Single-index composite quantile regression for massive data

Rong Jiang[a]; Keming Yu[b,1]

[a]Department of Statistics, Donghua University, Shanghai, 201620, China
[b]Brunel University London, UK

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, which have wide applications in many scientific fields including biostatistics, economics, and financial econometrics. 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.

Keywords: Single-index model; Composite quantile regression; Massive data.
2010 MSC: 60G08; 62G20.

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., 1993), we restrict our attention to the single-index model in the following form:

$$\mathbf{Y} = g_0(\mathbf{X}^\top \gamma_0) + \varepsilon, \tag{1.1}$$

where $\mathbf{Y}$ 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; γ_0 is the unknown index vector coefficient; and for the sake of identifiability, following Lin and Kulasekera (2007), we assume that $\|\gamma_0\| = 1$ and that the first component of γ_0 is positive, where $\|\cdot\|$ denotes the Euclidean norm and the error term ε is assumed to be independent of $\mathbf{X}$ with $E[\varepsilon] = 0$.

To estimate the parameters in model (1.1), Yu and Ruppert (2002) proposed the penalized spline estimation procedure, while Xia and Härdle (2006) applied the minimum average variance estimation (MAVE) method, which was originally introduced by Xia et al. (2002) for dimension reduction. Wu et al. (2010) studied quantile regression (QR), Feng et al. (2012) proposed the rank-based outer product of gradients estimator method, and Liu et al. (2013) applied the local linear model regression estimator method. 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. (2010) proposed a two-stage procedure and Liang et al. (2010) employed a profile least squares approach. Christou and Akritas (2016) developed a Nadaraya–Watson QR.

Existing non-iterative estimation procedures for single-index model were built on either least squares or quantile regression methods. However, the least squares method is sensitive to outliers and does not perform well when the error distribution is heavily skewed. The quantile regression method is an obvious alternative to the least squares.

[1]Corresponding author
E-mail address: jrtrying@dhu.edu.cn (Rong Jiang), keming.yu@brunel.ac.uk (Keming Yu)

However, the relative efficiency of the quantile regression can be arbitrarily small when compared with the least squares. In contrast to the above methods, the CQR method was first proposed by Zou and Yuan (2008) 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 $k = 1, \ldots, K$ and $0 < \tau_1 < \cdots < \tau_K < 1$. It is easy to see that the CQR method is a sum of different quantile regressions. Zou and Yuan (2008) showed that the CQR estimator shares robustness from QR and the relative efficiency of the CQR estimator compared with the least squares estimator is greater than 70% regardless of the error distribution. Jiang et al. (2012) proposed a CQR estimation for single-index model, and Jiang et al. (2016a) showed the relative efficiency of the CQR estimator compared with the least squares estimator for single-index model.

In this study, we use the CQR method to estimate the index coefficients γ_0 in model (1.1). Furthermore, we propose a non-iterative method based on the CQR method for estimating the parametric component of model (1.1) 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. (2012). 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. These barriers can be overcome with either newly developed statistical methodologies or computational methodologies. As a solution to the memory and storage limitation problems, the divide-and-conquer method (Lin and Xi, 2011; Chen and Xie, 2014; Schifano et al., 2016) 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 (2010) showed that such a meta-estimator asymptotically achieves the same asymptotic variance as that based on the entire dataset. Lin and Xi (2011) introduced an aggregated estimating equation estimator based on the Hessian matrix of the loss function. In our case with a composite quantile loss function, however, the Hessian matrix does not exist. Xie et al. (2012) developed a robust meta-analysis-type approach through the confidence distribution approach. Liu et al. (2015) 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., 2017), generalized linear model (Lin and Xi, 2011; Chen and Xie, 2014; Tang et al., 2016; Zhao et al., 2017), and nonparametric model (Lu et al., 2016; Kong and Xia, 2018). 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:

(1) Our estimation procedure directly targets the model parameter γ_0 in model (1.1) 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.

(2) 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 (1.1). 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 relegated to the Appendix.

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 γ_0 in model (1.1) solves the following minimization problem:

$$\gamma_0 = \arg\min_{\gamma} \sum_{k=1}^{K} E\left[\rho_{\tau_k}\left\{\mathbf{Y} - Q_{\tau_k}\left(\mathbf{Y}|\mathbf{X}^\top\gamma\right)\right\}\right], \tag{2.1}$$

where $Q_{\tau_k}(\mathbf{Y}|\mathbf{X}^\top\gamma) = c_k + g(\mathbf{X}^\top\gamma)$, $c_k = F^{-1}(\tau_k)$, $F(\cdot)$ is the cumulative distribution function of the model error ε, and $\rho_{\tau_k}(r) = \tau_k r - rI(r<0), k = 1, \ldots, K$, is the K check loss functions with $0 < \tau_1 < \cdots < \tau_K < 1$.

For model (1.1), $Q_{\tau_k}(\mathbf{Y}|\mathbf{x}^\top\gamma) = \inf\{y : P(Y \le y \mid \mathbf{X}^\top\gamma = \mathbf{x}^\top\gamma) \ge \tau_k\}$. Since in (2.1), $Q_{\tau_k}(\mathbf{Y}|\mathbf{X}^\top\gamma)$, $k = 1,\ldots,K$, are unknown, (2.1) should be minimized by solving one simple problem that $Q_{\tau_k}(\mathbf{Y}|\mathbf{X}^\top\gamma)$ must be replaced with an estimator for each k. However, there are no closed-form expressions for the estimator of $Q_{\tau_k}(\mathbf{Y}|\mathbf{X}^\top\gamma)$. Thus, this often leads to iterative algorithms for estimating γ_0, which raises the computational complexity. To overcome this difficulty, we define, for any given $\gamma \in R^p$, the function $H_{\tau_k}(t \mid \gamma) : R \to R$ as $H_{\tau_k}(t \mid \gamma) = E\left[Q_{\tau_k}(\mathbf{Y}|\mathbf{X}) \mid \mathbf{X}^\top\gamma = t\right]$, $k = 1,\ldots,K$, where $Q_{\tau_k}(\mathbf{Y}|\mathbf{x}) = \inf\{y : P(Y \le y \mid \mathbf{X} = \mathbf{x}) \ge \tau_k\}$. Hence, under single-index models, this specifies that $Q_{\tau_k}(\mathbf{Y}|\mathbf{X}^\top\gamma_0) = H_{\tau_k}(\mathbf{X}^\top\gamma_0 \mid \gamma_0), k = 1,\ldots,K$; thus γ_0 also satisfies

$$\gamma_0 = \arg\min_{\gamma} \sum_{k=1}^{K} E\left[\rho_{\tau_k}\left\{\mathbf{Y} - H_{\tau_k}(\mathbf{X}^\top\gamma \mid \gamma)\right\}\right]. \tag{2.2}$$

Let $\{Y_i, \mathbf{X}_i\}_{i=1}^n$ be an independent and identically distributed (i.i.d.) sample from $\{\mathbf{Y}, \mathbf{X}\}$. Thus, the right term of (2.2) can be approximated by

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

Thus, a non-iterative estimation process based on the Nadaraya–Watson estimator method can be constructed. First, we obtain the Nadaraya–Watson estimator of $H_{\tau_k}(\cdot \mid \gamma)$ for each k (see Christou and Akritas, 2016):

$$\hat{H}_{\tau_k}(t \mid \gamma) = \frac{\sum_{i=1}^{n}\hat{Q}_{\tau_k}(\mathbf{Y}|\mathbf{X}_i)\tilde{K}_{h_k}\left(\mathbf{X}_i^\top\gamma - t\right)}{\sum_{i=1}^{n}\tilde{K}_{h_k}\left(\mathbf{X}_i^\top\gamma - t\right)}, \tag{2.3}$$

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 γ_0 by solving the following minimization problem:

$$\hat{\gamma} = \arg\min_{\gamma} \frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{K}\rho_{\tau_k}\left\{Y_i - \hat{H}_{\tau_k}(\mathbf{X}_i^\top\gamma \mid \gamma)\right\}. \tag{2.4}$$

After obtaining the estimator $\hat{\gamma}$ of γ_0 in model (1.1), we can estimate $g_0(\cdot)$ in model (1.1). We used the weighted local CQR (WLCQR) proposed by Jiang et al. (2016b), which is valid without a symmetric error condition. For any given point u, the final estimate of $g_0(\cdot)$ is

$$\hat{g}(u \mid \hat{\gamma}) = \sum_{k=1}^{K} v_k \hat{a}_k, \tag{2.5}$$

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

$$(\hat{a}_1,\ldots,\hat{a}_K,\hat{b}) = \arg\min_{(a_1,\ldots,a_K,b)} \frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{K}\rho_{\tau_k}\left\{Y_i - a_k - b\left(\mathbf{X}_i^\top\hat{\gamma} - u\right)\right\}\tilde{K}_h\left(\mathbf{X}_i^T\hat{\gamma} - u\right). \tag{2.6}$$

Remark 2.1: In general, given K, one can use the equally spaced quantiles at $\tau_k = k/(K+1)$ for $k = 1,\ldots,K$; see Zou and Yuan (2008). Moreover, we can use redefined BIC (Tian et al., 2016) to select K as follows:

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

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

Remark 2.2: $\hat{Q}_{\tau_k}(\mathbf{Y}|\mathbf{X}_i)$, $k = 1, \ldots, K$, in (2.3) can be obtained by the D-vine copula proposed by Kraus and Czado (2017), as follows:

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

where $\hat{C}^{-1}_{F_Y|F_1,\ldots,F_p}$ is the estimator of the conditional Copula quantile function $C^{-1}_{F_Y|F_1,\ldots,F_p}$, $\hat{F}_Y(y) = \frac{1}{n+1}\sum_{i=1}^{n} I(Y_i \le y)$, and $\hat{F}_j(x_j) = \frac{1}{n+1}\sum_{i=1}^{n} I(X_{ij} \le x_j)$, $j = 1, \ldots, p$. The detailed estimation process can be found in Section 3.2 of Kraus and Czado (2017).

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 $\mathbf{U}_0 = \mathbf{X}^\top\gamma_0$. We choose the kernel $\tilde{K}(\cdot)$ as a symmetric density function and write $\mu_j = \int u^j\tilde{K}(u)du$, $\nu_j = \int u^j\tilde{K}^2(u)du$, $R_1 = \left\{\sum_{k=1}^{K} f(c_k)\right\}^{-2}\sum_{k=1}^{K}\sum_{k'=1}^{K}\tau_{kk'}$, $R_2(\mathbf{v}) = \sum_{k=1}^{K}\sum_{k'=1}^{K}\frac{v_k v_{k'}\tau_{kk'}}{f(c_k)f(c_{k'})}$, $\tau_{kk'} = \tau_k \wedge \tau_{k'} - \tau_k\tau_{k'}$, and $h_{\max} = \max_{1\le k\le K}\{h_k\}$, $h_{\min} = \min_{1\le k\le K}\{h_k\}$.

Theorem 2.1. Suppose that Conditions **C1–C4** given in the Appendix hold. $n \to \infty$, $nh_{\max}^4 \to 0$ and $nh_{\min} \to \infty$; then,

$$\sqrt{n}(\hat{\gamma} - \gamma_0) \xrightarrow{L} N\left(0, \mathbf{S}^{-}R_1\right), \tag{2.7}$$

where $\xrightarrow{L}$ stands for convergence in the distribution, $\mathbf{S} = E\left\{\{g_0'(\mathbf{X}^T\gamma_0)\}^2[\mathbf{X} - E(\mathbf{X}|\mathbf{X}^T\gamma_0)][\mathbf{X} - E(\mathbf{X}|\mathbf{X}^T\gamma_0)]^T\right\}$, and $\mathbf{S}^-$ is the Moore-Penrose inverse of symmetric matrix $\mathbf{S}$, since $\mathbf{S}$ is not full ranked (see Ma and He, 2016 and Tang, et al., 2018).

Theorem 2.2. Under the same conditions as in Theorem 2.1, if $n \to \infty, h \to 0$ and $nh \to \infty$, then for an interior point u of the support of $f_{\mathbf{U}_0}(\cdot)$,

$$\sqrt{nh}\left\{\hat{g}(u \mid \hat{\gamma}) - g_0(u) - \frac{1}{2}g_0''(u)\mu_2 h^2\right\} \xrightarrow{L} N\left(0, \frac{\nu_0 R_2(\mathbf{v})}{f_{\mathbf{U}_0}(u)}\right).$$

The bias of $\hat{g}(u \mid \hat{\gamma})$ is free 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 $\hat{g}(u \mid \hat{\gamma})$. Thus,

$$\mathbf{v}_{opt} = \arg\min_{\mathbf{v}} R_2(\mathbf{v}) = \frac{(\mathbf{c}^\top\mathbf{A}^{-1}\mathbf{c})\mathbf{A}^{-1}\mathbf{1} - (\mathbf{1}^\top\mathbf{A}^{-1}\mathbf{c})\mathbf{A}^{-1}\mathbf{c}}{(\mathbf{c}^\top\mathbf{A}^{-1}\mathbf{c})(\mathbf{1}^\top\mathbf{A}^{-1}\mathbf{1}) - (\mathbf{1}^\top\mathbf{A}^{-1}\mathbf{c})^2}, \tag{2.8}$$

where $\mathbf{c}$ is a K-dimensional column vector with the kth 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 $\tau_{kk'}/(f(c_k)f(c_{k'}))$, $k, k' = 1, \ldots, K$. Thus, with these optimal weights, the asymptotic variance of $\hat{g}(u \mid \hat{\gamma})$ is $f_{\mathbf{U}_0}^{-1}(u)\nu_0 R_2(\mathbf{v}_{opt})$.

Remark 2.3: For Theorem 2.1, the proposed estimator achieves the same efficiency as the iterative CQR estimator proposed by Jiang et al. (2012). The results of Theorem 2.2 are thus the same as those of Theorem 5 in Jiang et al. (2016b). Thus, the selection of the optimal bandwidth h can be found in Jiang et al. (2016b).

Remark 2.4: From (2.8), we can see that the optimal weight vector $\mathbf{v}_{opt}$ is complicated and involves the density of the errors $f(c_k)$, $k = 1, \ldots, K$. 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 $\hat{\varepsilon}$ can provide a consistent estimation $\hat{f}(\hat{c}_k)$ of $f(c_k)$, $k = 1, \ldots, K$. The details can be found in Section 2.3 of Jiang et al. (2016b).

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. (2002). 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

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

Note that $ARE_{\gamma_0}(NICQR, MAVE)$ is the same as the result obtained by Zou and Yuan (2008). Thus,

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

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. (2012). From a similar deduction to that of Jiang et al. (2016a), 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

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

where $R_3 = \sum_{k=1}^{K}\sum_{k'=1}^{K}[\tau_{kk'}/\{f(c_k)f(c_{k'})\}]/K^2$. Jiang et al. (2016a) showed that when the error distribution is symmetric, we have

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

Remark 2.5: To appreciate how much efficiency $ARE_{\gamma_0}(NICQR, MAVE)$, $ARE_{g_0}(WLCQR, MAVE)$, and $ARE_{g_0}(WLCQR, LCQR)$ is gained in practice, Figure 2 in Zou and Yuan (2008) reports $ARE_{\gamma_0}(NICQR, MAVE)$ with various error distributions for various K and Table 2 in Jiang et al. (2016a) reports $ARE_{g_0}(WLCQR, MAVE)$ and $ARE_{g_0}(WLCQR, LCQR)$ with various error distributions for various K.

2.4. Algorithm

To obtain the estimator $\hat{\gamma}$ by minimizing (2.4), we use the procedure introduced in Wang and Wu (2013), which consists of using a local linear approximation of $\hat{H}_{\tau_k}(\mathbf{X}_i^\top\gamma \mid \gamma)$ around an initial value $\tilde{\gamma}$ of γ_0. This yields

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

where $\hat{H}'_{\tau_k}(\mathbf{X}_i^\top\tilde{\gamma} \mid \tilde{\gamma}) = \left.\frac{\partial \hat{H}_{\tau_k}(\mathbf{X}_i^\top\gamma|\gamma)}{\partial\gamma}\right|_{\gamma=\tilde{\gamma}}$. Then, the proposed estimator is obtained from

$$\hat{\gamma} = \arg\min_{\gamma}\frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{K}\rho_{\tau_k}\left\{Y_i - \hat{H}_{\tau_k}(\mathbf{X}_i^\top\tilde{\gamma} \mid \tilde{\gamma}) - \hat{H}'_{\tau_k}(\mathbf{X}_i^\top\tilde{\gamma} \mid \tilde{\gamma})(\gamma - \tilde{\gamma})\right\}. \quad (2.9)$$

The steps of the NICQR procedure are summarized as follows.
Step 2.1 (Initialization step). Obtain an initial estimate $\tilde{\gamma}$ and $\tilde{g}(\cdot)$ by using the MAVE method, which can be obtained from the R package MAVE.
Step 2.2. $\hat{Q}_{\tau_k}(\mathbf{Y}|\mathbf{X}_i)$ in (2.3) is obtained by the D-vine copula proposed by Kraus and Czado (2017), which can be implemented by using the R package vinereg. Thus, we can obtain $\hat{H}_{\tau_k}(\mathbf{X}_i^\top\tilde{\gamma} \mid \tilde{\gamma})$ and $\hat{H}'_{\tau_k}(\mathbf{X}_i^\top\tilde{\gamma} \mid \tilde{\gamma})$.
Step 2.3 (Estimation of γ_0). Update $\hat{\gamma}$ by minimizing the objective function in (2.9), which can be achieved by using the cqr.cd in the R package cqrReg.
Step 2.4. Obtain the estimate $\hat{\mathbf{v}}$ of $\mathbf{v}_{opt}$ in (2.8). The estimation of $\hat{\mathbf{f}} = \left(\hat{f}(\hat{c}_1), \ldots, \hat{f}(\hat{c}_K)\right)^\top$ is as follows: compute $\hat{\varepsilon}_i = Y_i - \tilde{g}(\mathbf{X}_i^\top\tilde{\gamma})$ from the results in Step 2.1 and then use the kernel density estimation $\hat{f}(\cdot) = \frac{1}{n\tilde{h}}\sum_{i=1}^{n}\tilde{K}_{\tilde{h}}(\hat{\varepsilon}_i - \cdot)$ to estimate $f(\cdot)$. The $\hat{c}_k$ of c_k is the sample τ_k-quantile of $\{\hat{\varepsilon}_i\}_{i=1}^{n}$, where the bandwidth $\tilde{h}$ is chosen from $\tilde{h} = 0.9 \times \min\{std(\hat{\varepsilon}_1, \ldots, \hat{\varepsilon}_n), IQR(\hat{\varepsilon}_1, \ldots, \hat{\varepsilon}_n)/1.34\} \times n^{-1/5}$, std and IQR denote the sample standard deviation and sample interquantile, respectively (see Silverman, 1986).
Step 2.5 (Estimation of $g_0(\cdot)$). From the estimate $\hat{\gamma}$ in Step 2.3, for a given point u, estimate $g_0(u)$ from (2.6) with $\hat{\mathbf{v}}$ in Step 2.4, 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.

3.1. DC-NICQR method

It is infeasible to solve the optimization problem in (2.4) and (2.6) 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 mth subset contains n_m observations: $(\mathbf{X}_{m,i}, Y_{m,i})$, $i = 1, \ldots, n_m$, and $n = \sum_{m=1}^{M} n_m$. Here we assume M is fixed. Based on the asymptotic normality in (2.7), we form the asymptotic confidence density of γ_0 as

$$\hat{h}_n(\gamma_0) \propto \exp\left[-\frac{n}{2} R_1^{-1} (\gamma_0 - \hat{\gamma})^\top \mathbf{S} (\gamma_0 - \hat{\gamma})\right].$$

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

$$\hat{h}_n(\gamma_0) \propto \exp\left[-\frac{n}{2} \hat{R}_1^{-1} (\gamma_0 - \hat{\gamma})^\top \hat{S} (\gamma_0 - \hat{\gamma})\right],$$

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 $\hat{S} = \frac{1}{n} \sum_{i=1}^{n} \left[\hat{H}'(\mathbf{X}_i^\top \hat{\gamma} \mid \hat{\gamma})\right]^\top \hat{H}'(\mathbf{X}_i^\top \hat{\gamma} \mid \hat{\gamma})$ is the estimation of S. Note that $\hat{H}_{\tau_k}(\mathbf{X}_i^\top \gamma \mid \gamma)$ is the estimator of $H_{\tau_k}(\mathbf{X}^\top \gamma_0 \mid \gamma_0) = Q_{\tau_k}(\mathbf{Y} | \mathbf{X}^\top \gamma_0) = c_k + g(\mathbf{X}^\top \gamma_0)$. Thus, $\hat{H}'_{\tau_k}(\mathbf{X}^\top \gamma \mid \gamma)$ is the estimator of $g_0'(\mathbf{X}^T \gamma_0)[\mathbf{X} - E(\mathbf{X} | \mathbf{X}^T \gamma_0)]^T$, which is independent of τ_k. A similar result can be found in Section 4.1 of Christou and Akritas (2016). Therefore, we rewrite $\hat{H}'_{\tau_k}(\mathbf{X}^\top \gamma \mid \gamma)$ as $\hat{H}'(\mathbf{X}^\top \gamma \mid \gamma)$. In the simulation, we take $\hat{H}'(\mathbf{X}^\top \gamma \mid \gamma)$ with $\hat{H}'_{1/2}(\mathbf{X}^\top \gamma \mid \gamma)$.

It is also infeasible to obtain $\hat{h}_n(\gamma_0)$ 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 (2.7) to construct the asymptotic confidence density $\hat{h}_{n_m}(\gamma_0)$, $m = 1, \ldots, M$. Then, we can combine the M confidence densities to derive a combined estimator of γ_0. The combined estimator is denoted by $\hat{\gamma}^{DC}$ according to the following equation:

$$\hat{\gamma}^{DC} = \arg\max_{\gamma} \log \prod_{m=1}^{M} \hat{h}_{n_m}(\gamma) = \arg\min_{\gamma} \sum_{m=1}^{M} \frac{n_m}{2} \hat{R}_{1,m}^{-1} (\gamma - \hat{\gamma}_m)^\top \hat{S}_m (\gamma - \hat{\gamma}_m), \tag{3.1}$$

where $\hat{R}_{1,m} = \left\{\sum_{k=1}^{K} \hat{f}_m(\hat{c}_{k,m})\right\}^{-2} \sum_{k=1}^{K} \sum_{k'=1}^{K} \tau_{kk'}$, $\hat{S}_m = \frac{1}{n_m} \sum_{i=1}^{n_m} \left[\hat{H}'(\mathbf{X}_{m,i}^\top \hat{\gamma}_m \mid \hat{\gamma}_m)\right]^\top \hat{H}'(\mathbf{X}_{m,i}^\top \hat{\gamma}_m \mid \hat{\gamma}_m)$, and $\hat{\gamma}_m$ is the estimation of γ_0 by using the methodology to solve equation (2.4) for the mth subset. 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 (3.3) in Section 3.3). Thus, we use $\hat{R}_{1,m} = \hat{R}_1$ for $m = 1, \ldots, M$. By some simple algebra, the solution to the optimization problem in (3.1) can be expressed as a form of the weighted average of $\hat{\gamma}_m$, $m = 1, \ldots, M$, as

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

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

Step 3.1. Without loss of generality, the entire dataset is partitioned into M subsets; the mth subset contains n_m observations: $(\mathbf{X}_{m,i}, Y_{m,i})$, $i = 1, \ldots, n_m$, and $n = \sum_{m=1}^{M} n_m$.

Step 3.2. For each subset, obtain the estimators $\hat{\gamma}_m$, $m = 1, \ldots, M$, using the methodology to solve equation (2.4). The aggregated estimator for γ_0, as a weighted average of $\hat{\gamma}_m$, $m = 1, \ldots, M$, 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 3.3. After obtaining the estimation $\hat{\gamma}^{DC}$ of γ_0 in model (1.1), we can estimate $g_0(\cdot)$ in model (1.1). For any given point u, we find the estimators $\hat{g}_m(u)$ for each subset, using the methodology in (2.5). Then, the final estimate of $g_0(u \mid \hat{\gamma}^{DC})$, as a weighted average of $\hat{g}_m(u \mid \hat{\gamma}^{DC})$, $m = 1, \ldots, M$, 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 3.1. Assume that the conditions of Theorem 2.1 are satisfied; then,

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

Theorem 3.2. Assume that the conditions of Theorem 2.2 are satisfied with the optimal weight $\mathbf{v}_{opt}$ in (2.8); then,

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

Remark 3.1: The limiting distributions of $\hat{\gamma}^{DC}$ and $\hat{g}(\cdot)^{DC}$ in Theorems 3.1 and 3.2 are those of $\hat{\gamma}$ and $\hat{g}(\cdot)$ in Theorems 2.1 and 2.2, where all the data are analyzed. Thus, the DC-NICQR estimators are asymptotically equivalent to the corresponding estimator using the full datasets.

Remark 3.2: In Theorem 3.1 and Theorem 3.2, the resulting estimate is robust to the choice of block size M and subset size $n_m, m = 1, \ldots, M$. Thus, M and n_m are chosen so that the estimation of γ_0 can be easily handled within each block.

Remark 3.3: In Step 3.3, we use the same bandwidth h for each subset based on Theorem 3.2 following the method in Remark 2.3.

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

As mentioned in Remark 2.4. the optimal weight vector $\mathbf{v}_{opt}$ involves the density of the errors $f(c_k)$, $k = 1, \ldots, K$. We can use the kernel density estimation $\frac{1}{n}\sum_{i=1}^{n} \tilde{K}_{\tilde{h}}\left(\hat{\varepsilon}_i - \cdot\right)$ to estimate $f(\cdot)$. The estimator $\hat{c}_k$ of c_k is the sample τ_k-quantile of $\{\hat{\varepsilon}_i, i = 1, \ldots, n\}$, where $\hat{\varepsilon}_i = Y_i - \hat{g}^{DC}(\mathbf{X}_i^{\top}\hat{\gamma}^{DC})$. Therefore, we can obtain the estimation of $f(c_k)$ from $\hat{f}(\hat{c}_k)$ for $k = 1, \ldots, K$. However, when the available computer memory is much smaller than n, sorting $\{\hat{\varepsilon}_i, i = 1, \ldots, n\}$ becomes impossible. To overcome this difficulty, Li et al. (2013) 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. (2013), we can estimate c_k, $k = 1, \ldots, K$, as follows:

$$\hat{c}_k = \frac{1}{n}\sum_{m=1}^{M} n_m \hat{c}_{k,m}, \tag{3.2}$$

where $\hat{c}_{k,m}$ is the sample τ_k-quantile of $\{\hat{\varepsilon}_{m,i}, i = 1, \ldots, n_m\}$ and $\hat{\varepsilon}_{m,i} = Y_{m,i} - \hat{g}_m(\mathbf{X}_{m,i}^{\top}\hat{\gamma}^{DC})$. Then, from (3.2), the weighted combined estimator of $f(c_k)$, $k = 1, \ldots, K$, is given by

$$\hat{f}^{DC}(\hat{c}_k) = \frac{1}{n}\sum_{m=1}^{M} n_m \hat{f}_m(\hat{c}_k), \tag{3.3}$$

where $\hat{f}_m(\hat{c}_k) = \frac{1}{n_m}\sum_{i=1}^{n_m} \tilde{K}_{\tilde{h}_m}\left(\hat{\varepsilon}_{m,i} - \hat{c}_k\right), m = 1, \ldots, M$, are the kernel density estimations within each subset. Then, 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 kth element $\hat{c}_k$ and $\hat{\mathbf{A}}$ is a $K \times K$ matrix with the (k, k') element $\tau_{kk'} / \left(\hat{f}^{DC}(\hat{c}_k)\hat{f}^{DC}(\hat{c}_{k'})\right)$, $k, k' = 1, \ldots, K$.

Remark 3.4: The bandwidth selection is taken by $\tilde{h}_m = \left(\frac{n_m}{n}\right)^{1/5} h_m^*$ as selected by Li et al. (2013), where h_m^* is selected as $0.9 \times 1.06 \times \sigma_m \times n_m^{-1/5}$ and $\hat{\sigma}_m = std(\hat{\varepsilon}_{m,1}, \ldots, \hat{\varepsilon}_{m,n_m})$ to estimate σ_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. (2016) 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. (2016a), 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 $\tau_k = k/10, k = 1, \ldots, 9$. 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 H*ä*rdle, 2006);
(2) QR with $\tau = 0.5$ ($QR_{0.5}$) (see Wu et al., 2010);
(3) CQR with $K = 9$ (CQR_9) (see Jiang et al., 2012);
(4) Non-iterative least squares estimation (NILSE) (see Wang and Wu, 2013); and
(5) Non-iterative QR with $\tau = 0.5$ ($NIQR_{0.5}$) (see Christou and Akritas, 2016),
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:

$$\mathbf{Y} = \sin\{\pi(\mathbf{X}^\top\gamma_0 - A)/(B - A)\} + 0.2\varepsilon, \tag{4.1}$$

where $\mathbf{X}$ is uniformly distributed on $[0, 1]^3$, $\gamma_0 = (1, 1, 1)^\top/\sqrt{3}$, 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 ε: a standard normal distribution ($N(0, 1)$), uniformly distributed on $[-2, 2]$ ($U(-2, 2)$), 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 (MSEs) $\left(MSE = \sqrt{(\hat\gamma - \gamma_0)^\top(\hat\gamma - \gamma_0)}\right)$, and $Absolute\ Bias = |\hat\gamma - \gamma_0|$ of the estimate $\hat\gamma$ 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 are very small.

(ii) The $NICQR_9$ estimator performs better than $NIQR_9$ and is close to CQR_9 for different error distributions. Moreover, $NICQR_9$ is consistently superior to the other four methods except CQR_9 when the error is the Chi-square distribution.

(iii) t in Table 1 and Figure 1 are the average computing time in seconds used to estimate the index parameter. From t, we see that the operation time of $NICQR_9$ is faster than that of CQR_9 and with the increase of n, the gap is more obvious. Moreover, for the uniform and Chi-square error distributions, the operation time of $NICQR_9$ is faster than that of $QR_{0.5}$.

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

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

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

(i) When the error follows 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 U[-2,2], $QR_{0.5}$ and $NIQR_9$ perform worse than the other four methods. CQR_9 and $NICQR_9$ perform nearly as well as MAVE and NILSE.

(iii) For the asymmetric error $\chi^2(3)$, $NICQR_9$ performs the best.

Table 1 The means of Absolute Bias, MSE and t (standard deviation) for model (4.1).

Error	Methods	γ_1	γ_2	γ_3	MSE	t
N(0,1)	MAVE	0.0188 (0.0137)	0.0183 (0.0144)	0.0171 (0.0126)	0.0349 (0.0178)	0.2831
	$QR_{0.5}$	0.0213 (0.0160)	0.0195 (0.0152)	0.0194 (0.0144)	0.0386 (0.0204)	1.2297
	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
	$NIQR_{0.5}$	0.0333 (0.0259)	0.0322 (0.0261)	0.0322 (0.0239)	0.0629 (0.0337)	0.5427
	$NICQR_9$	0.0309 (0.0251)	0.0319 (0.0233)	0.0287 (0.0235)	0.0586 (0.0329)	2.8645
U(-2,2)	MAVE	0.0245 (0.0180)	0.0250 (0.0198)	0.0218 (0.0167)	0.0459 (0.0242)	0.2932
	$QR_{0.5}$	0.0328 (0.0243)	0.0317 (0.0254)	0.0306 (0.0230)	0.0609 (0.0328)	3.5831
	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
	$NIQR_{0.5}$	0.0486 (0.0352)	0.0443 (0.0361)	0.0479 (0.0367)	0.0906 (0.0478)	0.5757
	$NICQR_9$	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
	$QR_{0.5}$	0.0448 (0.0355)	0.0475 (0.0363)	0.0440 (0.0344)	0.0874 (0.0481)	3.8438
	CQR_9	0.0294 (0.0249)	0.0344 (0.0249)	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
	$NIQR_{0.5}$	0.0620 (0.0471)	0.0656 (0.0480)	0.0600 (0.0443)	0.1198 (0.0621)	0.7341
	$NICQR_9$	0.0298 (0.0252)	0.0306 (0.0236)	0.0297 (0.0242)	0.0618 (0.0335)	3.4304

Table 2 The mean of ASE (standard deviation) for model (4.1).

Methods	N(0,1)	U(-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)

4.1.2. Simulation example 2

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

$$\mathbf{Y} = \sin(2\mathbf{X}^\top\gamma_0) + \exp\{(\mathbf{X}^\top\gamma_0)^2\} + 0.5\cos\{2\pi(\mathbf{X}^\top\gamma_0)\}\varepsilon, \tag{4.2}$$

where the index parameter $\gamma_0 = (2, 2, 1)^\top/3$, and $\mathbf{X}$ is uniformly distributed on $[0, 1]^3$, and the residual ε 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 $WCQR_9$ proposed by Jiang et al. (2016a) 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 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. Real 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.

Table 3 The means of Absolute Bias, MSE, and t (standard deviation) for model (4.2).

Methods	γ_1	γ_2	γ_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

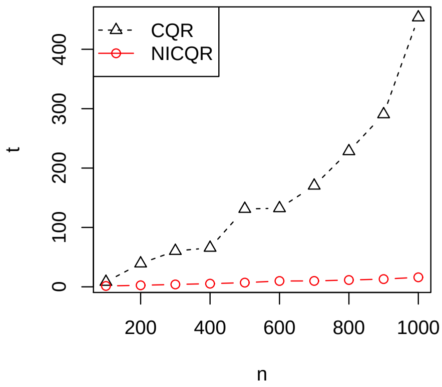


Figure 1: The average computing time (in seconds) of estimated γ_0 based on CQR_9 and $NICQR_9$ for different n.

Table 4 Single-index coefficient estimates and MSE 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
$QR_{0.5}$	0.5193	-0.8470	-0.1131	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

Many regression studies have used this dataset and found potential relationships among *medv* and RM, LSTAT and DIS (see Wu et al., 2010). 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 on 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. Figure 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.

4.2. *Example for massive datasets*

In this section, we investigate the performance of our DC-$NICQR_9$ method compared with the oracle full data $NICQR_9$ method. We evaluate the methods from two perspectives: (i) accuracy in estimating the oracle full data $NICQR_9$ statistics and (ii) computational efficiency in terms of run time.

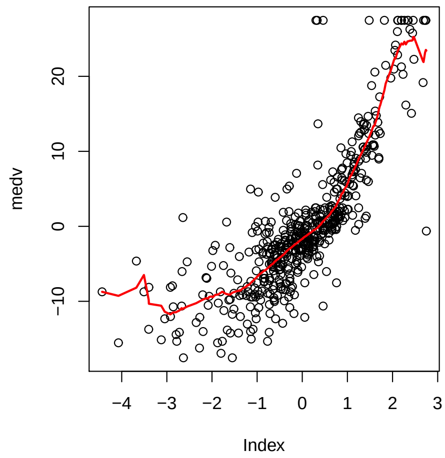


Figure 2: Estimated single index composite quantile regression for Boston housing data. The dots are the observations *medv* and the curve is the estimated *medv*.

Table 5 The means of MSE, Absolute Bias (standard deviation) and t_γ for model (4.3).

n	M	Absolute Bias (γ_1)	Absolute Bias (γ_2)	MSE	t_γ	AE	t_g
2000	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
5000	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
10000	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
100000	10	0.0045 (0.0030)	0.0022 (0.0015)	0.0050 (0.0033)	3977.09	0.0205 (0.0149)	1792.62

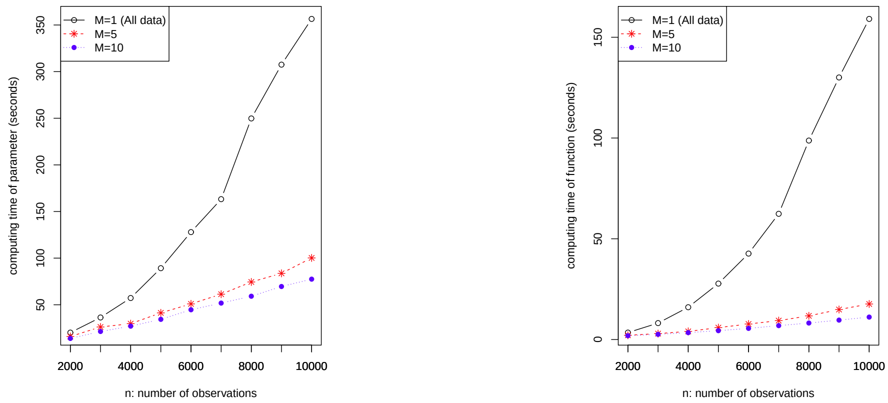
4.2.1. Simulation example 3

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

$$y = \cos\left(\mathbf{X}^\top\gamma_0\right) + \exp\left\{-\left(\mathbf{X}^\top\gamma_0\right)^2\right\} + 0.2\varepsilon, \tag{4.3}$$

where $\gamma_0 = (1,2)^\top/\sqrt{5}$, the covariate vector $\mathbf{X}$ is generated as a multivariate normal with mean zero and covariance matrix $\mathrm{Var}(\mathbf{X}) = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$, and the residual ε 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 Absolute Bias, MSE, computing time (t_γ) 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 $\mathbf{X}^\top\gamma_0$. According to Table 5, the performance of the DC-NICQR$_9$ estimators is similar to that of the estimators analyzing the full dataset ($M = 1$). Further, the computing time of γ and $g(\cdot)$ decreases drastically as M increases, as seen in Table 5 and Figure 3. With $n = 100000$, we cannot perform the DC-NICQR$_9$ method on the full dataset because of computer memory limitations. (The same situation also occurred in Chen and Xie (2014), who considered the CQR$_9$ method for the linear model.) However, we can obtain estimators by using the divide-and-conquer procedure with $M = 10$ (see Table 5).



(a) Mean computing time of γ

(b) Mean computing time of $g(\cdot)$

Figure 3: The curve is observation and the dots are the estimated values.

Table 6 The coefficient estimates and MSE for the airline on-time data.

Mthod	HD	DIS	NF	WF	MSE
LS	0.0044	-0.0505	-0.0004	-0.0451	0.0836
DC-NICQR$_9$	0.0608	-0.8030	0.0377	-0.5917	0.0831

4.2.2. Real 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 by Schifano et al. (2016). 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. (2016) developed a linear model that fit the data as follows:

$$AD = \gamma_1 HD + \gamma_2 DIS + \gamma_3 NF + \gamma_4 WF + \varepsilon, \tag{4.4}$$

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:

$$AD = g\{\gamma_1 HD + \gamma_2 DIS + \gamma_3 NF + \gamma_4 WF\} + \varepsilon. \tag{4.5}$$

For comparison purposes, we use the least squares method proposed by Draper and Smith (1998) to estimate $(\gamma_1, \gamma_2, \gamma_3, \gamma_4)^\top$ in model (4.4), and use the DC-NICQR$_9$ method proposed in Section 3 to estimate $(\gamma_1, \gamma_2, \gamma_3, \gamma_4)^\top$ in model (4.5). 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. Figure 4 illustrates the estimated AD along with the data, showing that the DC-NICQR$_9$ method performs well with smaller MSE.

5. Conclusion

We proposed an NICQR method to deal with massive datasets. The NICQR method is a non-iterative estimation algorithm that allows us to analyze massive datasets more quickly. Specifically, we used the divide-and-conquer

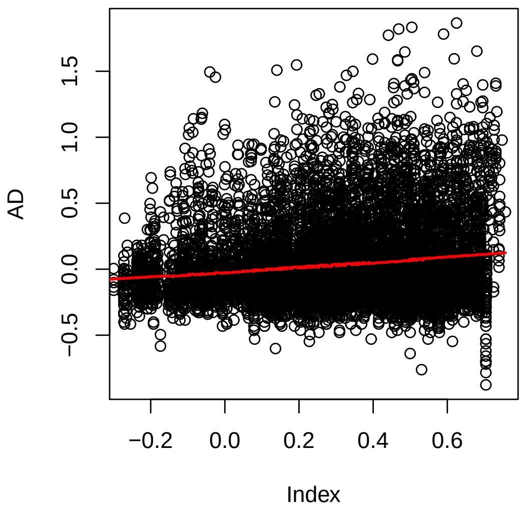


Figure 4: DC-NICQR$_9$ for airline on-time data for model (4.4). The dots are the observations *AD* and the curve is the estimated *AD*.

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.

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 Zhao et al., 2015 and Chen et al., 2019).

Acknowledgments

The three anonymous referees provided numerous valuable comments that improve the manuscript. This work was supported by the National Natural Science Foundation of China (No. 11801069) and the Shanghai Sailing Program (No. 17YF1400800).

Appendix

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 $\mathbf{U} = \mathbf{X}^\top\gamma$ is positive and uniformly continuous for γ in a neighborhood of γ_0. Further the density of $\mathbf{X}^\top\gamma_0$ is continuous and bounded away from 0 and ∞ on its support.
C3. The function $g_0(\cdot)$ has a continuous and bounded second derivative.
C4. Assume that the model error ε has a positive density $f(\cdot)$.
Remark: Conditions **C1-C4** are standard conditions, which are commonly used in single-index regression model, see Jiang et al. (2016b) and Christou and Akritas (2016).
Lemma 1. Let $(X_1, Y_1), \cdots, (X_n, Y_n)$ 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

$n^{2\varepsilon-1}h \to \infty$ for some $\varepsilon < 1 - s^{-1}$, then

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

Proof. This follows immediately from the result obtained by Mack and Silverman (1982).

Lemma 2. Suppose that conditions **C1-C2** given in the Appendix hold, and $nh_{\max}^4 \to 0$, then for any γ

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

Proof. Write

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

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

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

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

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

where

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

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

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. (2017). Under the condition **C2** and $nh^4_{\max} = o(1)$, $T_{12} = o_p(1)$. Thus, $T_1 = o_p(1)$. Next, to show that $T_2 = o_p(1)$.

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

By the U-statistics techniques and condition $nh^4_{\max} = o(1)$, we can proof that $T_2 = o_p(1)$.
Proof of Theorem 2.1. Set $\hat{\gamma}^* = \sqrt{n}(\hat{\gamma}-\gamma_0)$ and $\gamma^* = \sqrt{n}(\gamma-\gamma_0)$. Then, $\hat{\gamma}^*$ is also the minimizer of

$$L_n(\gamma^*) = \sum_{i=1}^{n}\sum_{k=1}^{K}\left\{\rho_{\tau_k}\left(Y^*_{i,\tau_k}-\tilde{H}_{\tau_k}(\mathbf{X}_i\mid\gamma^*/\sqrt{n}+\gamma_0)\right)-\rho_{\tau_k}\left(Y^*_{i,\tau_k}\right)\right\},$$

where $Y^*_{i,\tau_k} = Y_i - \hat{H}_{\tau_k}(\mathbf{X}_i^\top\gamma_0\mid\gamma_0)+O_p(n^{-1})$, and $\tilde{H}_{\tau_k}(\mathbf{X}_i\mid\gamma^*/\sqrt{n}+\gamma_0) = \hat{H}_{\tau_k}(\mathbf{X}_i^\top\gamma\mid\gamma)-\hat{H}_{\tau_k}(\mathbf{X}_i^\top\gamma_0\mid\gamma_0)$. Write $L_n(\gamma^*)$ as

$$L_n(\gamma^*) = E\left[L_n(\gamma^*)\mid\mathbf{X}\right]-\sum_{i=1}^{n}\sum_{k=1}^{K}\left\{\rho'_{\tau_k}\left(Y^*_{i,\tau_k}\right)-E\left[\rho'_{\tau_k}\left(Y^*_{i,\tau_k}\right)\mid\mathbf{X}\right]\right\}\tilde{H}_{\tau_k}(\mathbf{X}_i\mid\gamma^*/\sqrt{n}+\gamma_0)+R_n(\gamma^*),$$

where $R_n(\gamma^*)$ is the remainder term, and to save space, we obtain $R_n(\gamma^*) = o_p(1)$ by similar idea of proof in Fan et al. (1994). Note that

$$\begin{aligned}
E\left[L_n(\gamma^*)\mid\mathbf{X}\right] =&\sum_{i=1}^{n}\sum_{k=1}^{K}E\left[\rho_{\tau_k}\left(Y^*_{i,\tau_k}-\tilde{H}_{\tau_k}(\mathbf{X}_i\mid\gamma^*/\sqrt{n}+\gamma_0)\right)-\rho_{\tau_k}\left(Y^*_{i,\tau_k}\right)\mid\mathbf{X}\right]\\
=&-\sum_{i=1}^{n}\sum_{k=1}^{K}E\left[\rho'_{\tau_k}\left(Y^*_{i,\tau_k}\right)\mid\mathbf{X}\right]\tilde{H}_{\tau_k}(\mathbf{X}_i\mid\gamma^*/\sqrt{n}+\gamma_0)\\
&+\frac{1}{2}\sum_{i=1}^{n}\sum_{k=1}^{K}E\left[\rho''_{\tau_k}\left(Y^*_{i,\tau_k}\right)\mid\mathbf{X}\right]\tilde{H}^2_{\tau_k}(\mathbf{X}_i\mid\gamma^*/\sqrt{n}+\gamma_0)+o_p(1),
\end{aligned}$$

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

$$
\begin{aligned}
L_n(\gamma^*) =& -\sum_{k=1}^{K}\sum_{i=1}^{n}\rho'_{\tau_k}\left(Y^*_{i,\tau_k}\right)\tilde{H}_{\tau_k}(\mathbf{X}_i \mid \gamma^*/\sqrt{n}+\gamma_0) + \frac{1}{2}\sum_{k=1}^{K} f_\varepsilon(c_k)\sum_{i=1}^{n}\tilde{H}^2_{\tau_k}(\mathbf{X}_i \mid \gamma^*/\sqrt{n}+\gamma_0) + o_p(1)\\
=& -\sum_{k=1}^{K}\sum_{i=1}^{n}\rho'_{\tau_k}\left(Y^*_{i,\tau_k}\right)\left[\hat{H}_{\tau_k}\left(\mathbf{X}_i^\top(\gamma^*/\sqrt{n}+\gamma_0) \mid \gamma^*/\sqrt{n}+\gamma_0\right) - \hat{H}_{\tau_k}\left(\mathbf{X}_i^\top\gamma_0 \mid \gamma_0\right)\right]\\
&+ \frac{1}{2}\sum_{k=1}^{K} f_\varepsilon(c_k)\sum_{i=1}^{n}\left[\hat{H}_{\tau_k}\left(\mathbf{X}_i^\top(\gamma^*/\sqrt{n}+\gamma_0) \mid \gamma^*/\sqrt{n}+\gamma_0\right) - \hat{H}_{\tau_k}\left(\mathbf{X}_i^\top\gamma_0 \mid \gamma_0\right)\right]^2 + o_p(1).
\end{aligned}
$$

By Lemma 2, we can obtain

$$
\begin{aligned}
&\sum_{i=1}^{n}\left[\hat{H}_{\tau_k}\left(\mathbf{X}_i^\top(\gamma^*/\sqrt{n}+\gamma_0) \mid \gamma^*/\sqrt{n}+\gamma_0\right) - \hat{H}_{\tau_k}\left(\mathbf{X}_i^\top\gamma_0 \mid \gamma_0\right)\right]\\
=&\sum_{i=1}^{n}\left[H_{\tau_k}\left(\mathbf{X}_i^\top(\gamma^*/\sqrt{n}+\gamma_0) \mid \gamma^*/\sqrt{n}+\gamma_0\right) - H_{\tau_k}\left(\mathbf{X}_i^\top\gamma_0 \mid \gamma_0\right)\right] + o_p(n^{-1/2}),
\end{aligned}
$$

and

$$
\begin{aligned}
&H_{\tau_k}\left(\mathbf{X}_i^\top(\gamma^*/\sqrt{n}+\gamma_0) \mid \gamma^*/\sqrt{n}+\gamma_0\right) - H_{\tau_k}\left(\mathbf{X}_i^\top\gamma_0 \mid \gamma_0\right)\\
=&\frac{\gamma^*}{\sqrt{n}}\frac{\partial H_{\tau_k}(\mathbf{X}_i^\top\gamma \mid \gamma)}{\partial\gamma}\bigg|_{\gamma_0} + O_p(n^{-1}) = \frac{1}{\sqrt{n}}g'(\mathbf{X}_i^\top\gamma_0)\left(\mathbf{X}_i - E[\mathbf{X} \mid \mathbf{X}^\top\gamma_0]\right)^\top\gamma^* + O_p(n^{-1}).
\end{aligned}
$$

Thus,

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

where

$$
\begin{aligned}
W_n =&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{k=1}^{K}\rho'_{\tau_k}(Y^*_{i,\tau_k})g'(\mathbf{X}_i^\top\gamma_0)\left(\mathbf{X}_i - E[\mathbf{X} \mid \mathbf{X}^\top\gamma_0]\right)^\top,\\
S_n =&\frac{1}{n}\sum_{i=1}^{n}\left\{g'(\mathbf{X}_i^\top\gamma_0)\right\}^2\left(\mathbf{X}_i - E[\mathbf{X} \mid \mathbf{X}^\top\gamma_0]\right)\left(\mathbf{X}_i - E[\mathbf{X} \mid \mathbf{X}^\top\gamma_0]\right)^\top.
\end{aligned}
$$

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

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

It follows by the convexity lemma (see Pollard, 1991) that the quadratic approximation to $L_n(\gamma^*)$ holds uniformly for γ^* in any compact set. Thus, it follows that

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

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

Proof of Theorem 2.2. Note that

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

where $\hat{g}(u \mid \gamma_0)$ is a local linear estimator of $g_0(u)$ when the index coefficient γ_0 is known. For given u, for notational simplicity, we write $\sum_{k=1}^K v_k \hat{a}_{k,\hat{\gamma}} = \hat{g}(u \mid \hat{\gamma})$, $\hat{b}_{\hat{\gamma}} = \hat{g}'(u \mid \hat{\gamma})$, $\sum_{k=1}^K v_k \hat{a}_{k,\gamma_0} = \hat{g}(u \mid \gamma_0)$, and $\hat{b}_{\gamma_0} = \hat{g}'(u \mid \gamma_0)$, which are the solutions of the following minimization problems, respectively,

$$\min_{(a_1,\ldots,a_K,b)} \sum_{k=1}^K \sum_{i=1}^n \rho_{\tau_k} \left\{Y_i - a_k - b(X_i^T\hat{\gamma} - u)\right\} \tilde{K}_h \left(X_i^T\hat{\gamma} - u\right),$$

$$\min_{(a_1,\ldots,a_K,b)} \sum_{k=1}^K \sum_{i=1}^n \rho_{\tau_k} \left\{Y_i - a_k - b(X_i^T\gamma_0 - u)\right\} \tilde{K}_h \left(X_i^T\gamma_0 - u\right).$$

Denote

$$\begin{aligned}
\theta^* &= \sqrt{nh}\left\{a_{1,\hat{\gamma}} - g_0(u) - c_1, \ldots, a_{K,\hat{\gamma}} - g_0(u) - c_K, h[b_{\hat{\gamma}} - g_0'(u)]\right\}^T, \\
\bar{\theta}^* &= \sqrt{nh}\left\{\hat{a}_{1,\hat{\gamma}} - g_0(u) - c_1, \ldots, \hat{a}_{K,\hat{\gamma}} - g_0(u) - c_K, h[\hat{b}_{\hat{\gamma}} - g_0'(u)]\right\}^T, \\
\bar{\theta}^{**} &= \sqrt{nh}\left\{\hat{a}_{1,\gamma_0} - g_0(u) - c_1, \ldots, \hat{a}_{K,\gamma_0} - g_0(u) - c_K, h[\hat{b}_{\gamma_0} - g_0'(u)]\right\}^T, \\
Z_{i,k}^* &= \left\{e_k^T, (X_i^T\hat{\gamma} - u)/h\right\}^T, Z_{i,k}^{**} = \left\{e_k^T, (X_i^T\gamma_0 - u)/h\right\}^T,
\end{aligned}$$

where e_k is a K-vector with 1 on the kth position and 0 elsewhere. Further, write $K_i^* = \tilde{K}_h\left(X_i^T\hat{\gamma} - u\right)$, $K_i^{**} = \tilde{K}_h\left(X_i^T\gamma_0 - u\right)$, $\eta_{i,k} = I(\varepsilon_i \le c_k) - \tau_k$, $\eta_{i,k}^*(u) = I\{\varepsilon_i \le c_k - d_{i,k}^*\} - \tau_k$, $\eta_{i,k}^{**}(u) = I\{\varepsilon_i \le c_k - d_{i,k}^{**}\} - \tau_k$, where $d_{i,k}^*(u) = c_k + g_0(X_i^T\gamma_0) - g_0(u) - g_0'(u)(X_i^T\hat{\gamma} - u)$ and $d_{i,k}^{**}(u) = c_k + g_0(X_i^T\gamma_0) - g_0(u) - g_0'(u)(X_i^T\gamma_0 - u)$. Thus, $Y_i - a_k - b(X_i^T\hat{\gamma} - u) = \varepsilon_i - c_k + d_{i,k}^* - \Delta_{i,k}$, where $\Delta_{i,k} = \{Z_{i,k}^*\}^T\theta^*/\sqrt{nh}$. Then, $\bar{\theta}^*$ is also the minimizer of

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

By applying the identity (Knight, 1998)

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

we have

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

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

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

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

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

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

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

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

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

By applying the convexity lemma (Pollard, 1991) and the quadratic approximation lemma (Fan and Gijbels, 1996), the minimizer of $L_n^*(\theta^*)$ can be expressed as

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

$\bar\theta^{**}$ can be shown similarly as

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

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

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

Thus,

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

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

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

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

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

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

Proof of Theorem 3.1. By the similar proof of Theorem 2.1,

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

Moreover, from the proof of Theorem 2.1, for each subsets $n_m \to \infty$, $m = 1, \ldots, M$, we have

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

Thus, we can obtain

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

By the Cram$\acute{e}$r-Wald theorem and the Central Limit Theorem, the Theorem 3.1 can be proof.

Proof of Theorem 3.2. From the proof of Theorem 2.2, for each subsets $n_m \to \infty$, $m = 1, \ldots, M$, we have

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

Thus, we can obtain

$$\sqrt{nh}\left\{\hat{g}^{DC}(u \mid \hat{\gamma}^{DC}) - g_0(u)\right\} = \sqrt{nh}\left(\frac{1}{n}\sum_{m=1}^{M} n_m \hat{g}_m^{DC}(u \mid \hat{\gamma}^{DC}) - g_0(u)\right)$$

$$= -\frac{f_{\mathbf{U}_0}^{-1}(u)}{\sqrt{nh}}\sum_{k=1}^{K}\sum_{i=1}^{n} v_k f^{-1}(c_k)\left[I\left(\varepsilon_i \leq c_k - d_{i,k}^{**}\right) - \tau_k\right] K_i^{**} + o_p(1).$$

Thus, the Theorem 3.2 can be proof.

References

Chen, X. and Xie, M. (2014). A split-and-conquer approach for analysis of extraordinarily large data. *Statistica Sinica* **24** 1655–1684.

Chen, X., Liu, W. and Zhang, Y. (2019). Quantile regression under memory constraint. *The Annals of Statistics* **6** 3244–3273.

Christou, E. and Akritas, M. G. (2016). Single index quantile regression for heteroscedastic data. *Journal of Multivariate Analysis* **150** 169–182.

Draper, N. R. and Smith, H. (1998). *Applied regression analysis*. NewYork: John Wiley & Sons.

Fan, J., Hu, T. C. and Truong Y. K. (1994). Robust nonparametric function estimation. *Scandinavian Journal of Statistics* **21** 433–446.

Fan, J. and Gijbels, I. (1996). *Local polynomial modelling and its applications*. Chapman and Hall, London.

Feng, L., Zou, C. L. and Wang, Z. J. (2012). Rank-based inference for the single-index model. *Statistics and Probability Letters* **82** 535–541.

Härdle, W., Hall, P. and Ichimura, H. (1993). Optimal smoothing in single-index models. *The Annals of Statistics* **21** 157–178.

Jiang, R., Zhou, Z. G., Qian, W. M. and Shao, W. Q. (2012). Single-index composite quantile regression. *Journal of the Korean Statistical Society* **3** 323–332.

Jiang, R., Qian, W. M. and Zhou, Z. G. (2016a). Weighted composite quantile regression for single-index models. *Journal of Multivariate Analysis* **148** 34–48.

Jiang, R., Qian, W. M. and Zhou, Z. G. (2016b). Single-index composite quantile regression with heteroscedasticity and general error distributions. *Statistical Papers* **57** 185–203.

Kong, E. and Xia, Y. (2018). On the efficiency of online approach to nonparametric smoothing of big data. *Statistica Sinica* DOI: 10.5705/ss.202015.0365.

Knight, K. (1998). Limiting distributions for L_1 regression estimators under general conditions. *The Annals of Statistics* **26** 755–770.

Kraus, D. and Czado, C. (2017). D-vine copula based quantile regression. *Computational Statistics and Data Analysis* **110** 1–18.

Lee, J. D., Liu, Q., Sun, Y. and Taylor, J. E. (2017). Communication-efficient sparse regression. *Journal of Machine Learning Research* **18** 1–30.

Li, R., Lin, D. and Li, B. (2013). Statistics inference in massive data sets. *Applied Stochastic Models in Business and Industry* **29** 399–409.

Liang, H., Liu, X., Li, R. Z. and Tsai, C. L. (2010). Estimation and testing for partially linear single-index models. *The Annals of Statistics* **6** 3811–3836.

Lin, W. and Kulasekera, K. B. (2007). Identifiability of single-index models and additive-index models. *Biometrika* **94** 496–501.

Lin, N. and Xi, R. (2011). Aggregated estimating equation estimation. *Statistics and Its Interface* **4** 73–83.

Lin, D. and Zeng, D. (2010). On the relative efficiency of using summary statistics versus individual-level data in meta-analysis. *Biometrika* **97** 321–332.

Liu, J., Zhang, R., Zhao, W. and Lv, Y. (2013). A robust and efficient estimation method for single index models. *Journal of Multivariate Analysis* **122** 226–238.

Liu, D. G., Liu R. Y. and Xie, M. G. (2015). Multivariate meta-analysis of heterogeneous studies using only summary statistics: efficiency and robustness. *Journal of the American Statistical Association* **110** 326–340.

Lu, J., Cheng, G. and Liu. H. (2016). Nonparametric heterogeneity testing for massive data. *arXiv:1601.06212*.

Ma, S. J. and He, X. M. (2016). Inference for single-index quantile regression models with profile optimization. *The Annals of Statistics* **4** 1234–1268.

Mack, Y. and Silverman, B. (1982). Weak and strong uniform consistency of kernel regression estimates. *Z.Wahrscheinlichkeitstheori Verw. Geb.* **61** 405–415.

Pollard, D. (1991). Asymptotics for least absolute deviation regression estimators. *Econometric Theory* **7** 186–199.

Rémillard, B., Nasri, B. and Bouezmarni, T. (2017). On copula-based conditional quantile estimators. *Statistics and Probability Letters* **128** 14–20.

Silverman, B. W. (1986). *Density Estimation*. Chapman and Hall, London.

Schifano, E. D., Wu, J., Wang, C., Yan, J. and Chen, M. H. (2016). Online updating of statistical inference in the big data setting. *Technometrics* **58** 393–403.

Tang, L., Zhou, L. and Song, X. K. (2016). Method of divide-and-combine in regularised generalised linear models for big data. *arXiv:1611.06208*.

Tang, Y. L., Wang, H. and Liang, H. (2018). Composite estimation for single-index models with responses subject to detection limits. *Scandinavian Journal of Statistics* **45** 444–464.

Tian, Y., Zhu, Q. and Tian, M. (2016). Estimation of linear composite quantile regression using EM algorithm. *Statistics and Probability Letters* **117** 183–191.

Wang, J. L., Xue, L. G., Zhu, L. X. and Chong, Y. S. (2010). Estimation for a partial-linear single-index model. *The Annals of Statistics* **1** 246–274.

Wang, Q. and Wu, R. (2013). Shrinkage estimation of partially linear single-index models. *Statistics and Probability Letters* **83** 2324–2331.

Wu, T. Z., Yu, K. and Yu, Y. (2010). Single-index quantile regression. *Journal of Multivariate analysis* **101** 1607–1621.

Xia, Y. and Härdle, W. (2006). Semi-parametric estimation of partially linear single-index models. *Journal of Multivariate Analysis* **97** 1162–1184.

Xia, Y., Tong, H., Li, W. K. and Zhu, L. (2002). An adaptive estimation of dimension reduction space. *Journal of the Royal Statistical Society: Series B* **64** 363–410.

Xie, M., Singh, K. and Strawderman, W. E. (2012). Confidence distributions and a unifying framework for meta-analysis. *Journal of the American Statistical Association* **106** 320–333.

Yu, Y. and Ruppert, D. (2002). Penalized spline estimation for partially linear single-index models. *Journal of the American Statistical Association* **97** 1042–1054.

Zhao, T., Kolar, M. and Liu, H. (2015). A general framework for robust testing and confidence regions in high-dimensional quantile regression. *arXiv:1412.8724*.

Zhao, Y., Zou, C. and Wang, Z. (2017). An adaptive lack of fit test for big data. *Statistical Theory and Related Fields* **1** 59–68.

Zou, H. and Yuan, M. (2008). Composite quantile regression and the oracle model selection theory. *The Annals of Statistics* **36** 1108–1126.