hat{b}_{\hat{\gamma}} - g'_0(u)\}]^\top, \\ \bar{\theta}^{**} &= \sqrt{nh} [\hat{a}_{1,\gamma_0} - g_0(u) - c_1, \dots, \hat{a}_{K,\gamma_0} - g_0(u) - c_K, h\{\hat{b}_{\gamma_0} - g'_0(u)\}]^\top, \\ Z_{i,k}^* &= \{e_k^\top, (X_i^\top \hat{\gamma} - u)/h\}^\top, Z_{i,k}^{**} = \{e_k^\top, (X_i^\top \gamma_0 - u)/h\}^\top, \end{aligned}$$

where  $e_k$  is a  $K$ -vector with 1 on the  $k$ th position and 0 elsewhere. Further, write  $K_i^* = \tilde{K}_h(X_i^\top \hat{\gamma} - u)$ ,  $K_i^{**} = \tilde{K}_h(X_i^\top \gamma_0 - u)$ ,  $\eta_{i,k} = I(\varepsilon_i \leq c_k) - \tau_k$ ,  $\eta_{i,k}^*(u) = I(\varepsilon_i \leq c_k - d_{i,k}^*) - \tau_k$ ,  $\eta_{i,k}^{**}(u) = I(\varepsilon_i \leq c_k - d_{i,k}^{**}) - \tau_k$ , where  $d_{i,k}^*(u) = c_k + g_0(X_i^\top \hat{\gamma}) - g_0(u) - g'_0(u)(X_i^\top \hat{\gamma} - u)$  and  $d_{i,k}^{**}(u) = c_k + g_0(X_i^\top \gamma_0) - g_0(u) - g'_0(u)(X_i^\top \gamma_0 - u)$ . Thus,  $Y_i - a_k - b(X_i^\top \hat{\gamma} - u) = \varepsilon_i - c_k + d_{i,k}^* - \Delta_{i,k}$ , where  $\Delta_{i,k} = \{Z_{i,k}^*\}^\top \theta^* / \sqrt{nh}$ . Then,  $\bar{\theta}^*$  is also the minimizer of

$$L_n^*(\theta^*) = \sum_{k=1}^K \sum_{i=1}^n \{ \rho_{\tau_k}(\varepsilon_i - c_k + d_{i,k}^* - \Delta_{i,k}) - \rho_{\tau_k}(\varepsilon_i - c_k + d_{i,k}^*) \} K_i^*.$$

By applying the identity (Knight [10])

$$\rho_\tau(x - y) - \rho_\tau(x) = y \{I(x \leq 0) - \tau\} + \int_0^y \{I(x \leq z) - I(x \leq 0)\} dz,$$

we have

$$\begin{aligned} L_n^*(\theta^*) &= \sum_{k=1}^K \sum_{i=1}^n K_i^* \Delta_{i,k} \{I(\varepsilon_i \leq c_k - d_{i,k}^*) - \tau_k\} + \sum_{k=1}^K \sum_{i=1}^n K_i^* \int_0^{\Delta_{i,k}} \{I(\varepsilon_i \leq c_k - d_{i,k}^* + z) - I(\varepsilon_i \leq c_k - d_{i,k}^*)\} dz \\ &\equiv W_n^{*\top} \theta^* + \sum_{k=1}^K B_{n,k}^*(\theta^*), \end{aligned}$$

where  $W_n^* = (nh)^{-1/2} \sum_{k=1}^K \sum_{i=1}^n \eta_{i,k}^*(u) Z_{i,k}^* K_i^*$  and  $B_{n,k}^*(\theta^*) = \sum_{i=1}^n K_i^* \int_0^{\Delta_{i,k}} \{I(\varepsilon_i \leq c_k - d_{i,k}^* + z) - I(\varepsilon_i \leq c_k - d_{i,k}^*)\} dz$ . Since  $B_{n,k}^*(\theta^*)$  is a summation of i.i.d. random variables of the kernel form, it follows by Lemma 1 that

$$B_{n,k}^*(\theta^*) = E\{B_{n,k}^*(\theta^*)\} + O_p(\log^{1/2}(1/h)/\sqrt{nh}).$$



The conditional expectation of  $\sum_{k=1}^K B_{n,k}^*(\theta^*)$  can be calculated as

$$\begin{aligned} \sum_{k=1}^K E\{B_{n,k}^*(\theta^*)|\mathbf{U}_0\} &= \sum_{k=1}^K \sum_{i=1}^n K_i^* \int_0^{\Delta_{i,k}} \{F(c_k - d_{i,k}^* + z) - F(c_k - d_{i,k}^*)\} dz \\ &= \frac{1}{2} \theta^{*\top} \left\{ \frac{1}{nh} \sum_{k=1}^K \sum_{i=1}^n K_i^* f(c_k - d_{i,k}^*) Z_{i,k}^* Z_{i,k}^{*\top} \right\} \theta^* + O_p(\log^{1/2}(1/h)/\sqrt{nh}) \\ &\equiv \frac{1}{2} \theta^{*\top} S_n^* \theta^* + O_p(\log^{1/2}(1/h)/\sqrt{nh}), \end{aligned}$$

where  $S_n^* = (nh)^{-1} \sum_{k=1}^K \sum_{i=1}^n K_i^* f(c_k - d_{i,k}^*) Z_{i,k}^* Z_{i,k}^{*\top}$ . Then,

$$\begin{aligned} L_n^*(\theta^*) &= W_n^{*\top} \theta^* + \sum_{k=1}^K E\{B_{n,k}^*(\theta^*)\} + O_p(\log^{1/2}(1/h)/\sqrt{nh}) \\ &= W_n^{*\top} \theta^* + \sum_{k=1}^K E[E\{B_{n,k}^*(\theta^*)|\mathbf{U}_0\}] + O_p(\log^{1/2}(1/h)/\sqrt{nh}) \\ &= W_n^{*\top} \theta^* + \frac{1}{2} \theta^{*\top} E(S_n^*) \theta^* + O_p(\log^{1/2}(1/h)/\sqrt{nh}). \end{aligned}$$

It can be shown that  $E(S_n^*) = f_{\mathbf{U}_0}(u) S^* + O(h^2)$ , where

$$S^* = \begin{pmatrix} C^* & 0 \\ 0 & \mu_2 \sum_{k=1}^K f(c_k) \end{pmatrix}.$$

where  $C^*$  is a  $K \times K$  diagonal matrix with  $C_{jj}^* = f(c_j)$ . Therefore, we can write  $L_n^*(\theta^*)$  as

$$L_n^*(\theta^*) = W_n^{*\top} \theta^* + \frac{1}{2} f_{\mathbf{U}_0}(u) \theta^{*\top} S^* \theta^* + O_p(\log^{1/2}(1/h)/\sqrt{nh}).$$

By applying a convexity lemma (Pollard [22]), the minimizer of  $L_n^*(\theta^*)$  can be expressed as

$$\bar{\theta}^* = -\{f_{\mathbf{U}_0}(u) S^*\}^+ W_n^* + o_p(1).$$

The estimator  $\bar{\theta}^{**}$  can similarly be shown to equal

$$\bar{\theta}^{**} = -\{f_{\mathbf{U}_0}(u) S^*\}^+ W_n^{**} + o_p(1),$$

where  $W_n^{**} = (nh)^{-1/2} \sum_{k=1}^K \sum_{i=1}^n \eta_{i,k}^{**}(u) Z_{i,k}^{**} K_i^{**}$ . Thus, by the conditions  $\sum_{k=1}^K v_k = 1$  and  $\sum_{k=1}^K v_k c_k = 0$ , we can obtain

$$\begin{aligned} \sqrt{nh} \{\hat{g}(u | \gamma_0) - g_0(u)\} &= \sqrt{nh} \left\{ \sum_{k=1}^K v_k \hat{a}_{k,\gamma_0} - g_0(u) \right\} = \sum_{k=1}^K v_k \sqrt{nh} \{\hat{a}_{k,\gamma_0} - g_0(u) - c_k\} \\ &= -\frac{f_{\mathbf{U}_0}^{-1}(u)}{\sqrt{nh}} \sum_{k=1}^K \sum_{i=1}^n v_k f^{-1}(c_k) \{I(\varepsilon_i \leq c_k - d_{i,k}^{**}) - \tau_k\} K_i^{**} + o_p(1), \\ \sqrt{nh} \{\hat{g}(u | \hat{\gamma}) - g_0(u)\} &= -\frac{f_{\mathbf{U}_0}^{-1}(u)}{\sqrt{nh}} \sum_{k=1}^K \sum_{i=1}^n v_k f^{-1}(c_k) \{I(\varepsilon_i \leq c_k - d_{i,k}^*) - \tau_k\} K_i^* + o_p(1). \end{aligned}$$

Thus,

$$\begin{aligned} \sqrt{nh} \{\hat{g}(u | \hat{\gamma}) - \hat{g}(u | \gamma_0)\} &= \sqrt{nh} \{\hat{g}(u | \hat{\gamma}) - g_0(u)\} - \sqrt{nh} \{\hat{g}(u | \gamma_0) - g_0(u)\} \\ &= -\frac{f_{\mathbf{U}_0}^{-1}(u)}{\sqrt{nh}} \sum_{k=1}^K \sum_{i=1}^n v_k f^{-1}(c_k) [I(\varepsilon_i \leq c_k - d_{i,k}^*) - \tau_k] K_i^* - \{I(\varepsilon_i \leq c_k - d_{i,k}^{**}) - \tau_k\} K_i^{**} + o_p(1). \end{aligned}$$

When  $\|\hat{\gamma} - \gamma_0\| = O_p(n^{-1/2})$ , we can obtain  $\sqrt{nh} \{\hat{g}(u | \hat{\gamma}) - \hat{g}(u | \gamma_0)\} = o_p(1)$ . For above, it is easy to see that

$$\begin{aligned} \sqrt{nh} \{\hat{g}(u | \hat{\gamma}) - g_0(u)\} &= \sqrt{nh} \{\hat{g}(u | \hat{\gamma}) - \hat{g}(u | \gamma_0)\} + \sqrt{nh} \{\hat{g}(u | \gamma_0) - g_0(u)\} \\ &= -\frac{f_{\mathbf{U}_0}^{-1}(u)}{\sqrt{nh}} \sum_{k=1}^K \sum_{i=1}^n v_k f^{-1}(c_k) \{I(\varepsilon_i \leq c_k - d_{i,k}^{**}) - \tau_k\} K_i^{**} + o_p(1). \end{aligned}$$

Then, by the conditions  $\sum_{k=1}^K v_k = 1$  and  $\sum_{k=1}^K v_k c_k = 0$ , we have

$$\begin{aligned} \text{bias}\{\hat{g}(u | \hat{\gamma}) | \mathbf{U}_0\} &= E\{\hat{g}(u | \hat{\gamma}) - g_0(u) | \mathbf{U}_0\} \\ &= -\frac{f_{\mathbf{U}_0}^{-1}(u)}{nh} \sum_{k=1}^K \sum_{i=1}^n v_k f^{-1}(c_k) \{F(c_k - d_{i,k}^{**}) - F(c_k)\} K_i^{**} + o_p(1) \\ &= -\frac{f_{\mathbf{U}_0}^{-1}(u)}{nh} \sum_{k=1}^K \sum_{i=1}^n v_k d_{i,k}^{**} K_i^{**} \{1 + o_p(1)\} + o_p(1) \\ &= -\frac{f_{\mathbf{U}_0}^{-1}(u)}{nh} \sum_{k=1}^K \sum_{i=1}^n v_k \{c_k + g_0(X_i^T \gamma_0) - g_0(u) - g'_0(u)(X_i^T \gamma_0 - u)\} K_i^{**} \{1 + o_p(1)\} + o_p(1) \\ &= -\frac{f_{\mathbf{U}_0}^{-1}(u)}{nh} \sum_{i=1}^n \{g_0(X_i^T \gamma_0) - g_0(u) - g'_0(u)(X_i^T \gamma_0 - u)\} K_i^{**} \{1 + o_p(1)\} + o_p(1) \\ &= -\frac{1}{2} g''_0(u) \mu_2 h^2 \{1 + o_p(1)\} + o_p(1) = -\frac{1}{2} g''_0(u) \mu_2 h^2 + o_p(h^2). \end{aligned}$$

Furthermore,  $\text{Var}\{\hat{g}(u | \hat{\gamma}) | \mathbf{U}_0\} = v_0 R_2(\mathbf{v}) / \{nh f_{\mathbf{U}_0}(u)\}$ . This completes the proof.

**Proof of Theorem 3.** By a similar proof as of Theorem 1,

$$\begin{aligned} \frac{1}{n} \sum_{m=1}^M n_m \hat{S}_m &= \frac{1}{n} \sum_{m=1}^M \sum_{i=1}^{n_m} \{g'(\mathbf{X}_{m,i}^\top \gamma_0)\}^2 \{\mathbf{X}_{m,i} - E(\mathbf{X}_m | \mathbf{X}_m^\top \gamma_0)\} \{\mathbf{X}_{m,i} - E(\mathbf{X}_m | \mathbf{X}_m^\top \gamma_0)\}^\top + \frac{1}{n} \sum_{m=1}^M o_p(n_m) \\ &= \frac{1}{n} \sum_{i=1}^n \{g'(\mathbf{X}_i^\top \gamma_0)\}^2 \{\mathbf{X}_i - E(\mathbf{X} | \mathbf{X}^\top \gamma_0)\} \{\mathbf{X}_i - E(\mathbf{X} | \mathbf{X}^\top \gamma_0)\}^\top + o_p(1) \\ &= \mathbf{S}_n + o_p(1) = \mathbf{S} + o_p(1). \end{aligned}$$

Moreover, from the proof of Theorem 1, for each subsets  $n_m \rightarrow \infty$ ,  $m \in \{1, \dots, M\}$ , we have

$$n_m \mathbf{S}_m(\hat{\gamma}_m - \gamma_0) = -\left\{ \sum_{k=1}^K f(c_k) \right\}^{-1} \sum_{i=1}^{n_m} \sum_{k=1}^K \rho'_{\tau_k}(Y_{i,\tau_k}^*) g'(\mathbf{X}_{m,i}^\top \gamma_0) \{\mathbf{X}_{m,i} - E(\mathbf{X}_m | \mathbf{X}_m^\top \gamma_0)\}^\top + o_p(\sqrt{n_m}).$$

Thus, we can obtain

$$\begin{aligned} \sqrt{n}(\hat{\gamma}^{DC} - \gamma_0) &= \sqrt{n} \left\{ \left( \sum_{m=1}^M n_m \hat{S}_m \right)^+ \sum_{m=1}^M n_m \hat{S}_m \hat{\gamma}_m - \gamma_0 \right\} = \left( \frac{1}{n} \sum_{m=1}^M n_m \hat{S}_m \right)^+ \left\{ \frac{1}{\sqrt{n}} \sum_{m=1}^M n_m \hat{S}_m (\hat{\gamma}_m - \gamma_0) \right\} \\ &= \{\mathbf{S}^+ + o_p(1)\} \left\{ \frac{1}{\sqrt{n}} \sum_{m=1}^M n_m \mathbf{S}_m (\hat{\gamma}_m - \gamma_0) + \frac{1}{\sqrt{n}} \sum_{m=1}^M o_p(\sqrt{n_m}) \right\} \\ &= -\mathbf{S}^+ \left\{ \sum_{k=1}^K f(c_k) \right\}^{-1} \frac{1}{\sqrt{n}} \sum_{m=1}^M \sum_{i=1}^{n_m} \sum_{k=1}^K \rho'_{\tau_k}(Y_{i,\tau_k}^*) g'(\mathbf{X}_{m,i}^\top \gamma_0) \{\mathbf{X}_{m,i} - E(\mathbf{X}_m | \mathbf{X}_m^\top \gamma_0)\}^\top + o_p(1) \\ &= -\mathbf{S}^+ \left\{ \sum_{k=1}^K f(c_k) \right\}^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^n \sum_{k=1}^K \rho'_{\tau_k}(Y_{i,\tau_k}^*) g'(\mathbf{X}_i^\top \gamma_0) \{\mathbf{X}_i - E(\mathbf{X} | \mathbf{X}^\top \gamma_0)\}^\top + o_p(1). \end{aligned}$$

By the Cramér–Wald theorem and the Central Limit Theorem, Theorem 3 can be proved.

**Proof of Theorem 4.** From the proof of Theorem 2, for each subsets  $n_m \rightarrow \infty$ ,  $m \in \{1, \dots, M\}$ , we have

$$\sqrt{n_m h} \{\hat{g}_m(u | \hat{\gamma}^{DC}) - g_0(u)\} = -\frac{f_{\mathbf{U}_0}^{-1}(u)}{\sqrt{n_m h}} \sum_{k=1}^K \sum_{i=1}^{n_m} v_k f^{-1}(c_k) \{I(\varepsilon_i \leq c_k - d_{i,k}^{**}) - \tau_k\} K_i^{**} + o_p(1).$$

Thus, we can obtain

$$\begin{aligned}\sqrt{nh} \{ \hat{g}^{DC}(u | \hat{\gamma}^{DC}) - g_0(u) \} &= \sqrt{nh} \left\{ \frac{1}{n} \sum_{m=1}^M n_m \hat{g}_m^{DC}(u | \hat{\gamma}^{DC}) - g_0(u) \right\} \\ &= - \frac{f_{u_0}^{-1}(u)}{\sqrt{nh}} \sum_{k=1}^K \sum_{i=1}^n v_k f^{-1}(c_k) \{ I(\varepsilon_i \leq c_k - d_{i,k}^{**}) - \tau_k \} K_i^{**} + o_p(1).\end{aligned}$$

Thus, Theorem 4 can be proved.

### CRedit authorship contribution statement

**Rong Jiang:** Major analysis, Numerical computation, First draft of the paper. **Keming Yu:** Conceptualization, Formal analysis, Pointing out the real data sources, Revision of the draft.

### Acknowledgments

The three anonymous referees provided numerous valuable comments that improved the manuscript. This work was supported by the National Natural Science Foundation of China (No. 11801069) and the Fundamental Research Funds for the Central Universities of China (No. 2232020D-43).

### References

- [1] X. Chen, W. Liu, Y. Zhang, Quantile regression under memory constraint, *Ann. Statist.* 47 (2019) 3244–3273.
- [2] X. Chen, M. Xie, A split-and-conquer approach for analysis of extraordinarily large data, *Statist. Sinica* 24 (2014) 1655–1684.
- [3] E. Christou, M. Akritas, Single index quantile regression for heteroscedastic data, *J. Multivariate Anal.* 150 (2016) 169–182.
- [4] J. Fan, T.-C. Hu, Y. Truong, Robust nonparametric function estimation, *Scand. J. Stat.* 21 (1994) 433–446.
- [5] W. Hardle, P. Hall, H. Ichimura, Optimal smoothing in single-index models, *Ann. Statist.* 21 (1993) 157–178.
- [6] R. Jiang, M.-F. Guo, X. Liu, Composite quasi-likelihood for single-index models with massive datasets, *Comm. Statist. Simulation Comput.* (2020) 1–17.
- [7] R. Jiang, W.-M. Qian, Z.-G. Zhou, Single-index composite quantile regression with heteroscedasticity and general error distributions, *Statist. Papers* 57 (2016) 185–203.
- [8] R. Jiang, W.-M. Qian, Z.-G. Zhou, Weighted composite quantile regression for single-index models, *J. Multivariate Anal.* 148 (2016) 34–48.
- [9] R. Jiang, Z.-G. Zhou, W.-M. Qian, W.-Q. Shao, Single-index composite quantile regression, *J. Korean Stat. Soc.* 41 (2012) 323–332.
- [10] K. Knight, Limiting distributions for L1 regression estimators under general conditions, *Ann. Statist.* 26 (1998) 755–770.
- [11] E. Kong, Y. Xia, On the efficiency of online approach to nonparametric smoothing of big data, *Statist. Sinica* 29 (2019) 185–201.
- [12] D. Kraus, C. Czado, D-vine copula based quantile regression, *Comput. Statist. Data Anal.* 110 (2017) 1–18.
- [13] J. Lee, Q. Liu, Y. Sun, J. Taylor, Communication-efficient sparse regression, *J. Mach. Learn. Res.* 18 (2017) 1–30.
- [14] R. Li, D. Lin, B. Li, Statistical inference in massive data sets, *Appl. Stoch. Models Bus. Ind.* 29 (2013) 399–409.
- [15] H. Liang, X. Liu, R. Li, C.-L. Tsai, Estimation and testing for partially linear single-index models, *Ann. Statist.* 38 (2010) 3811–3836.
- [16] W. Lin, K. Kulasekera, Identifiability of single-index models and additive-index models, *Biometrika* 94 (2007) 496–501.
- [17] N. Lin, R. Xi, Aggregated estimating equation estimation, *Stat. Interface* 1 (2011) 73–83.
- [18] D. Lin, D. Zeng, On the relative efficiency of using summary statistics versus individual level data in meta-analysis, *Biometrika* 97 (2010) 321–332.
- [19] D. Liu, R. Liu, M. Xie, Multivariate meta-analysis of heterogeneous studies using only summary statistics: Efficiency and robustness, *J. Amer. Statist. Assoc.* 110 (2015) 326–340.
- [20] S. Ma, X. He, Inference for single-index quantile regression models with profile optimization, *Ann. Statist.* 44 (2016) 1234–1268.
- [21] Y. Mack, B. Silverman, Weak and strong uniform consistency of kernel regression estimates, *Probab. Theory Related Fields* 61 (1982) 405–415.
- [22] D. Pollard, Asymptotics for least absolute deviation regression estimators, *Econom. Theory* 7 (1991) 186–199.
- [23] B. Remillard, B.R. Nasri, T. Bouezmarni, On copula-based conditional quantile estimators, *Statist. Probab. Lett.* 128 (2017) 14–20.
- [24] E. Schifano, J. Wu, C. Wang, J. Yan, M.-H. Chen, Online updating of statistical inference in the big data setting, *Technometrics* 58 (2016) 393–403.
- [25] Y. Tang, H. Wang, H. Liang, Composite estimation for single-index models with responses subject to detection limits, *Scand. J. Stat.* 45 (2018) 444–464.
- [26] Y. Tian, Z. Qianqian, M. Tian, Estimation of linear composite quantile regression using EM algorithm, *Statist. Probab. Lett.* 117 (2016) 183–191.
- [27] Q. Wang, R. Wu, Shrinkage estimation of partially linear single-index models, *Statist. Probab. Lett.* 83 (2013) 2324–2331.
- [28] J.-L. Wang, L. Xue, L. Zhu, Y. Chong, Estimation for a partial-linear single-index model, *Ann. Statist.* 38 (2010) 246–274.
- [29] T. Wu, K. Yu, Y. Yu, Single-index quantile regression, *J. Multivariate Anal.* 101 (2010) 1607–1621.
- [30] Y. Xia, W.K. Hardle, Semi-parametric estimation of partially linear single-index models, *J. Multivariate Anal.* 97 (2006) 1162–1184.
- [31] Y. Xia, H. Tong, W. Li, L.-X. Zhu, An adaptive estimation of dimension reduction space, *J. R. Stat. Soc. Ser. B* 64 (2002) 363–410.
- [32] M. Xie, W. Strawderman, Confidence distributions and a unifying framework for meta-analysis, *J. Amer. Statist. Assoc.* 106 (2011) 320–333.
- [33] H. Zou, M. Yuan, Composite quantile regression and the oracle model selection theory, *Ann. Statist.* 36 (2008) 1108–1126.