Journal of Multivariate Analysis 101 (2010) 1607–1621

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

Single-index quantile regression[☆]

Tracy Z. Wu [a], Keming Yu [b], Yan Yu [c,*]

[a] Consumer Risk Modeling and Analytics Group at JPMorgan Chase Bank, 1111 Polaris Parkway, Columbus, OH 43240, United States
[b] Brunel University, Uxbridge, Middlesex UB8 3PH, London, UK
[c] University of Cincinnati, PO BOX 210130, Cincinnati, OH 45221, United States

ARTICLE INFO

Article history:
Received 25 March 2009
Available online 19 February 2010

AMS subject classifications:
62G08
62G20
62G35

Keywords:
Conditional quantile
Dimension reduction
Local polynomial smoothing
Nonparametric model
Semiparametric model

ABSTRACT

Nonparametric quantile regression with multivariate covariates is a difficult estimation problem due to the "curse of dimensionality". To reduce the dimensionality while still retaining the flexibility of a nonparametric model, we propose modeling the conditional quantile by a single-index function $g_0(\mathbf{x}^\top\boldsymbol{\gamma}_0)$, where a univariate link function $g_0(\cdot)$ is applied to a linear combination of covariates $\mathbf{x}^\top\boldsymbol{\gamma}_0$, often called the single-index. We introduce a practical algorithm where the unknown link function $g_0(\cdot)$ is estimated by local linear quantile regression and the parametric index is estimated through linear quantile regression. Large sample properties of estimators are studied, which facilitate further inference. Both the modeling and estimation approaches are demonstrated by simulation studies and real data applications.

1. Introduction

In this paper, we introduce single-index quantile regression for nonparametric estimation with multivariate covariates. Given $\tau \in (0, 1)$, we propose a single-index model for the τth conditional quantile $\theta_\tau(\mathbf{x})$ of y given $\mathbf{x}$,

$$\theta_\tau(\mathbf{x}) = g_0(\mathbf{x}^\top\boldsymbol{\gamma}_0), \tag{1}$$

where $\mathbf{x}$ is a vector of d-dimensional covariates, y is a real valued dependent variable, $g_0(\cdot)$ is the unknown univariate link function, and $\boldsymbol{\gamma}_0$ is the unknown single-index vector coefficient satisfying $\|\boldsymbol{\gamma}_0\| = 1$ and the first component $\gamma_1 > 0$ for identifiability [1]. Here $\|\cdot\|$ is the Euclidean norm. Single-index quantile regression model (1) generalizes the seminal work of linear quantile regression of Koenker and Bassett [2] by replacing the linear combination $\mathbf{x}^\top\boldsymbol{\gamma}_0$ with a nonparametric counterpart $g_0(\mathbf{x}^\top\boldsymbol{\gamma}_0)$.

The single-index approach has proven to be an efficient way to cope with high-dimensional nonparametric estimation problems in conditional mean regression (e.g. [3–8]). When used to model conditional quantiles involving multivariate covariates, single-index models inherit the same advantages as in the mean regression context: (i) the unspecified link function allows model flexibility and thus has less risk of mis-specification; (ii) the single-index in the link function projects multivariate covariates onto one-dimensional variate, effectively reducing the dimensionality in nonparametric estimation; (iii) the single-index structure together with the nonlinear link function can model some interactions among the covariates

☆ This manuscript is based in part on the Ph.D. Thesis of the first author at University of Cincinnati. This paper formerly circulated under the title "Local Linear Estimation for Single-index Conditional Quantiles", which was presented in JSM in Seattle, August 2006.

* Corresponding address: Department of Quantitative Analysis Operations Management, University of Cincinnati, PO BOX 210130, Cincinnati, OH 45221, United States.
E-mail address: Yan.Yu@uc.edu (Y. Yu).

0047-259X/$ – see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmva.2010.02.003

implicitly, which is more realistic to real applications; (iv) the interpretation of covariate effects is easy because of the linear structure of the index. Model (1) is very general. Under certain assumptions (e.g. monotonicity) the model can be used to model the quantiles for several important cases such as survival and transformation models and location-scale model, as pointed out in [9]. These cases would be interesting for future research. Indeed, based on our estimation equation (6), Kong and Xia [10] recently investigated the Bahadur representation of single-index parameter estimators.

Research on nonparametric quantile regression is relatively sparse in contrast to that for mean regression. Yu and Jones [11] developed local linear approaches for univariate quantile regression. Besides local estimation approaches, spline approach is another stream of nonparametric quantile regression. See [12,13] for more details. Stone [14] and Chaudhuri [15] considered fully nonparametric quantile regression in a general multivariate setting. They are flexible but usually unattractive in practice due to the well-known "curse-of-dimensionality".

Recently, dimension-reduction techniques for nonparametric quantile regression models have attracted a lot of attention in the literature. These include additive models and partially linear models (e.g. [16–18]). But these models do not incorporate interactions, nor nest with single-index models. A close alternative to our approach is the important work of average derivative models (e.g. Chaudhuri et al. [9] for quantile regression; Härdle and Stoker [19] for mean regression). They estimate the single-index vector by taking an expectation of the vector of partial derivatives of the conditional quantile with respect to the covariates $\mathbf{x}$. Though theoretically appealing, it requires relevant multi-dimensional quantiles to be obtained nonparametrically before the index can be estimated directly.

In this article, we present an overall treatment of estimation and inference along with an application using the proposed single-index quantile regression model (1). To the best of our knowledge, this appears to be the first paper in the scientific literature that presents a comprehensive study of single-index quantile regression. In practice, we introduce an algorithm that is particularly tailored to model (1), which is based on the local linear approach to estimate the nonparametric part $g_0(\cdot)$, and linear quantile regression to solve the parametric index part $\boldsymbol{\gamma}_0$. With this algorithm, single-index models can be estimated quite expediently as shown in both simulation study and real data applications. It is natural to adopt the local linear approach as in Yu and Jones[11] for univariate conditional quantiles. Not surprisingly, Yu and Jones [20] also find that local linear method is more advantageous over local constant approaches in quantile regression because of their lower bias and nicer boundary performances. In theory, we obtain the asymptotic properties for the proposed nonparametric estimator $\hat{g}(\cdot)$, conditional quantile estimator $\hat{\theta}_\tau(\mathbf{x})$, and the parametric single-index vector estimate $\hat{\boldsymbol{\gamma}}$, which naturally facilitate further inference. Confidence intervals for conditional quantiles are also readily available. In addition, we derive an approximate simple-to-calculate rule-of-thumb bandwidth selector based on the large sample theory. Optimal bandwidth by minimizing the asymptotic mean squared error for single-index quantile regression might be otherwise computationally intensive, particularly when a number of quantiles need to be estimated. A Monte Carlo simulation study and an application to Boston Housing data show promises of our proposed approach.

The rest of the paper is organized as follows. In Section 2, we propose a local linear estimation method as well as some computational algorithms. Asymptotic properties of the estimators are obtained in Section 3. We present both simulation examples and real data applications in Section 4. Section 5 concludes the paper. All the proofs are relegated to the Appendix.

2. The model and estimation

2.1. The model and local linear estimation

For our single-index quantile regression model (1), note $g_0(\cdot)$ should really be $g_{0,\tau}(\cdot)$ and $\boldsymbol{\gamma}_0$ should be $\boldsymbol{\gamma}_{0,\tau}$, both unique to the given quantile. We omit the subscript τ for notational convenience. The linear combination of the covariates $\mathbf{x}^{\mathsf{T}}\boldsymbol{\gamma}_0$ is often called the single-index. Mathematically, the true parameter vector $\boldsymbol{\gamma}_0$ solves the following minimization problem:

$$\boldsymbol{\gamma}_0 = \arg\min_{\boldsymbol{\gamma}} E\left[\rho_\tau\left(y - g(\mathbf{x}^{\mathsf{T}}\boldsymbol{\gamma})\right)\right] \text{ subject to } \|\boldsymbol{\gamma}\| = 1, \quad \gamma_1 > 0, \tag{2}$$

where the loss function (also called the "check" function) $\rho_\tau(u) = |u| + (2\tau - 1)u$ and $g(\cdot)$ is the unknown link function. The right-hand side of the above equation is the expected loss which can be equivalently written as

$$E\left[\rho_\tau\left(y - g(\mathbf{x}^{\mathsf{T}}\boldsymbol{\gamma})\right)\right] = E\left\{E\left[\rho_\tau\left(y - g(\mathbf{x}^{\mathsf{T}}\boldsymbol{\gamma})\right)|\mathbf{x}^{\mathsf{T}}\boldsymbol{\gamma}\right]\right\}, \tag{3}$$

where $E\left[\rho_\tau\left(y - g(\mathbf{x}^{\mathsf{T}}\boldsymbol{\gamma})\right)|\mathbf{x}^{\mathsf{T}}\boldsymbol{\gamma}\right]$ is the conditional expected loss and $g(\cdot)$ is the τth conditional quantile function.

Let $\{\mathbf{x}_i, y_i\}_{i=1}^n$ be an independent identically distributed (i.i.d.) sample from $(\mathbf{x}, y)$. For $\mathbf{x}_i^{\mathsf{T}}\boldsymbol{\gamma}$ "close" to u, the τth conditional quantile at $\mathbf{x}_i^{\mathsf{T}}\boldsymbol{\gamma}$ can be approximated linearly by

$$g(\mathbf{x}_i^{\mathsf{T}}\boldsymbol{\gamma}) \approx g(u) + g'(u)(\mathbf{x}_i^{\mathsf{T}}\boldsymbol{\gamma} - u) = a + b(\mathbf{x}_i^{\mathsf{T}}\boldsymbol{\gamma} - u), \tag{4}$$

where $a \overset{def}{\equiv} g(u)$ and $b \overset{def}{\equiv} g'(u)$. Following (4), we minimize the following local linear sample analogue of $L_{\boldsymbol{\gamma}}(u)$ [11] with respect to (a, b) to obtain $\hat{g}(u) = \hat{a}$,

$$\sum_{i=1}^n \rho_\tau\left(y_i - a - b(\mathbf{x}_i^{\mathsf{T}}\boldsymbol{\gamma} - u)\right) K\left(\frac{\mathbf{x}_i^{\mathsf{T}}\boldsymbol{\gamma} - u}{h}\right), \tag{5}$$

where $K(\cdot)$ is the kernel weight function and h is the bandwidth.

We average (5) over u and obtain the sample analog of (3), i.e. the objective function that is used to estimate model (1),

$$\sum_{j=1}^{n}\sum_{i=1}^{n}\rho_\tau\left(y_i-a_j-b_j(\mathbf{x}_i^\top\boldsymbol{\gamma}-\mathbf{x}_j^\top\boldsymbol{\gamma})\right)\frac{K_h(\mathbf{x}_i^\top\boldsymbol{\gamma}-\mathbf{x}_j^\top\boldsymbol{\gamma})}{\sum_{l=1}^{n}K_h(\mathbf{x}_l^\top\boldsymbol{\gamma}-\mathbf{x}_j^\top\boldsymbol{\gamma})},\tag{6}$$

where $K_h(\cdot)=K(\cdot/h)/h$. In practice, minimization of (6) is done by iteratively solving two simple problems, one with respect to a_j's and b_j's, and the other with respect to $\boldsymbol{\gamma}$.

We rewrite (6) as

$$\sum_{j=1}^{n}\sum_{i=1}^{n}\rho_\tau\left(y_i-a_j-b_j(\mathbf{x}_i^\top\boldsymbol{\gamma}-\mathbf{x}_j^\top\boldsymbol{\gamma})\right)\omega_{ij},\tag{7}$$

where $\omega_{ij}=\frac{K_h(\mathbf{x}_i^\top\boldsymbol{\gamma}-\mathbf{x}_j^\top\boldsymbol{\gamma})}{\sum_{l=1}^{n}K_h(\mathbf{x}_l^\top\boldsymbol{\gamma}-\mathbf{x}_j^\top\boldsymbol{\gamma})}$. We decompose (7) into:

P1. Given $\boldsymbol{\gamma}$,

$$(\hat{a}_j,\hat{b}_j)_{j=1}^n=\arg\min_{(a_j,b_j)_{j=1}^n}\sum_{j=1}^{n}\sum_{i=1}^{n}\rho_\tau\left(y_i-a_j-b_j(\mathbf{x}_i^\top\boldsymbol{\gamma}-\mathbf{x}_j^\top\boldsymbol{\gamma})\right)\omega_{ij}.$$

So for any $j\in\{1,2,\ldots,n\}$,

$$(\hat{a}_j,\hat{b}_j)=\arg\min_{(a_j,b_j)}\sum_{i=1}^{n}\rho_\tau\left(y_i-a_j-b_j(\mathbf{x}_i^\top\boldsymbol{\gamma}-\mathbf{x}_j^\top\boldsymbol{\gamma})\right)\omega_{ij}.$$

P2. Given a_j's and b_j's,

$$\begin{aligned}\hat{\boldsymbol{\gamma}}&=\arg\min_{\boldsymbol{\gamma}}\sum_{j=1}^{n}\sum_{i=1}^{n}\rho_\tau\left(y_i-a_j-b_j(\mathbf{x}_i-\mathbf{x}_j)^\top\boldsymbol{\gamma}\right)\omega_{ij}\\&=\arg\min_{\boldsymbol{\gamma}}\sum_{j=1}^{n}\sum_{i=1}^{n}\rho_\tau\left(y_{ij}^*-\mathbf{x}_{ij}^{*\top}\boldsymbol{\gamma}\right)\omega_{ij}^*,\end{aligned}$$

where $y_{ij}^*=y_i-a_j$, $\mathbf{x}_{ij}^*=b_j(\mathbf{x}_i-\mathbf{x}_j)$, and $\omega_{ij}^*=\omega_{ij}$ evaluated at the current estimate of $\boldsymbol{\gamma}$, $i,j=1,\ldots,n$.

The subproblem P1 deals with estimating (a_j,b_j), $j=1,\ldots,n$, as if $\boldsymbol{\gamma}$ is known. While in P2, $\boldsymbol{\gamma}$ is estimated through usual linear quantile regression without intercept (*regression-through-origin*) on n^2 "observations" $\{y_{ij}^*,\mathbf{x}_{ij}^*\}_{i,j=1}^n$ with known weights $\{\omega_{ij}^*\}_{i,j=1}^n$ evaluated at the estimate of $\boldsymbol{\gamma}$ from the previous iteration.

An algorithm for estimating $\boldsymbol{\gamma}$ is as follows.

Step 0. Obtain initial $\hat{\boldsymbol{\gamma}}^{(0)}$ from average derivative estimate (ADE) of Chaudhuri et al. [9]. Standardize the initial estimate such that $\|\hat{\boldsymbol{\gamma}}\|=1$ and $\hat{\gamma}_1>0$ (Initialization step).

Step 1. Given $\hat{\boldsymbol{\gamma}}$, obtain $\{\hat{a}_j,\hat{b}_j\}_{j=1}^n$ by solving a series of the following

$$\min_{(a_j,b_j)}\sum_{i=1}^{n}\rho_\tau\left(y_i-a_j-b_j(\mathbf{x}_i-\mathbf{x}_j)^\top\hat{\boldsymbol{\gamma}}\right)\omega_{ij},\tag{8}$$

with the bandwidth h chosen optimally.

Step 2. Given $\{\hat{a}_j,\hat{b}_j\}_{j=1}^n$, obtain $\hat{\boldsymbol{\gamma}}$ by solving

$$\min_{\boldsymbol{\gamma}}\sum_{j=1}^{n}\sum_{i=1}^{n}\rho_\tau\left(y_i-\hat{a}_j-\hat{b}_j(\mathbf{x}_i-\mathbf{x}_j)^\top\boldsymbol{\gamma}\right)\omega_{ij},\tag{9}$$

with ω_{ij} evaluated at $\boldsymbol{\gamma}$ and h from step 1.

Step 3. Repeat Steps 1 and 2 until convergence.

In the above algorithm, $\hat{\boldsymbol{\gamma}}$ is standardized this way: $\boldsymbol{\gamma}=\mathrm{sign}_1\boldsymbol{\gamma}/\|\boldsymbol{\gamma}\|$, where sign_1 is the sign of the first component of $\boldsymbol{\gamma}$. Only a starting value for $\hat{\boldsymbol{\gamma}}$ is needed. The initial estimator $\hat{\boldsymbol{\gamma}}^{(0)}$ from average derivative estimate (ADE) of Chaudhuri et al. [9] has nice properties and has been proven to be root-n consistent. However, multi-dimensional kernel estimation is involved, which may be computationally intensive. Alternatively, in practice, one may obtain initial $\hat{\boldsymbol{\gamma}}^{(0)}$ from $\min_{a,\boldsymbol{\gamma}}\sum_{i=1}^{n}\rho_\tau(y_i-a-\mathbf{x}_i^\top\boldsymbol{\gamma})$, where a is the quantile regression intercept. A similar algorithm was introduced by Xia et al. [21] in the mean regression context.

Finally, we estimate $g(\cdot)$ at any u by $\hat{g}(\cdot; h, \hat{\boldsymbol{\gamma}}) = \hat{a}$ where

$$(\hat{a}, \hat{b}) = \arg\min_{(a,b)} \sum_{i=1}^{n} \rho_\tau \left(y_i - a - b(\mathbf{x}_i^\top \hat{\boldsymbol{\gamma}} - u)\right) K_h(\mathbf{x}_i^\top \hat{\boldsymbol{\gamma}} - u). \tag{10}$$

A program written in the `R environment` which executes the algorithm is downloadable from the following link http://statqa.cba.uc.edu/~yuy/SINDEXQ.rar.

2.2. Selection of bandwidth

Bandwidth selection is always crucial in local smoothing as it governs the curvature of the fitted function. Theoretically, when the sample size is large, the optimal bandwidth could be derived by minimizing the asymptotic mean squared error (AMSE) from Theorem 1 in Section 3.1. However, the optimal bandwidth can not be calculated directly due to several unknown quantities. Implementation is also computationally expensive, particularly when we would like to estimate several quantiles. Yu and Jones [11] derived an approximate optimal bandwidth under moderate assumptions. In fact, by noting that the similarities between our AMSE and the expression given by Yu and Jones [11] and following the same argument, we obtain the following rule-of-thumb bandwidth h_τ:

$$h_\tau = h_m \left\{\tau(1-\tau)/\phi\left(\Phi^{-1}(\tau)\right)^2\right\}^{1/5}, \tag{11}$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ are the probability density function and the cumulative distribution function of the standard normal distribution respectively. The bandwidth h_τ has a nice property of relating to the optimal bandwidth $h_m = \left\{\frac{[\int K^2(v)dv][var(y|\mathbf{x}^\top\boldsymbol{\gamma}=u)]}{n[\int v^2K(v)dv]^2[\frac{d^2}{du^2}E(y|\mathbf{x}^\top\boldsymbol{\gamma}=u)]^2[f_{U_0}(u)]}\right\}^{1/5}$ used in mean regression via a multiplying factor involving only τ. Since there are many existing algorithms for h_m (see e.g. [22]), h_τ is also readily available.

The approximation by (11) provides a computationally easy way to calculate the otherwise difficult-to-obtain optimal bandwidth for quantile regression. However, the approximation is based on several key assumptions, notably among which are that the curvatures of the conditional median and the conditional mean are similar and that the conditional density function of the dependent variable can be approximated by normal densities. Details regarding the approximation can be found in [11,17].

3. Large sample properties

3.1. Asymptotics for nonparametric part

This section aims to derive the distribution theory for the nonparametric estimator $\hat{g}(\cdot)$. This requires that $\boldsymbol{\gamma}_0$ is either given or estimated with reasonable accuracy. A degenerate case of scalar $\boldsymbol{\gamma}_0$ has been addressed in [23,11].

We assume that the parametric part $\boldsymbol{\gamma}$ can be estimated to the order $O_p(n^{-1/2})$. Indeed, the root-n consistent average derivative quantile estimator (ADE, [9]) is used as "pilot" estimator $\hat{\boldsymbol{\gamma}}^{(0)}$, though initial multi-dimensional smoothing is involved. The root-n neighborhood assumption is common in single-index mean regression literatures, see e.g. [24,6]. The final estimator of the nonparametric part is obtained by minimizing (10) and can be as efficient as the case when $\boldsymbol{\gamma}_0$ is known.

Assumptions

(i) The kernel $K(\cdot) \geq 0$ has a compact support and its first derivative is bounded. It satisfies $\int_{-\infty}^{\infty} K(z)dz = 1$, $\int_{-\infty}^{\infty} zK(z)dz = 0$, $\int_{-\infty}^{\infty} z^2K(z)dz < \infty$, and $|\int_{-\infty}^{\infty} z^jK^2(z)dz| < \infty, j = 0, 1, 2$.

(ii) The density function of $\mathbf{x}^\top\boldsymbol{\gamma}$ 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 ∞ on its support.

(iii) The conditional density function of y given u, $f_y(y|u)$ is continuous in u for each y. Moreover, there exist positive constants ϵ and δ and a positive function $G(y|u)$ such that

$$\sup_{|u_n-u|\leq\epsilon} f_y(y|u_n) \leq G(y|u) \quad \text{and that} \int |\rho_\tau'(y-g_0(u))|^{2+\delta}G(y|u)d\mu(y) < \infty, \text{ and}$$

$$\int (\rho_\tau(y-t) - \rho_\tau(y) - \rho_\tau'(y)t)^2G(y|u)d\mu(y) = o(t^2) \quad \text{as } t \to 0.$$

(iv) The function $g_0(\cdot)$ has a continuous and bounded second derivative.

The assumptions above are commonly used in the literature and are satisfied in many applications. Assumption (i) simply requires that the kernel function is a proper density with finite second moment that is required for the asymptotic variance of estimators; Assumption (ii) guarantees the existence of any ratio terms with the density appearing as part of the denominator; Assumption (iii) is weaker than the Lipschitz continuity of the function $\rho_\tau'(\cdot)$; Assumption (iv) is a common assumption for a link function.

Remark 1. (i) The loss function $\rho_\tau(\cdot)$ is piecewise linear and non-differentiable at 0. We have $\rho'_\tau(y) = 2(\tau - I(y < 0))$ at $y \neq 0$ and may set $\rho'_\tau(y) = 0$ at $y = 0$. In fact, for any continuous random variable y, $y = 0$ occurs with a zero probability.

(ii) Since $\rho'_\tau(\cdot)$ is piecewisely constant with limited number of values, the condition $\int |\rho'_\tau(y - g_0(u))|^{2+\delta} G(y|u) \mathrm{d}\mu(y) < \infty$ in Assumption (iii) can be reduced to $\int G(y|u)\mathrm{d}\mu(y) < \infty$.

Theorem 1. *Under Assumptions* (i)–(iv), *if* $n \to \infty$, $h \to 0$, *and* $nh \to \infty$, *then for an interior point* u,

$$(nh)^{1/2}\{\hat{g}(u; h, \hat{\boldsymbol{\gamma}}) - g_0(u) - \beta(u)h^2\} \xrightarrow{D} N(0, \alpha^2(u)),$$

where $\beta(u) = \frac{g_0''(u)\int v^2 K(v)\mathrm{d}v}{2}$, $\alpha^2(u) = \frac{\int K^2(v)\mathrm{d}v}{f_{U_0}(u)} \frac{\tau(1-\tau)}{[f_y(g_0(u)|u)]^2}$, $f_{U_0}(\cdot)$ *is the density of* $U_0 = \mathbf{x}^\mathsf{T}\boldsymbol{\gamma}_0$ *and* $f_y(\cdot|u)$ *is the conditional density of* y *given* u.

Proof. See Appendix. □

In Theorem 1, we consider the difference between the estimated link function and true link function, both evaluated at the same index value u. However pointwise accuracy is based on the quantity $\hat{\theta}_\tau(\mathbf{x}) - \theta_\tau(\mathbf{x}) = \hat{g}(\mathbf{x}^\mathsf{T}\hat{\boldsymbol{\gamma}}; \hat{\boldsymbol{\gamma}}) - g_0(\mathbf{x}^\mathsf{T}\boldsymbol{\gamma}_0)$. Here both the estimated link function and true link function are evaluated at the same covariate value $\mathbf{x}$, while the quantile estimate takes both index coefficient and link function as estimated. The scaled pointwise error term can be written as $(nh)^{1/2}\{\hat{\theta}_\tau(\mathbf{x}) - \theta_\tau(\mathbf{x})\} = (nh)^{1/2}\{\hat{g}(\mathbf{x}^\mathsf{T}\hat{\boldsymbol{\gamma}}; \hat{\boldsymbol{\gamma}}) - \hat{g}(\mathbf{x}^\mathsf{T}\boldsymbol{\gamma}_0; \hat{\boldsymbol{\gamma}})\} + (nh)^{1/2}\{\hat{g}(\mathbf{x}^\mathsf{T}\boldsymbol{\gamma}_0; \hat{\boldsymbol{\gamma}}) - g_0(\mathbf{x}^\mathsf{T}\boldsymbol{\gamma}_0)\}$, where $\hat{g}(\mathbf{x}^\mathsf{T}\hat{\boldsymbol{\gamma}}; \hat{\boldsymbol{\gamma}})$ is $\hat{g}(\cdot; \hat{\boldsymbol{\gamma}})$ evaluated at $\mathbf{x}^\mathsf{T}\hat{\boldsymbol{\gamma}}$ and $\hat{g}(\mathbf{x}^\mathsf{T}\boldsymbol{\gamma}_0; \hat{\boldsymbol{\gamma}})$ is $\hat{g}(\cdot; \hat{\boldsymbol{\gamma}})$ evaluated at $\mathbf{x}^\mathsf{T}\boldsymbol{\gamma}_0$. In the above equation, the first part is handled by the Taylor expansion and the second part is handled by Theorem 1. Thus, we also have the following result.

Theorem 2. *Under the same conditions as in Theorem* 1,

$$(nh)^{1/2}\{\hat{\theta}_\tau(\mathbf{x}) - \theta_\tau(\mathbf{x}) - \beta(\mathbf{x}^\mathsf{T}\boldsymbol{\gamma}_0)h^2\} \xrightarrow{D} N(0, \alpha^2(\mathbf{x}^\mathsf{T}\boldsymbol{\gamma}_0)), \tag{12}$$

where $\hat{\theta}_\tau(\mathbf{x}) = \hat{g}(\mathbf{x}^\mathsf{T}\hat{\boldsymbol{\gamma}}; \hat{\boldsymbol{\gamma}})$, $\theta_\tau(\mathbf{x}) = g_0(\mathbf{x}^\mathsf{T}\boldsymbol{\gamma}_0)$, $\beta(\cdot)$ *and* $\alpha(\cdot)$ *are defined as in Theorem* 1, *and* $\beta(\mathbf{x}^\mathsf{T}\boldsymbol{\gamma}_0)h^2$ *and* $\alpha^2(\mathbf{x}^\mathsf{T}\boldsymbol{\gamma}_0)/nh$ *are the asymptotic bias and variance respectively.*

Proof. See Appendix. □

3.2. Asymptotics for parametric part

We need the following additional assumption.

Assumption

(v) The following expectations exist.

$$\mathbf{C}_0 = E\left\{g_0'(\mathbf{x}^\mathsf{T}\boldsymbol{\gamma}_0)^2\left[\mathbf{x} - E(\mathbf{x}|\mathbf{x}^\mathsf{T}\boldsymbol{\gamma}_0)\right]\left[\mathbf{x} - E(\mathbf{x}|\mathbf{x}^\mathsf{T}\boldsymbol{\gamma}_0)\right]^\mathsf{T}\right\},$$

$$\mathbf{C}_1 = E\left\{f_y(g_0(\mathbf{x}^\mathsf{T}\boldsymbol{\gamma}_0))g_0'(\mathbf{x}^\mathsf{T}\boldsymbol{\gamma}_0)^2\left[\mathbf{x} - E(\mathbf{x}|\mathbf{x}^\mathsf{T}\boldsymbol{\gamma}_0)\right]\left[\mathbf{x} - E(\mathbf{x}|\mathbf{x}^\mathsf{T}\boldsymbol{\gamma}_0)\right]^\mathsf{T}\right\}.$$

Theorem 3. *Under Assumptions* (i)–(v), *if* $n \to \infty$, $h \to 0$, *and* $nh \to \infty$, *and if* $\hat{\boldsymbol{\gamma}}$ *is the minimizer of* (9), *then we have the following,*

$$\sqrt{n}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0) \xrightarrow{D} N(0, \tau(1-\tau)\mathbf{C}_1^{-1}\mathbf{C}_0\mathbf{C}_1^{-1}). \tag{13}$$

Proof. See Appendix. □

Remark 2. In (13), generalized inverse $\mathbf{C}_1^{-1}$ is taken, since $\mathbf{C}_1$ is not full ranked.

4. Numerical studies

4.1. Simulations

We use several simulation examples to study the properties of the estimators.

4.1.1. Example 1

The first example is a sine-bump model with homoscedastic errors. The resulting quantiles are the sums of the sine function and a constant. The design is similar to that of Carroll et al. [6] in mean regression context but without the partially linear term included in the original study:

$$y = \sin\left(\frac{\pi(u-A)}{C-A}\right) + 0.1Z, \tag{14}$$

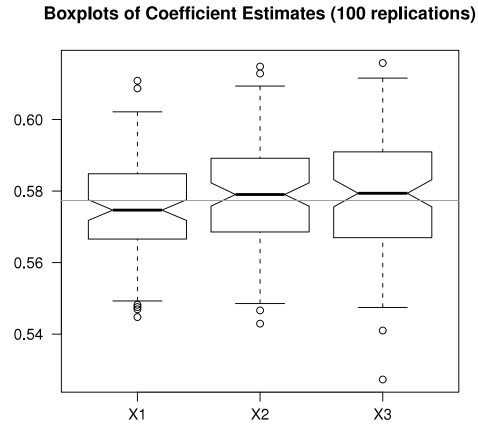

Fig. 1. Boxplot of single-index coefficient estimates for Simulation Example 1. True $\boldsymbol{\gamma}_0 \approx (0.5774, 0.5774, 0.5774)^{\mathsf{T}}$ (horizontal line).

where $u = \mathbf{x}^{\mathsf{T}}\boldsymbol{\gamma}$, $\mathbf{x} = (x_1, x_2, x_3)^{\mathsf{T}}$, $\boldsymbol{\gamma}_0 = \frac{1}{\sqrt{3}}(1, 1, 1)^{\mathsf{T}}$, $A = \frac{\sqrt{3}}{2} - \frac{1.645}{\sqrt{12}}$, $C = \frac{\sqrt{3}}{2} + \frac{1.645}{\sqrt{12}}$; x_i i.i.d. $\sim$ $Unif(0, 1)$, $i = 1, 2, 3$; $Z \sim N(0, 1)$; x_i's and Z are mutually independent.

The index coefficients $\boldsymbol{\gamma}$ are estimated via a series of quantile regressions with $\tau = 0.1, 0.3, 0.5, 0.7$ and 0.9 respectively. For each τ, we simulate 100 random samples, each with sample size $n = 200$. Table 1 shows average estimates, sample standard errors (s.e.), bias and mean squared error (MSE) of the single-index coefficient estimate $\hat{\boldsymbol{\gamma}}$. Fig. 1 shows the box plots of the 100 coefficient estimates from single-index median regression ($\tau = 0.5$). One can see that the distributions of estimates are centered around the true values.

We further conduct a Monte Carlo variance study. The study is designed as follows: for each simulated design matrix $\mathbf{x}$, we simulate 100 response vectors $\mathbf{y}$, then compute 100 estimates of the parameter vector $\hat{\boldsymbol{\gamma}}$. Denote the Monte Carlo sample of parameter estimates as $\left\{\hat{\boldsymbol{\gamma}}\right\}_1^{100}$. Next, we compare the sample standard error $MCse$ of $\left\{\hat{\boldsymbol{\gamma}}\right\}_1^{100}$ with the estimated asymptotic standard error and bootstrap standard error estimators. $MCse$ is a Monte Carlo estimate of the true standard error and is used in place of the true standard error in assessing the performance of the estimated asymptotic standard error and bootstrap standard error. The relative difference between the standard error estimates $\widehat{se}$ and Monte Carlo standard error is measured by

$$D = \frac{\text{norm}(\widehat{se} - MCse)}{\text{norm}(MCse)}. \tag{15}$$

This generic formula is applied to two standard vector norms. D_2 denotes the above formula applied using the L_2 (Euclidean) norm while D_1 indicates the L_1 norm is used.

Asymptotic standard errors obtained from (13) of Theorem 3 involves several unknown quantities, such as the density function $f_y(\cdot)$ and the expectation $E(x|\cdot)$. We may replace these quantities with an estimated density and unconditional sample averages in practice. In the simulation study, the true density function is known, which has been used here.

The procedure used for bootstrap standard error is similar to that described in [16]. In particular, for a given τ:

1. Compute the global error, $\widehat{\epsilon}_{\tau,i} = y_i - \hat{g}_\tau(\mathbf{x}^{\mathsf{T}}\hat{\boldsymbol{\gamma}})$. Center the errors to have zero mean.
2. Resample from $\{\widehat{\epsilon}_{\tau,i}\}$ to form $\{\epsilon^*_{\tau,i}\}$.
3. Create new "observations" $y^*_i = \hat{g}_\tau(\mathbf{x}^{\mathsf{T}}\hat{\boldsymbol{\gamma}}) + \epsilon^*_{\tau,i}$.
4. Obtain an estimate of single-index conditional quantile for (x_i, y^*_i).

We repeat the last 3 steps $B = 100$ times and then take the sample standard error of the $B = 100$ single-index coefficient estimates as the bootstrap standard error (Bootstrap s.e.).

Table 2 shows the estimated average bootstrap standard error, the estimated asymptotic standard error and the Monte Carlo sample standard error over 100 simulations. They share some common features — standard errors of the estimates are greater for estimates estimated from extreme quantiles than for estimates estimated from central quantiles. Compared to the magnitude of true coefficients, the bias can be considered negligible.

Table 2 also gives the relative distance between the estimated and sample standard error. We observe that both bootstrap and asymptotic standard error estimates applied to our single-index quantile regression model give good estimates of the true standard error. However, in practice, we would recommend the bootstrap standard error because the expression for the asymptotic variance formula (13) involves an unknown density function, which may be computationally intensive to estimate. This concurs with the conclusions of De Gooijer and Zerom [16].

Table 1
Monte Carlo study for Simulation Example 1. True $\boldsymbol{\gamma}_0 \approx (0.5774, 0.5774, 0.5774)^{\mathsf{T}}$. The sample average, sample standard error (s.e.), Bias, and Mean Squared Error (MSE) of the single-index parameter estimates. "MSE" is the mean squared error, the squared bias plus squared s.e.

	Estimate	$\hat{\gamma}_1$	$\hat{\gamma}_2$	$\hat{\gamma}_3$
$\tau = 0.1$	Average	0.5728	0.5774	0.5803
	s.e.	0.0198	0.0234	0.0280
	Bias	−0.0046	0.0000	0.0029
	MSE	0.0004	0.0005	0.0008
$\tau = 0.3$	Average	0.5748	0.5801	0.5765
	s.e.	0.0141	0.0146	0.0172
	Bias	−0.0026	0.0027	−0.0009
	MSE	0.0002	0.0002	0.0003
$\tau = 0.5$	Average	0.5748	0.5786	0.5781
	s.e.	0.0131	0.0153	0.0166
	Bias	−0.0026	0.0012	0.0007
	MSE	0.0002	0.0002	0.0003
$\tau = 0.7$	Average	0.5720	0.5830	0.5763
	s.e.	0.0164	0.0156	0.0170
	Bias	−0.0054	0.0056	−0.0011
	MSE	0.0003	0.0003	0.0003
$\tau = 0.9$	Average	0.5643	0.5830	0.5829
	s.e.	0.0195	0.0225	0.0208
	Bias	−0.0131	0.0056	0.0055
	MSE	0.0006	0.0005	0.0005

Table 2
Monte Carlo Variance Study for Simulation Example 1. True $\boldsymbol{\gamma}_0 \approx (0.5774, 0.5774, 0.5774)^{\mathsf{T}}$. Relative distance between the bootstrap standard error (Bootstrap s.e.) and sample standard error (MC s.e.); the asymptotic standard error (Asym. s.e.) and sample standard error (MC s.e.) are measured by D_2 (L_2 Euclidean norm) and D_1 (L_1 norm) defined in Eq. (15).

	Standard error	$\hat{\gamma}_1$	$\hat{\gamma}_2$	$\hat{\gamma}_3$	D_2	D_1
$\tau = 0.1$	MC s.e.	0.01980	0.02339	0.02799		
	Bootstrap s.e.	0.01929	0.02032	0.02489	0.10568	0.09371
	Asym. s.e.	0.01990	0.01955	0.01937	0.22728	0.17639
$\tau = 0.3$	MC s.e.	0.01415	0.01461	0.01722		
	Bootstrap s.e.	0.01550	0.01511	0.01759	0.05569	0.04828
	Asym. s.e.	0.01534	0.01507	0.01494	0.09833	0.08576
$\tau = 0.5$	MC s.e.	0.01314	0.01529	0.01662		
	Bootstrap s.e.	0.01386	0.01330	0.01469	0.10972	0.10318
	Asym. s.e.	0.01459	0.01433	0.01420	0.11404	0.10729
$\tau = 0.7$	MC s.e.	0.01643	0.01563	0.01700		
	Bootstrap s.e.	0.01401	0.01496	0.01636	0.09150	0.07614
	Asym. s.e.	0.01534	0.01507	0.01494	0.08465	0.07560
$\tau = 0.9$	MC s.e.	0.01951	0.02254	0.02082		
	Bootstrap s.e.	0.01950	0.02068	0.01977	0.05873	0.04650
	Asym. s.e.	0.01990	0.01955	0.01937	0.09195	0.07679

4.1.2. Example 2

We consider a location-scale model, where both the location and the scale depend on a common index u. The quantiles are "almost-linear-in-index" [11] when the single-index u is close to zero:

$$y = 10\sin(0.75u) + \sqrt{\sin(u) + 1}Z, \tag{16}$$

where $u = \mathbf{x}^{\mathsf{T}}\boldsymbol{\gamma}$, $\mathbf{x} = (x_1, x_2)^{\mathsf{T}}$, $\boldsymbol{\gamma}_0 = (1, 2)^{\mathsf{T}}/\sqrt{5}$; x_i i.i.d. $\sim N(0, 0.25^2)$, $i = 1, 2$, $Z \sim N(0, 1)$, x_i's and Z are mutually independent.

We generate data from (16) with sample size $n = 400$. Fig. 2 shows boxplots of 100 estimates obtained from single-index quantile regression with $\tau = 0.5$. The Monte Carlo estimates are $\hat{\boldsymbol{\gamma}} = (0.4508, 0.8918)^{\mathsf{T}}$ with Monte Carlo standard errors $(0.0347, 0.0180)^{\mathsf{T}}$. The true values $\boldsymbol{\gamma}_0 = (1, 2)^{\mathsf{T}}/\sqrt{5} \approx (0.4472, 0.8944)^{\mathsf{T}}$. Again one can see that the single-index estimates are close to and are centered around the true parameters.

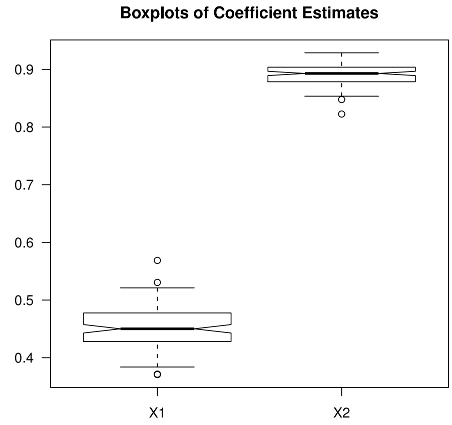

Fig. 2. Boxplot of single-index coefficient estimates for Simulation Example 2. True $\boldsymbol{\gamma}_0 \approx (0.4472, 0.8944)^{\mathsf{T}}$.

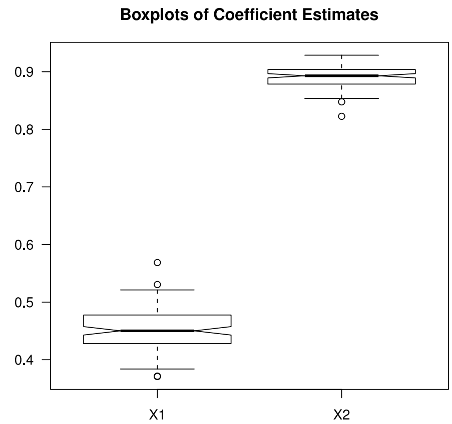

Fig. 3. Boxplot of single-index coefficient estimates for Simulation Example 3. True $\boldsymbol{\gamma}_0 \approx (0.4472, 0.8944)^{\mathsf{T}}$.

4.1.3. Example 3

Quantile regression does not require strict assumptions on error distributions. Here we consider an asymmetric (exponential) distribution:

$$y = 5\cos(u) + \exp(-u^2) + E, \tag{17}$$

where $u = \mathbf{x}^{\mathsf{T}}\boldsymbol{\gamma}, \mathbf{x} = (x_1, x_2)^{\mathsf{T}}, \boldsymbol{\gamma}_0 = (1, 2)^{\mathsf{T}}/\sqrt{5}$; x_i i.i.d. $\sim N(0, 1), i = 1, 2$, the residual E follows an exponential distribution with mean 2, x_i's and E are mutually independent.

Fig. 3 shows the boxplots of single-index coefficient estimates from 100 replications when $\tau = 0.5$ and sample size $n = 400$. Again we observe similar pattern as in the previous examples.

4.2. An application to Boston housing data

We consider an application regarding Boston housing data. The data contain 506 observations on 14 variables, the dependent variable of interest is *medv*, the median value of owner-occupied homes in \$1000's. Thirteen other statistical measurements on the 506 census tracts in suburban Boston from the 1970 census are also included. This data can be found in the *StatLib* library maintained at Carnegie Mellon University.

Many regression studies have used this data set and found potential relationship between *medv* and *RM, TAX, PTRATIO, LSTAT* [25,17]; *RM, LSTAT, DIS* [9]. In this study, we first focus on the following four covariates:

RM: average number of rooms per dwelling;
TAX: full-value property tax (in dollar) per \$10,000;
PTRATIO: pupil-teacher ratio by town;
LSTAT: percentage of lower status of the population.

We follow previous studies and take logarithmic transformations on *TAX* and *LSTAT*. The dependent variable is centered around zero. No cleaning is done on the covariates. We notice that the dependent variable is censored from above and

Table 3
Single-index coefficient estimates (standard errors) for Boston housing data.

τ	RM	log (TAX)	PTRATIO	log (LSTAT)
0.10	0.3380 (0.0364)	−0.5702 (0.0495)	−0.0527 (0.0061)	−0.7469 (0.0392)
0.25	0.3360 (0.0516)	−0.5362 (0.0526)	−0.0669 (0.0057)	−0.7714 (0.0749)
0.50	0.3687 (0.0285)	−0.4515 (0.0454)	−0.0718 (0.0075)	−0.8093 (0.0205)
0.75	0.2406 (0.0581)	−0.1969 (0.0115)	−0.0946 (0.0141)	−0.9457 (0.0259)
0.90	0.0776 (0.0165)	−0.2809 (0.0265)	−0.0714 (0.0043)	−0.9539 (0.1016)

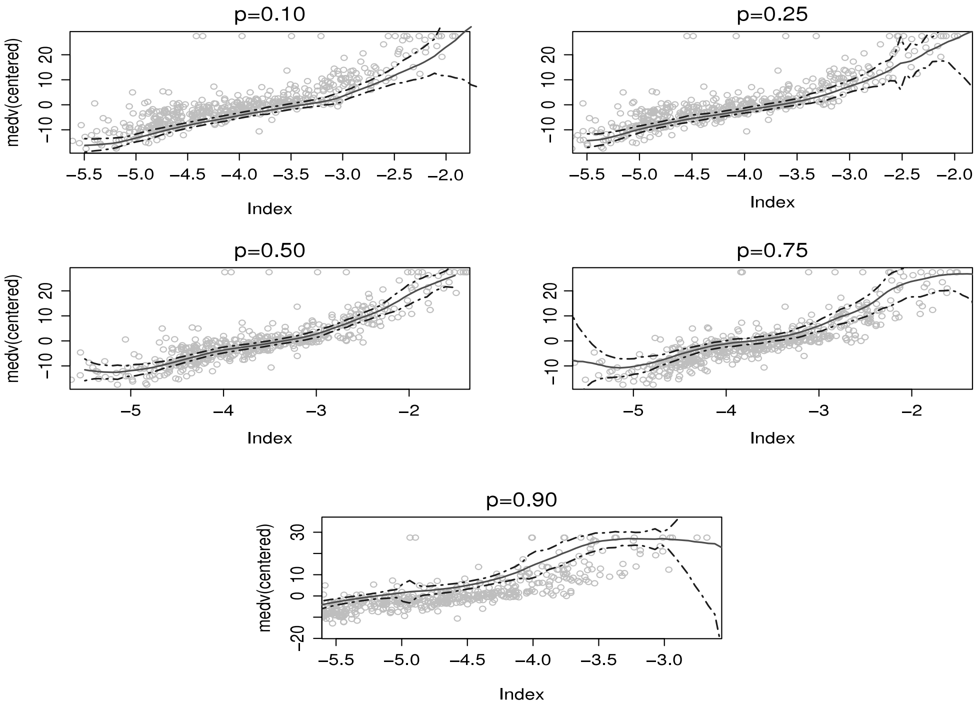

Fig. 4. Quantiles and their 95% pointwise confidence intervals for Boston housing data.

modeling of conditional quantiles is more appropriate than modeling of averages. In this study each conditional quantile is modeled by a single-index model:

$$\theta_\tau(medv|RM, TAX, PTRATIO, LSTAT) = g\left(\gamma_1 RM + \gamma_2 \log(TAX) + \gamma_3 PTRATIO + \gamma_4 \log(LSTAT)\right). \tag{18}$$

Typical estimates from various combinations of starting values and convergence criteria are presented in Table 3. The coefficients indicate relative effects of the four covariates on a particular percentile as well as relative effects of a given covariate on different percentiles. log *(LSTAT)* seems to be the most important covariate for all percentile levels comparing the absolute values of the normalized coefficients. Even though the relationship between *RM* and lower quantiles of *medv* is substantial, we see only marginal effect of *RM* on the upper quantiles. A similar pattern is observed for log *(TAX)*. The pattern is reversed for log *(LSTAT)*. Similar to [16], because the asymptotic variance formulas (12) and (13) are rather complicated, we use bootstrap standard errors instead of the asymptotic estimates in real data applications. Bootstrap standard errors for the estimates are shown in parenthesis in Table 3. The estimated 10th, 25th, 50th 75th, 90th quantiles and their 95% pointwise confidence intervals are shown in Fig. 4 together with scatter plots of y and the estimated indices. We notice possible quantile curves crossing at both tails, which reflect a paucity of data in the region concerned.

We then fit another single-index quantile model using the three covariates as used in Chaudhuri et al. [9] on average derivative quantile regression (ADE): *RM*, *LSTAT* and *DIS*.

$$\theta_\tau(medv|RM, LSTAT, DIS) = g\left(\gamma_1 RM + \gamma_2 LSTAT + \gamma_3 DIS\right). \tag{19}$$

The variable *DIS* is weighted distances to five Boston employment centers. All three covariates are standardized to have zero mean and unit variance. The estimates for the single-index coefficient are given in Table 4. Again to those observed in

Table 4
Single-index coefficient estimates with [9] covariates from Boston housing data.

τ	RM	LSTAT	DIS
0.10	0.2506	−0.9676	−0.0292
0.25	0.2259	−0.9729	−0.0499
0.50	0.2682	−0.9577	−0.1043
0.75	0.2988	−0.9418	−0.1540
0.90	0.2737	−0.8193	−0.5038

Table 5
Model average sum of check function based (absolute) residuals comparison for Boston housing data.

τ	Model (18)	Model (19)	Model (ADE)
0.10	1.102	1.228	1.559
0.25	2.105	2.229	2.696
0.50	2.845	2.874	3.042
0.75	2.577	2.490	2.430
0.90	1.749	3.320	3.126

Chaudhuri et al. [9], LSTAT seems to be the most important covariate for all percentile levels comparing the absolute values of the normalized coefficients. We observe that the effects of *RM* and *LSTAT* are stable across different quantiles, but the effect of *DIS* varies quite much for different quantiles. A similar change pattern is observed in Chaudhuri et al. [9]: the effects of *RM* and *LSTAT* are more significant in magnitude and for each of the two variables; the effects on different quantiles are very similar, but the effects of *DIS* on different quantiles can vary from −0.292 to 0.593 [9] compared to a varying range from −0.504 to −0.029 using our estimation approach.

To assess the relative success of the two model specifications in (18) and (19), we compare the average sum of check function based (absolute) residuals $R_\tau = \frac{1}{n}\sum_i \rho_\tau(y_i - \hat{\theta}_\tau(\mathbf{x}_i))$, where $\rho_\tau(u) = |u| + (2\tau - 1)u$. This is analogous to average sum of absolute residuals (or mean absolute deviation) in mean regression. Table 5 gives model R_τ for three different model fits, where column two follows model (18) specification with four covariates; column three follows model (19) specification with three covariates as in [9]; and column four corresponds to model (19) specification but with exact average derivative estimate (ADE) from [9]. Table 5 suggests that model (18) gives the smallest average sum of ρ_τ (absolute) residuals R_τ except when $\tau = 0.75$. Model (19) with the proposed single-index estimate yields smaller average sum of ρ_τ (absolute) residuals than that using exact average derivative estimate (ADE) from [9], for lower quantiles $\tau = 0.1, 0.25, 0.50$, where the sign of *DIS* coefficient takes a different direction. For larger quantiles $\tau = 0.75, 0.90$, ADE gives smaller R_τ instead. Note that model (18) uses four covariates; model (19) and ADE use three covariates. Model complexity measures, such as model degree of freedom, are not taken into consideration in R_τ here. Although in mean regression there are different measures of model fit, such theory is very limited in quantile regression. For linear quantile regression with an asymmetric Laplace error distribution, more exact test statistics have been studied by Koenker and Machado [26]. Model complexity measure, such as the notion "effective degree of freedom" in nonparametric mean regression, may be an interesting future research topic to pursue for nonparametric quantile estimation.

5. Conclusions

We have proposed a single-index quantile regression model that effectively reduces dimensionality and is parsimonious and flexible. The local linear estimation approach is adopted. We have derived the large sample properties of the estimates and further studied the inferences. We illustrated the proposed approach with both simulation examples and real data applications. The approach described here may be extended to a generalized single-index quantile regression framework with a known link function, which would be considered in future research.

Appendix

A.1. Proof of Theorem 1

We first quote the following lemma, which will be later used in our proof.

Quadratic Approximation Lemma (*[27]*). *Suppose $A_n(s)$ is convex and can be represented as $\frac{1}{2}s'Vs + U_n's + C_n + r_n(s)$, where V is symmetric and positive definite, U_n is stochastically bounded, C_n is arbitrary, and $r_n(s)$ goes to zero in probability for each s. Then α_n, the argmin of A_n, is only $o_p(1)$ away from $\beta_n = -V^{-1}U_n$, the argmin of of $\frac{1}{2}s'Vs + U_n's + C_n$. If also $U_n \to_d U$, then $\alpha_n \to_d -V^{-1}U$.*

We explicitly write $\hat{g}(u; h, \hat{\boldsymbol{\gamma}}) := \hat{g}(u; \hat{\boldsymbol{\gamma}})$ to indicate the dependence on h.

$$(nh)^{1/2}\{\hat{g}(u; h, \hat{\boldsymbol{\gamma}}) - g_0(u)\} = (nh)^{1/2}\{\hat{g}(u; h, \hat{\boldsymbol{\gamma}}) - \hat{g}(u; h, \boldsymbol{\gamma}_0)\} + (nh)^{1/2}\{\hat{g}(u; h, \boldsymbol{\gamma}_0) - g_0(u)\}, \tag{20}$$

where $\hat{g}(\cdot; h, \boldsymbol{\gamma}_0)$ is a local linear estimator of $g_0(\cdot)$ if the index coefficient $\boldsymbol{\gamma}_0$ is known. According to Theorem 3 in [23], we have

$$(nh)^{1/2}\{\hat{g}(u; h, \boldsymbol{\gamma}_0) - g_0(u) - \beta(u)h^2\} \xrightarrow{D} N\left(0, \alpha^2(u)\right). \tag{21}$$

The first part on the right-hand side of (20), $(nh)^{1/2}\{\hat{g}(u; h, \hat{\boldsymbol{\gamma}}) - \hat{g}(u; h, \boldsymbol{\gamma}_0)\}$ can be shown $o_p(1)$. The details are given below.

For given u, for notational simplicity, we write $\hat{a}_{\hat{\boldsymbol{\gamma}}} := \hat{g}(u; h, \hat{\boldsymbol{\gamma}})$, $\hat{b}_{\hat{\boldsymbol{\gamma}}} := \hat{g}'(u; h, \hat{\boldsymbol{\gamma}})$, $\hat{a}_{\boldsymbol{\gamma}_0} := \hat{g}(u; h, \boldsymbol{\gamma}_0)$, and $\hat{b}_{\boldsymbol{\gamma}_0} := \hat{g}'(u; h, \boldsymbol{\gamma}_0)$ which are the solutions of the following minimization problems respectively,

$$\min_{a,b} \sum_{i=1}^{n} \rho_\tau\left(y_i - a - b(\mathbf{x}_i^\mathsf{T}\hat{\boldsymbol{\gamma}} - u)\right) K((\mathbf{x}_i^\mathsf{T}\hat{\boldsymbol{\gamma}} - u)/h), \tag{22}$$

$$\min_{a,b} \sum_{i=1}^{n} \rho_\tau\left(y_i - a - b(\mathbf{x}_i^\mathsf{T}\boldsymbol{\gamma}_0 - u)\right) K((\mathbf{x}_i^\mathsf{T}\boldsymbol{\gamma}_0 - u)/h). \tag{23}$$

We assume that the minimizers to both (22) and (23) exist for the following discussion to be meaningful and we want to show whether $(nh)^{1/2}\left(\hat{a}_{\hat{\boldsymbol{\gamma}}} - \hat{a}_{\boldsymbol{\gamma}_0}\right)$ is $o_p(1)$.

Each of the objective functions in (22) and (23) is non-differentiable and the resulting estimator is implicit. We consider quadratic approximation for each objective function. Under the root-n assumption on a preliminary estimator $\hat{\boldsymbol{\gamma}}$, we show quadratic approximations for (22) and (23) are "close enough" that their minimizers are "close enough" to each other in a sense defined later.

Denote

$$\bar{\boldsymbol{\theta}}_n^* = (nh)^{1/2}\left(\hat{a}_{\hat{\boldsymbol{\gamma}}} - g_0(u), h(\hat{b}_{\hat{\boldsymbol{\gamma}}} - g_0'(u))\right)^\mathsf{T},$$

$$\bar{\boldsymbol{\theta}}_n^{**} = (nh)^{1/2}\left(\hat{a}_{\boldsymbol{\gamma}_0} - g_0(u), h(\hat{b}_{\boldsymbol{\gamma}_0} - g_0'(u))\right)^\mathsf{T},$$

$$\mathbf{Z}_i^* = \left(1, (\mathbf{x}_i^\mathsf{T}\hat{\boldsymbol{\gamma}} - u)/h\right)^\mathsf{T}, \qquad \mathbf{Z}_i^{**} = \left(1, (\mathbf{x}_i^\mathsf{T}\boldsymbol{\gamma}_0 - u)/h\right)^\mathsf{T},$$

and denote

$$y_i^* = y_i - g_0(u) - g_0'(u)(\mathbf{x}_i^\mathsf{T}\hat{\boldsymbol{\gamma}} - u), \qquad y_i^{**} = y_i - g_0(u) - g_0'(u)(\mathbf{x}_i^\mathsf{T}\boldsymbol{\gamma}_0 - u),$$
$$K_i^* = K((\mathbf{x}_i^\mathsf{T}\hat{\boldsymbol{\gamma}} - u)/h), \qquad K_i^{**} = K((\mathbf{x}_i^\mathsf{T}\boldsymbol{\gamma}_0 - u)/h).$$

Thus $\bar{\boldsymbol{\theta}}_n^*$ and $\bar{\boldsymbol{\theta}}_n^{**}$ minimize

$$Q_n^*(\boldsymbol{\theta}) = \sum_{i=1}^{n}\left[\rho_\tau(y_i^* - \boldsymbol{\theta}^\mathsf{T}\mathbf{Z}_i^*/\sqrt{nh}) - \rho_\tau(y_i^*)\right]K_i^* \quad \text{and}$$

$$Q_n^{**}(\boldsymbol{\theta}) = \sum_{i=1}^{n}\left[\rho_\tau(y_i^{**} - \boldsymbol{\theta}^\mathsf{T}\mathbf{Z}_i^{**}/\sqrt{nh}) - \rho_\tau(y_i^{**})\right]K_i^{**} \text{ respectively.}$$

Both $Q_n^*(\boldsymbol{\theta})$ and $Q_n^{**}(\boldsymbol{\theta})$ are convex in $\boldsymbol{\theta}$ and they converge pointwise to their conditional expectations whose quadratic approximations can be more easily derived. The convergence is also uniform on any compact set of $\boldsymbol{\theta}$ [23]. Following [23], we can show in the same way

$$Q_n^*(\boldsymbol{\theta}) = \frac{1}{2}\boldsymbol{\theta}^\mathsf{T}\mathbf{S}^*\boldsymbol{\theta} + \mathbf{W}_n^{*\mathsf{T}}\boldsymbol{\theta} + r_n^*(\boldsymbol{\theta}), \quad r_n^*(\boldsymbol{\theta}) = o_p(1), \tag{24}$$

$$Q_n^{**}(\boldsymbol{\theta}) = \frac{1}{2}\boldsymbol{\theta}^\mathsf{T}\mathbf{S}^{**}\boldsymbol{\theta} + \mathbf{W}_n^{**\mathsf{T}}\boldsymbol{\theta} + r_n^{**}(\boldsymbol{\theta}), \quad r_n^{**}(\boldsymbol{\theta}) = o_p(1), \tag{25}$$

where

$$\mathbf{S}^* = \mathbf{S}^{**} = f_{U_0}(u)\varphi''(0|u)\begin{pmatrix}1 & 0\\ 0 & \int_{\mathcal{V}} K(\mathcal{V})\mathcal{V}^2 d\mathcal{V}\end{pmatrix},$$

$$\mathbf{W}_n^* = -(nh)^{-1/2}\sum \rho_\tau'(y_i^*)\mathbf{Z}_i^* K_i^*, \qquad \mathbf{W}_n^{**} = -(nh)^{-1/2}\sum \rho_\tau'(y_i^{**})\mathbf{Z}_i^{**} K_i^{**}.$$

Here $\varphi''(0|u)$ is the second derivative of $\varphi(t|u) = E(\rho_\tau(y - g_0(u) + t)|U = u)$ with respect to t evaluated at $t = 0$. The first and second derivatives of $\varphi(t|u)$ with respect to t, $\varphi'(t|u)$ and $\varphi''(t|u)$, are assumed to exist. And $\mathcal{V} \in [-M, M]$, where M is such a real number that $[-M, M]$ contains the support of $K(\cdot)$.

Write

$$Q_n^*(\boldsymbol{\theta}) = E(Q_n^*(\boldsymbol{\theta})|\mathcal{X}) - (nh)^{-1/2}\left(\sum \rho_\tau'(y_i^*)\mathbf{Z}_i^{*\mathsf{T}}K_i^* - E(\rho_\tau'(y_i^*)|u_i)\mathbf{Z}_i^{*\mathsf{T}}K_i^*\right)\boldsymbol{\theta} + R_n^*(\boldsymbol{\theta}), \tag{26}$$

where $\rho_\tau'(y_i^*)$ only exists for $y_i^* \neq 0$, however we set $\rho_\tau'(0)$ to 0 since $y_i^* = 0$ occurs only with a probability of zero. Write $u_i = \mathbf{x}_i^{\mathsf{T}}\hat{\boldsymbol{\gamma}}$. Then we have the following result,

$$\begin{aligned} E(Q_n^*(\boldsymbol{\theta})|\mathcal{X}) &= [\varphi(g_0(u_i) - g_0(u) - g_0'(u)(u_i-u) - \boldsymbol{\theta}^{\mathsf{T}}\mathbf{Z}_i^*/\sqrt{nh}|u_i) - \varphi(g_0(u_i) - g_0(u) - g_0'(u)(u_i-u)|u_i)]K_i^* \\ &= -(nh)^{-1/2}\sum \varphi'(g_0(u_i) - g_0(u) - g_0'(u)(u_i-u)|u_i)(\boldsymbol{\theta}^{\mathsf{T}}\mathbf{Z}_i^*)K_i^* \\ &\quad + (2nh)^{-1}\boldsymbol{\theta}^{\mathsf{T}}\left(\sum K_i^*\varphi''(g_0(u_i) - g_0(u) - g_0'(u)(u_i-u)|u_i)\mathbf{Z}_i^*\mathbf{Z}_i^{*\mathsf{T}}\right)\boldsymbol{\theta}(1+o_p(1)) \\ &= -(nh)^{-1/2}\sum E(\rho_\tau'(y_i^*)|u_i)(\boldsymbol{\theta}^{\mathsf{T}}\mathbf{Z}_i^*)K_i^* \\ &\quad + (2nh)^{-1}\boldsymbol{\theta}^{\mathsf{T}}\left(\sum K_i^*\varphi''(g_0(u_i) - g_0(u) - g_0'(u)(u_i-u)|u_i)\mathbf{Z}_i^*\mathbf{Z}_i^{*\mathsf{T}}\right)\boldsymbol{\theta}(1+o_p(1)), \end{aligned} \tag{27}$$

and

$$\begin{aligned} K_i^*\varphi''(g_0(u_i) - g_0(u) - g_0'(u)(u_i-u)|u_i) &= K_i^*(\varphi''(0 + O_p(h^2)|u)) \\ &= K_i^*(\varphi''(0|u) + O_p(h^2)). \end{aligned} \tag{28}$$

The last equality requires $\varphi'''(0|u)$ exists and is bounded. The second term in the last equality of (27) can be rewritten as

$$\begin{aligned} (nh)^{-1}\sum K_i^*\varphi''(g_0(u_i) - g_0(u) - g_0'(u)(u_i-u)|u_i)\mathbf{Z}_i^*\mathbf{Z}_i^{*\mathsf{T}} &= (nh)^{-1}\sum K_i^*\left(\varphi''(0|u) + O_p(h^2)\right)\mathbf{Z}_i^*\mathbf{Z}_i^{*\mathsf{T}} \\ &= (nh)^{-1}\sum K_i^*\varphi''(0|u)\mathbf{Z}_i^*\mathbf{Z}_i^{*\mathsf{T}}\left(1 + O_p(h^2)\right) \\ &= \mathbf{S}\left(1 + O_p(h^2)\right), \end{aligned} \tag{29}$$

where

$$\begin{aligned} \mathbf{S} &= (nh)^{-1}\sum K_i^*\varphi''(0|u)\mathbf{Z}_i^*\mathbf{Z}_i^{*\mathsf{T}} \\ &= \begin{pmatrix} s_0 & s_1 \\ s_1 & s_2 \end{pmatrix}, \end{aligned}$$

with the matrix components $s_j = (nh)^{-1}\sum K((u_i-u)/h)\varphi''(0|u)((u_i-u)/h)^j, j = 0, 1, 2.$

$$\begin{aligned} E(s_j) &= h^{-1}\varphi''(0|u)\int_{\mathcal{U}} K((\mathcal{U}-u)/h)((\mathcal{U}-u)/h)^j f_U(\mathcal{U})\mathrm{d}\mathcal{U} \\ &= \varphi''(0|u)\int_{\mathcal{V}} K(\mathcal{V})\mathcal{V}^j f_{U_0}(\mathcal{V}h+u)\mathrm{d}\mathcal{V}(1+o(1)) \\ &\qquad \text{recall } \mathcal{V} \in [-M, M];\, f_U(\cdot) = f_{U_0}(\cdot)(1+o(1)) \text{ by Assumption (ii),} \\ &\qquad \text{where } U = \mathbf{x}^{\mathsf{T}}\hat{\boldsymbol{\gamma}},\, U_0 = \mathbf{x}^{\mathsf{T}}\boldsymbol{\gamma}_0,\, f_U(\cdot) \text{ and } f_{U_0}(\cdot) \text{ are their densities respectively.} \\ &= f_{U_0}(u)\varphi''(0|u)\left[\int_{\mathcal{V}} K(\mathcal{V})\mathcal{V}^j\mathrm{d}\mathcal{V}\right](1+o(1)) \\ &= f_{U_0}(u)\varphi''(0|u)c_j(1+o(1)), \quad \text{where } c_j = \int_{\mathcal{V}} K(\mathcal{V})\mathcal{V}^j\mathrm{d}\mathcal{V};\, c_0 = 1, c_1 = 0 \text{ from Assumption (i).} \end{aligned}$$

One can verify $var(s_j) = o(1)$. Therefore

$$\mathbf{S} = f_{U_0}(u)\varphi''(0|u)\begin{pmatrix} 1 & 0 \\ 0 & c_2 \end{pmatrix} + o_p(1) := \mathbf{S}^* + o_p(1), \quad \text{where } c_2 = \int_{\mathcal{V}} K(\mathcal{V})\mathcal{V}^2\mathrm{d}\mathcal{V}. \tag{30}$$

By (27)–(30), we have

$$E(Q_n^*(\boldsymbol{\theta})|\mathcal{X}) = -(nh)^{-1/2}\sum E(\rho_\tau'(y_i^*)|u_i)(\boldsymbol{\theta}^{\mathsf{T}}\mathbf{Z}_i^*)K_i^* + \frac{1}{2}\boldsymbol{\theta}^{\mathsf{T}}\mathbf{S}^*\boldsymbol{\theta}(1+o_p(1)). \tag{31}$$

For $R_n^*(\boldsymbol{\theta})$ defined by (26), $R_n^*(\boldsymbol{\theta}) = o_p(1)$. To save space, we refer readers to [23] for similar idea of proof. Substitute (31) into (26), we have

$$Q_n^*(\boldsymbol{\theta}) = \frac{1}{2}\boldsymbol{\theta}^{\mathsf{T}}\mathbf{S}^*\boldsymbol{\theta} + \mathbf{W}_n^{*\mathsf{T}}\boldsymbol{\theta} + r_n^*(\boldsymbol{\theta}), \tag{32}$$

where $\mathbf{W}_n^* = -(nh)^{-1/2}\sum \rho'_\tau(y_i^*)\mathbf{Z}_i^* K_i^*$ is stochastically bounded, and $r_n^*(\boldsymbol{\theta}) = o_p(1)$. We outline the proof for stochastic boundedness of $\mathbf{W}_n^*$. By change of variable and existence of $\int K^2(\mathcal{V})\mathcal{V}^j \mathrm{d}\mathcal{V}, j = 0, 1, 2$ in Assumption (i), for some $c > 0$,

$$\begin{aligned} E(\mathbf{W}_n^*\mathbf{W}_n^{*\mathsf{T}}) &\le cE\left((nh)^{-1}\sum(\rho'_\tau(y_i^*))^2(K_i^*)^2\mathbf{Z}_i^*\mathbf{Z}_i^{*\mathsf{T}}\right) \\ &= O\left(h^{-1}E\left((K_i^*)^2\mathbf{Z}_i^*\mathbf{Z}_i^{*\mathsf{T}}\right)\right) = O(1), \end{aligned} \tag{33}$$

which also implies $E(\mathbf{W}_n^*) = O(1)$ as a result of Jensen's inequality. Bounded second moment implies that $\mathbf{W}_n^*$ is stochastically bounded. The asymptotic normality of $\hat{\boldsymbol{\theta}}_n^* = -\mathbf{S}^{*-1}\mathbf{W}^*$ follows from that of $\mathbf{W}_n^*$ as a result of Quadratic Approximation Lemma. Eq. (25) can be shown similarly with $\mathbf{S}^{**} = \mathbf{S}^*$, $\mathbf{W}_n^{**} = -(nh)^{-1/2}\sum \rho'_\tau(y_i^{**})\mathbf{Z}_i^{**}K_i^{**}$, and $r_n^{**}(\boldsymbol{\theta}) = o_p(1)$.

According to the Quadratic Approximation Lemma, $\bar{\boldsymbol{\theta}}_n^* - \hat{\boldsymbol{\theta}}_n^* = o_p(1)$ and $\bar{\boldsymbol{\theta}}_n^{**} - \hat{\boldsymbol{\theta}}_n^{**} = o_p(1)$, where $\hat{\boldsymbol{\theta}}_n^* = -\mathbf{S}^{*-1}\mathbf{W}_n^*$ and $\hat{\boldsymbol{\theta}}_n^{**} = -\mathbf{S}^{**-1}\mathbf{W}_n^{**}$. Because of the root-$n$ assumption on $\hat{\boldsymbol{\gamma}}$, we further show $\hat{\boldsymbol{\theta}}_n^* - \hat{\boldsymbol{\theta}}_n^{**} = o_p(1)$.

Write $\mathbf{S}_0 = \mathbf{S}^* = \mathbf{S}^{**}$,

$$\begin{aligned} \hat{\boldsymbol{\theta}}_n^* - \hat{\boldsymbol{\theta}}_n^{**} &= -\mathbf{S}^{-1}(\mathbf{W}_n^* - \mathbf{W}_n^{**}) \\ &= \mathbf{S}^{-1}(nh)^{-1/2}\sum \rho'_\tau(y_i^*)\mathbf{Z}_i^*K_i^* - \rho'_\tau(y_i^{**})\mathbf{Z}_i^{**}K_i^{**} \\ &= \mathbf{S}^{-1}(nh)^{-1/2}\sum \rho'_\tau(y_i^*)(\mathbf{Z}_i^*K_i^* - \mathbf{Z}_i^{**}K_i^{**}). \end{aligned}$$

The last equality is due to the fact that y_i^{**} has the same sign as y_i^* a.s. when $\|\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0\| = O_p(n^{-1/2})$. For some $c > 0$,

$$\begin{aligned} E\left((\hat{\boldsymbol{\theta}}_n^* - \hat{\boldsymbol{\theta}}_n^{**})(\hat{\boldsymbol{\theta}}_n^* - \hat{\boldsymbol{\theta}}_n^{**})^{\mathsf{T}}\right) &\le c\mathbf{S}^{-1}h^{-1}E\left((\rho'_\tau(y_i^*))^2(\mathbf{Z}_i^*K_i^* - \mathbf{Z}_i^{**}K_i^{**})(\mathbf{Z}_i^*K_i^* - \mathbf{Z}_i^{**}K_i^{**})^{\mathsf{T}}\right)(\mathbf{S}^{-1})^{\mathsf{T}}, \\ &= O\left(h^{-1}E\left((\mathbf{Z}_i^*K_i^* - \mathbf{Z}_i^{**}K_i^{**})(\mathbf{Z}_i^*K_i^* - \mathbf{Z}_i^{**}K_i^{**})^{\mathsf{T}}\right)\right) = O\left(o(1)\right) = o(1), \end{aligned} \tag{34}$$

which also implies $E(\hat{\boldsymbol{\theta}}_n^* - \hat{\boldsymbol{\theta}}_n^{**}) = o(1)$. Thus, $\hat{\boldsymbol{\theta}}_n^* - \hat{\boldsymbol{\theta}}_n^{**} = E(\hat{\boldsymbol{\theta}}_n^* - \hat{\boldsymbol{\theta}}_n^{**}) + o_p(1) = o_p(1)$ according to its first and second moment. Therefore $\bar{\boldsymbol{\theta}}_n^* - \bar{\boldsymbol{\theta}}_n^{**} = o_p(1)$. Take the first component of the vector, we obtain $(nh)^{1/2}\{\hat{g}(u; h, \hat{\boldsymbol{\gamma}}) - \hat{g}(u; h, \boldsymbol{\gamma}_0)\} = o_p(1)$.

Furthermore, when the bandwidth is taken at usual optimal rate, i.e. $h \in \mathcal{H}_n = \{h : C_1 n^{-1/5} \le h \le C_2 n^{-1/5}\}$ [24], we have $(nh)^{1/2}\{\hat{g}(u; h, \hat{\boldsymbol{\gamma}}) - \hat{g}(u; h, \boldsymbol{\gamma}_0)\} = o_p(1)$. That proves $(nh)^{1/2}\{\hat{g}(u; h, \hat{\boldsymbol{\gamma}}) - g_0(u) - \beta(u)h^2\} \xrightarrow{D} N(0, \alpha^2(u))$. □

A.2. *Proof of* Theorem 2

For given $\mathbf{x}$,

$$\begin{aligned} (nh)^{1/2}\{\hat{\theta}_\tau(\mathbf{x}) - \theta_\tau(\mathbf{x})\} &= (nh)^{1/2}\{\hat{g}(\mathbf{x}^{\mathsf{T}}\hat{\boldsymbol{\gamma}}; \hat{\boldsymbol{\gamma}}) - \hat{g}(\mathbf{x}^{\mathsf{T}}\boldsymbol{\gamma}_0; \hat{\boldsymbol{\gamma}}) + \hat{g}(\mathbf{x}^{\mathsf{T}}\boldsymbol{\gamma}_0; \hat{\boldsymbol{\gamma}}) - g_0(\mathbf{x}^{\mathsf{T}}\boldsymbol{\gamma}_0)\} \\ &= A + (nh)^{1/2}\{\hat{g}(\mathbf{x}^{\mathsf{T}}\boldsymbol{\gamma}_0; \hat{\boldsymbol{\gamma}}) - g_0(\mathbf{x}^{\mathsf{T}}\boldsymbol{\gamma}_0)\}, \end{aligned}$$

by the Taylor theorem, $A = (nh)^{1/2}\{\hat{g}(\mathbf{x}^{\mathsf{T}}\hat{\boldsymbol{\gamma}}; \hat{\boldsymbol{\gamma}}) - \hat{g}(\mathbf{x}^{\mathsf{T}}\boldsymbol{\gamma}_0; \hat{\boldsymbol{\gamma}})\} = (nh)^{1/2}\hat{g}'(\mathbf{x}^{\mathsf{T}}\boldsymbol{\gamma}_0; \hat{\boldsymbol{\gamma}})O_p(\|\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0\|) = o_p(n^{1/2}h^{5/2})$, which is $o_p(1)$ when the bandwidth is taken at usual optimal rate. Then Theorem 2 is a result of of Theorem 1. □

A.3. *Proof of* Theorem 3

We outline our approach to show asymptotic normality of $\hat{\boldsymbol{\gamma}}$. The proof will again rely on quadratic approximation and literally follow a similar logic as in the proof of Theorem 2. Given $(\hat{a}_j, \hat{b}_j)$, minimize the following to obtain $\hat{\boldsymbol{\gamma}}$,

$$\sum_{j=1}^n\sum_{i=1}^n \rho_\tau\left(y_i - \hat{a}_j - \hat{b}_j(\mathbf{x}_i - \mathbf{x}_j)^{\mathsf{T}}\boldsymbol{\gamma}\right)\omega_{ij}. \tag{35}$$

Write $\hat{\boldsymbol{\gamma}}^* = \sqrt{n}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0)$, then $\hat{\boldsymbol{\gamma}}^*$ minimizes the following,

$$\mathcal{Q}_n(\boldsymbol{\gamma}^*) = \sum_{j=1}^n\sum_{i=1}^n\left[\rho_\tau\left(y_{ij} - \frac{1}{\sqrt{n}}\hat{b}_j\mathbf{x}_{ij}^{\mathsf{T}}\boldsymbol{\gamma}^*\right) - \rho_\tau(y_{ij})\right]\omega_{ij}, \tag{36}$$

where $y_{ij} = y_i - \hat{a}_j - \hat{b}_j\mathbf{x}_{ij}^{\mathsf{T}}\boldsymbol{\gamma}_0$. It can be shown that $\mathcal{Q}_n(\boldsymbol{\gamma}^*) = \frac{1}{2}\boldsymbol{\gamma}^{*\mathsf{T}}\mathbf{T}_n\boldsymbol{\gamma}^* + \mathbf{V}_n^{\mathsf{T}}\boldsymbol{\gamma}^* + o_p(1)$, where $\mathbf{T}_n = 2\mathbf{C}_1$, $\mathbf{V}_n = (4\tau(1-\tau))^{1/2}\mathbf{C}_0{}^{1/2}\mathcal{Z}_n$, $\mathcal{Z}_n \to_D N(0, I)$, and, $\mathbf{C}_0$ and $\mathbf{C}_1$ are defined in Assumption (v). In fact,

$$\mathcal{Q}_n(\boldsymbol{\gamma}^*) = E(\mathcal{Q}_n(\boldsymbol{\gamma}^*)) - \frac{1}{\sqrt{n}}\sum_{j=1}^n\sum_{i=1}^n\left[\omega_{ij}\rho'_\tau(y_{ij})\hat{b}_j\mathbf{x}_{ij}^{\mathsf{T}} - \omega_{ij}E(\rho'_\tau(y_{ij}))\hat{b}_j\mathbf{x}_{ij}^{\mathsf{T}}\right]\boldsymbol{\gamma}^* + \mathcal{R}(\boldsymbol{\gamma}^*),$$

where $\mathcal{R}(\boldsymbol{\gamma}^*) = o_p(1)$.

Write $\varphi(t|\cdot) = \varphi(t|\mathbf{x}^\top\boldsymbol{\gamma}_0 = x^\top\boldsymbol{\gamma}_0) = E(\rho_\tau(y-\theta_\tau(x)+t)|\mathbf{x}^\top\boldsymbol{\gamma}_0 = x^\top\boldsymbol{\gamma}_0)$, we have

$$\begin{aligned}
E(\mathcal{Q}_n(\boldsymbol{\gamma}^*)) &= \sum_{j=1}^n\sum_{i=1}^n\left[E\rho_\tau\left(y_{ij}-\frac{1}{\sqrt{n}}\hat{b}_j\mathbf{x}_{ij}^\top\boldsymbol{\gamma}^*\right)-E\rho_\tau(y_{ij})\right]\omega_{ij}\\
&= \sum_{j=1}^n\sum_{i=1}^n E\rho_\tau\left(y_{ij}-\frac{1}{\sqrt{n}}\hat{b}_j\mathbf{x}_{ij}^\top\hat{\boldsymbol{\gamma}}^*+\frac{1}{\sqrt{n}}\hat{b}_j\mathbf{x}_{ij}^\top(\hat{\boldsymbol{\gamma}}^*-\boldsymbol{\gamma}^*)\right)\omega_{ij}\\
&\quad -\sum_{j=1}^n\sum_{i=1}^n E\rho_\tau\left(y_{ij}-\frac{1}{\sqrt{n}}\hat{b}_j\mathbf{x}_{ij}^\top\hat{\boldsymbol{\gamma}}^*+\frac{1}{\sqrt{n}}\hat{b}_j\mathbf{x}_{ij}^\top\hat{\boldsymbol{\gamma}}^*\right)\omega_{ij}\\
&= \sum_{j=1}^n\sum_{i=1}^n E\rho_\tau\left(y_i-\hat{a}_j-\hat{b}_j\mathbf{x}_{ij}^\top\hat{\boldsymbol{\gamma}}+\frac{1}{\sqrt{n}}\hat{b}_j\mathbf{x}_{ij}^\top(\hat{\boldsymbol{\gamma}}^*-\boldsymbol{\gamma}^*)\right)\omega_{ij}\\
&\quad -\sum_{j=1}^n\sum_{i=1}^n E\rho_\tau\left(y_i-\hat{a}_j-\hat{b}_j\mathbf{x}_{ij}^\top\hat{\boldsymbol{\gamma}}+\frac{1}{\sqrt{n}}\hat{b}_j\mathbf{x}_{ij}^\top\hat{\boldsymbol{\gamma}}^*\right)\omega_{ij}\\
&= \sum_{j=1}^n\sum_{i=1}^n E\rho_\tau\left(y_i-\hat{g}(\mathbf{x}_i^\top\hat{\boldsymbol{\gamma}}|\mathbf{x}^\top\boldsymbol{\gamma}_0 = x^\top\boldsymbol{\gamma}_0)+\frac{1}{\sqrt{n}}\hat{b}_j\mathbf{x}_{ij}^\top(\hat{\boldsymbol{\gamma}}^*-\boldsymbol{\gamma}^*)\right)\omega_{ij}\\
&\quad -\sum_{j=1}^n\sum_{i=1}^n E\rho_\tau\left(y_i-\hat{g}(\mathbf{x}_i^\top\hat{\boldsymbol{\gamma}}|\mathbf{x}^\top\boldsymbol{\gamma}_0 = x^\top\boldsymbol{\gamma}_0)+\frac{1}{\sqrt{n}}\hat{b}_j\mathbf{x}_{ij}^\top\hat{\boldsymbol{\gamma}}^*\right)\omega_{ij}\\
&= \sum_{j=1}^n\sum_{i=1}^n\left\{\varphi\left(\frac{1}{\sqrt{n}}\hat{b}_j\mathbf{x}_{ij}^\top(\hat{\boldsymbol{\gamma}}^*-\boldsymbol{\gamma}^*)|\mathbf{x}^\top\boldsymbol{\gamma}_0 = x^\top\boldsymbol{\gamma}_0\right)-\varphi\left(\frac{1}{\sqrt{n}}\hat{b}_j\mathbf{x}_{ij}^\top\hat{\boldsymbol{\gamma}}^*|\mathbf{x}^\top\boldsymbol{\gamma}_0 = x^\top\boldsymbol{\gamma}_0\right)\right\}(1+o_p(1))\omega_{ij}.
\end{aligned}$$

By the Taylor expansion, we have

$$\begin{aligned}
E(\mathcal{Q}_n(\boldsymbol{\gamma}^*)) &= -\sum_{j=1}^n\sum_{i=1}^n\left\{\varphi'\left(\frac{1}{\sqrt{n}}\hat{b}_j\mathbf{x}_{ij}^\top\hat{\boldsymbol{\gamma}}^*|\mathbf{x}^\top\boldsymbol{\gamma}_0 = x^\top\boldsymbol{\gamma}_0\right)\cdot\frac{1}{\sqrt{n}}\hat{b}_j\mathbf{x}_{ij}^\top\boldsymbol{\gamma}^*\right\}(1+o_p(1))\omega_{ij}\\
&\quad +\sum_{j=1}^n\sum_{i=1}^n\left\{\frac{1}{2n}\boldsymbol{\gamma}^{*\top}\left[\varphi''\left(\frac{1}{\sqrt{n}}\hat{b}_j\mathbf{x}_{ij}^\top\hat{\boldsymbol{\gamma}}^*|\mathbf{x}^\top\boldsymbol{\gamma}_0 = x^\top\boldsymbol{\gamma}_0\right)\hat{b}_j^2\mathbf{x}_{ij}\mathbf{x}_{ij}^\top\right]\boldsymbol{\gamma}^*\right\}(1+o_p(1))\omega_{ij}\\
&= -\frac{1}{\sqrt{n}}\sum_{j=1}^n\sum_{i=1}^n\left\{\varphi'(0|\mathbf{x}^\top\boldsymbol{\gamma}_0 = x^\top\boldsymbol{\gamma}_0)\hat{b}_j\mathbf{x}_{ij}^\top\omega_{ij}(1+o_p(1))\right\}\boldsymbol{\gamma}^*\\
&\quad +\frac{1}{2n}\boldsymbol{\gamma}^{*\top}\sum_{j=1}^n\sum_{i=1}^n\left\{\varphi''(0|\mathbf{x}^\top\boldsymbol{\gamma}_0 = x^\top\boldsymbol{\gamma}_0)\hat{b}_j^2\mathbf{x}_{ij}\mathbf{x}_{ij}^\top\omega_{ij}(1+o_p(1))\right\}\boldsymbol{\gamma}^*.
\end{aligned}$$

The last equation holds because of root-n assumption. Some simple calculation shows $\varphi'(0|\cdot) = 2(\tau - F_y(g(\mathbf{x}_i^\top\hat{\boldsymbol{\gamma}}|\cdot)))$ and $\varphi''(0|) = 2f_y(g(\mathbf{x}_i^\top\hat{\boldsymbol{\gamma}}|\cdot))$, where $F_y(\cdot)$ is the CDF and $f_y(\cdot)$ is the p.d.f. On the other hand, for each $E\rho'_\tau(y_{ij})$, $E\rho'_\tau(y_{ij}) = 2\tau(1-F_{y_{ij}}(y_{ij}))+2(\tau-1)F_{y_{ij}}(y_{ij}) = 2(\tau - F_{y_{ij}}(y_{ij}))$. So we have the following representation,

$$\mathcal{Q}_n(\boldsymbol{\gamma}^*) = -\frac{1}{\sqrt{n}}\left[\sum_{j=1}^n\sum_{i=1}^n\rho'_\tau(y_{ij})\hat{b}_j\mathbf{x}_{ij}^\top\omega_{ij}\right]\boldsymbol{\gamma}^*+\frac{1}{2n}\boldsymbol{\gamma}^{*\top}\left[\sum_{j=1}^n\sum_{i=1}^n 2f_y(.)(\hat{b}_j^2\mathbf{x}_{ij}\mathbf{x}_{ij}^\top)\omega_{ij}\right]\boldsymbol{\gamma}^*+o_p(1).$$

Due to root-n consistency assumption, we have $\rho'_\tau(y_{ij})$ given $\hat{a}_j$, $\hat{b}_j$ has asymptotic distribution of $\rho'_\tau(y-\theta_\tau(x))$, i.e. equal to 2τ with a probability of $1-\tau$, and equal to $2(\tau-1)$ with a probability of τ, and has asymptotic mean of zero and variance of $4\tau(1-\tau)$. Thus under Assumption (v), we have the following approximation, $\mathcal{Q}_n(\boldsymbol{\gamma}^*) = \frac{1}{2}\boldsymbol{\gamma}^{*\top}\mathbf{T}_n\boldsymbol{\gamma}^*+\mathbf{V}_n^\top\boldsymbol{\gamma}^*+o_p(1)$, where $\mathbf{T}_n = 2\mathbf{C}_1$, $\mathbf{V}_n = (4\tau(1-\tau))^{1/2}\mathbf{C}_0{}^{1/2}(\mathcal{Z}_n)$ and $\mathcal{Z}_n \to_D N(0, I)$. Finally, by Quadratic Approximation Lemma and Slutsky's Theorem, we have Theorem 3. Note that C_1 is not positive definite due to norm 1 identifiability constraint. Generalized inverse of C_1 is used instead. The proof of the Quadratic Approximation Lemma indicates that the positive definiteness of the matrix can be reduced into semi-positive definiteness, even to the existence of a generalized inverse. □

References

[1] W. Lin, K.B. Kulasekera, Indentifiability of single-index models and additive-index models, Biometrica 94 (2007) 496–501.

[2] R. Koenker, G.S. Bassett, Regression quantiles, Econometrica 46 (1978) 33–50.

[3] H. Ichimura, Semiparametric Least Squares (SLS) and weighted SLS estimation of single-index models, Journal of Econometrics 58 (1993) 71–120.

[4] J.L. Horowitz, W. Härdle, Direct semiparametric estimation of single-index models with discrete covariates, Journal of the American Statistical Association 91 (1996) 1623–1629.
[5] J.L. Horowitz, Semiparametric Methods in Econometrics, Springer, New York, 1998.
[6] R.J. Carroll, J. Fan, I. Gijbels, M.P. Wand, Generalized partially linear single-index models, Journal of the American Statistical Association 92 (1997) 477–489.
[7] Y. Yu, D. Ruppert, Penalized spline estimation for partially linear single-index models, Journal of the American Statistical Association 97 (2002) 1042–1054.
[8] Y. Xia, W. Härdle, Semi-parametric estimation of partially linear single index models, Journal of Multivariate Analysis 97 (2006) 1162–1184.
[9] P. Chaudhuri, K. Doksum, A. Samarov, On average derivative quantile regression, Annals of Statistics 25 (1997) 715–744.
[10] E. Kong, Y. Xia, Quantile estimation of a general single-index model, Working paper, 2010. http://arxiv.org/abs/0803.2474.
[11] K. Yu, M.C. Jones, Local linear quantile regression, Journal of the American Statistical Association 93 (1998) 228–237.
[12] R. Koenker, P. Ng, S. Portnoys, Quantile smoothing splines, Biometrika 81 (1994) 673–680.
[13] R. Koenker, Quantile Regression, Cambridge University Press, 2005.
[14] C.J. Stone, Consistent nonparametric regression, with discussion, Annals of Statistics 5 (1977) 595–645.
[15] P. Chaudhuri, Nonparametric estimates of regression quantiles and their local bahadur representation, Annals of Statistics 19 (1991) 760–777.
[16] J.G. De Gooijer, D. Zerom, On additive conditional quantiles with high-dimensional covariates, Journal of the American Statistical Association 98 (2003) 135–146.
[17] K. Yu, Z. Lu, Local linear additive quantile regression, Scandinavian Journal of Statistics 31 (2004) 333–346.
[18] J.L. Horowitz, S. Lee, Nonparametric estimation of an additive quantile regression model, Journal of the American Statistical Association 100 (2005) 1238–1249.
[19] W. Härdle, T.M. Stoker, Investing smooth multiple regression by the method of average derivatives, Journal of the American Statistical Association 84 (1989) 986–995.
[20] K. Yu, M.C. Jones, A comparison of local constant and local linear regression quantile estimation, Computational Statistics and Data Analysis 25 (1997) 159–166.
[21] Y. Xia, H. Tong, W.K. Li, L. Zhu, An adaptive estimation of dimension reduction space (with discussion), Journal of the Royal Statistical Society Series B 64 (2002) 363–410.
[22] D. Ruppert, S.J. Sheather, M.P. Wand, An effective bandwidth selector for local least squares regression, Journal of the American Statistical Association 90 (1995) 1257–1270.
[23] J. Fan, T.C. Hu, Y.K. Truong, Robust nonparametric function estimation, Scandinavian Journal of Statistics 21 (1994) 433–446.
[24] W. Härdle, P. Hall, H. Ichimura, Optimal smoothing in single-index models, Annals of Statistics 21 (1993) 157–178.
[25] J.D. Opsomer, D. Ruppert, A fully automated bandwidth selection for additive regression model, Journal of the American Statistical Association 93 (1998) 605–618.
[26] R. Koenker, J. Machado, Goodness of fit and related inference processes for quantile regression, Journal of the American Statistical Association 94 (1999) 1296–1310.
[27] N.L. Hjort, D. Pollard, Asymptotics for minimisers of convex processes, 1993, Preprint.