Statistica Sinica Preprint No: SS-2025-0069	
Title	Quantile Index Regression
Manuscript ID	SS-2025-0069
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202025.0069
Complete List of Authors	Yingying Zhang, Qianqian Zhu, Yuefeng Si and Guodong Li
Corresponding Authors	Qianqian Zhu
E-mails	zhu.qianqian@mail.shufe.edu.cn

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0069

Statistica Sinica

QUANTILE INDEX REGRESSION

Yingying Zhang[1], Qianqian Zhu[2], Yuefeng Si[3] and Guodong Li[3]

East China Normal University[1]

Shanghai University of Finance and Economics[2]

The University of Hong Kong[3]

Abstract: Estimating the structures at high or low quantiles has become an important subject and attracted increasing attention across numerous fields. However, due to data sparsity at tails, it usually is a challenging task to obtain reliable estimation, especially for high-dimensional data. This paper suggests a flexible parametric structure to tails, and this enables us to conduct the estimation at quantile levels with rich observations and then to extrapolate the fitted structures to far tails. The proposed model depends on some quantile indices and hence is called the quantile index regression. Moreover, the composite quantile regression method is employed to obtain non-crossing quantile estimators, and this paper further establishes their theoretical properties, including asymptotic normality for the case with low-dimensional covariates and non-asymptotic error bounds for that with high-dimensional covariates. Simulation studies and an empirical example are presented to illustrate the usefulness of the new model.

Key words and phrases: Asymptotic normality, High-dimensional analysis, Non-asymptotic property, Partially parametric model, Quantile regression.

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0069

1. Introduction

Quantile regression proposed by Koenker and Bassett (1978) has been widely used across various fields such as biological science, ecology, economics, finance, and machine learning, etc.; see, e.g., Cade and Noon (2003), Yu et al. (2003), Meinshausen and Ridgeway (2006), Linton and Xiao (2017) and Koenker (2017). More references on quantile regression can be found in the books of Koenker (2005) and Davino et al. (2014). Quantile regression has also been studied for high-dimensional data; see, e.g., Belloni and Chernozhukov (2011), Wang et al. (2012) and Zheng et al. (2015). On the other hand, due to practical needs, it is increasingly becoming a popular subject to estimate the structures at high or low quantiles, such as the risk of high loss for investments in finance (Kuester et al., 2006; Zheng et al., 2018), high tropical cyclone intensity and extreme waves in climatology (Elsner et al., 2008; Jagger and Elsner, 2008; Lobeto et al., 2021), and low infant birth weights in medicine (Abrevaya, 2001; Chernozhukov et al., 2022). It hence is natural to make inference at these extreme quantiles for high-dimensional data, while this is still an open problem.

There are two types of approaches in the literature to model the structures at tails. The first one is based on the conditional distribution function (CDF) of the response Y for a given set of covariates $\boldsymbol{X}$, and it is usually assumed to have a semiparametric structure at tails; see, e.g., Pareto-type structures in Beirlant and Goegebeur (2004) and Wang and Tsai (2009). While this method cannot provide conditional quantiles in explicit forms. Later, Noufaily and Jones (2013) considered a full parametric form, the generalized gamma distribution, to the CDF and then

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0069

inverted the fitted distribution into a conditional quantile distribution. However, as indicated in Racine and Li (2017), indirect inverse-CDF-based estimators may not be efficient in tail regions when the data has unbounded support.

The second approach is extremal quantile regression, which combines quantile regression with extreme value theory to estimate the conditional quantile at a very high or low level of τ_n^*, which satisfies $(1-\tau_n^*) = O(n)$ with n being the sample size; see Chernozhukov (2005). Specifically, this is a two-stage approach: (i.) performing the estimation at intermediate quantiles τ_n with $(1-\tau_n)^{-1} = o(n)$; and (ii.) extrapolating the fitted quantile structures to those at extreme quantiles by assuming the extreme value index that is associated to the tails of conditional distributions; see Wang et al. (2012) and Wang and Li (2013) for details. The key of this method is to make use of the feasible estimation at intermediate levels since there are relatively more observations. However, intermediate quantiles are also at the far tails, and the corresponding data points may still not be rich enough for the case with many covariates.

In order to handle the case with high-dimensional covariates, along the lines of extremal quantile regression, this paper suggests to conduct estimation at quantile levels with much richer observations, say some fixed levels around τ_0, and then extrapolate the estimated results to extreme quantiles by fully or partially assuming a form of conditional quantile functions on $[\tau_0, 1)$. Note that there exist many quantile functions, which have explicit forms, such as the generalized lambda and Burr XII distributions (Gilchrist, 2000). Especially the generalized lambda distribution can provide a very accurate approximation to some Pareto-type and extreme

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0069

value distributions, as well as some commonly used distributions such as Gaussian distribution (Vasicek, 1976; Gilchrist, 2000). These flexible parametric forms can be assumed to the quantile function on $[\tau_0, 1)$, and the drawback of inverting a distribution function hence can be avoided.

Specifically, for a predetermined interval $\mathcal{I} \subset (0, 1)$, the quantile function of response Y is assumed to have an explicit form of $Q(\tau, \boldsymbol{\theta})$ for each level $\tau \in \mathcal{I}$, up to unknown parameters or indices $\boldsymbol{\theta}$. By further letting $\boldsymbol{\theta}$ be a function of covariates $\mathbf{X}$, we then can define the conditional quantile function as follows:

$$Q_Y(\tau|\mathbf{X}) = \inf\{y : F_Y(y|\mathbf{X}) \geq \tau\} = Q(\tau, \boldsymbol{\theta}(\mathbf{X})), \quad \tau \in \mathcal{I}, \tag{1.1}$$

where $F_Y(\cdot|\mathbf{X})$ is the distribution of Y conditional on $\mathbf{X}$, and $\boldsymbol{\theta}(\mathbf{X})$ is a d-dimensional parametric function. Since $\boldsymbol{\theta}(\mathbf{X})$ can be referred to d indices, model (1.1) can then be called the quantile index regression (QIR) for simplicity. The proposed model has a form of single- or multi-index quantile regression models (Zhang et al., 2020). The partial or full parametric form used in model (1.1) makes possible the prediction at levels beyond those for estimation, while the single- or multi-index model conducts the estimation and prediction at the same quantile levels. In practice, to handle high quantiles, we may take $\mathcal{I} = [\tau_0, 1)$ with a fixed value of τ_0 and then conduct a conditional quantile regression (CQR) estimation for model (1.1) at levels within $\mathcal{I}$ but with richer observations. Subsequently, the fitted QIR model can be used to predict extreme quantiles. More importantly, since the estimation is conducted at fixed quantile levels, there is no difficulty to handle the case with high-dimensional covariates. In addition, comparing with the aforementioned two types of approaches in the literature, the proposed method can

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0069

not only estimate quantile regression functions effectively, but also forecast extreme quantiles directly. Finally, the QIR model naturally yields quantile estimators that guarantee non-crossing conditional quantiles since its quantile function is nondecreasing with respect to τ.

The proposed model is introduced in details at Section 2, and the three main contributions can be summarized below:

(a) When conducting the CQR estimation, we encounter the first challenge on model identification, and this problem is carefully studied for three commonly used quantile functions in Section 2.2. The model misspecification problem is also investigated.

(b) Section 2.3 derives the asymptotic normality of CQR estimators for the case with low-dimensional covariates. This is a challenging task since the corresponding objective function is non-convex and non-differentiable, and we overcome the difficulty by adopting the bracketing method in Pollard (1985).

(c) Section 2.4 establishes non-asymptotic properties of a regularized high-dimensional estimation. This is also not trivial due to the problem at (b).

The rest of this paper is organized as follows. Section 3 discusses some implementation issues in searching for these estimators. Numerical studies, including simulation experiments and a real analysis, are given in Sections 4 and 5, and Section 6 provides a short conclusion and discussion. All technical details are relegated to the Supplementary Material.

For the sake of convenience, this paper denotes vectors and matrices by boldface letters,

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0069

e.g., $\mathbf{X}$ and $\mathbf{Y}$, and denotes scalars by regular letters, e.g., X and Y. In addition, for any two real-valued sequences $\{a_n\}$ and $\{b_n\}$, denote $a_n \gtrsim b_n$ (or $a_n \lesssim b_n$) if there exists a constant c such that $a_n \geq cb_n$ (or $a_n \leq cb_n$) for all n, and denote $a_n \asymp b_n$ if $a_n \gtrsim b_n$ and $a_n \lesssim b_n$. For a generic vector $\mathbf{X}$ and matrix $\mathbf{Y}$, let $\|\mathbf{X}\|$, $\|\mathbf{X}\|_1$ and $\|\mathbf{Y}\|_{\mathrm{F}}$ represent the Euclidean norm, ℓ_1-norm and Frobenius norm, respectively.

2. Quantile index regression

2.1 Quantile index regression model

Consider a response Y and a p-dimensional vector of covariates $\mathbf{X} = (X_1, ..., X_p)'$. We then rewrite the quantile function of Y conditional on $\mathbf{X}$ at (1.1) with an explicit form of $\boldsymbol{\theta}(\mathbf{X}, \boldsymbol{\beta})$,

$$Q_Y(\tau|\mathbf{X}) = Q(\tau, \boldsymbol{\theta}(\mathbf{X}, \boldsymbol{\beta})), \quad \tau \in \mathcal{I}, \tag{2.1}$$

where $\mathcal{I} \subset (0, 1)$ is an interval or the union of multiple disjoint intervals, the d indices are included in $\boldsymbol{\theta}(\mathbf{X}, \boldsymbol{\beta}) = (\theta_1(\mathbf{X}, \boldsymbol{\beta}), \cdots, \theta_d(\mathbf{X}, \boldsymbol{\beta}))'$, $\boldsymbol{\beta} = (\boldsymbol{\beta}_1', ..., \boldsymbol{\beta}_d')'$, $\boldsymbol{\beta}_j = (\beta_{j1}, \cdots, \beta_{jp})'$, $\theta_j(\mathbf{X}, \boldsymbol{\beta}) = g_j(\mathbf{X}'\boldsymbol{\beta}_j)$, the user-specified link functions $g_j^{-1}(\cdot)$s with $1 \leq j \leq d$ are all monotonic, and the intercept term can be included by letting $X_1 = 1$. We call model (2.1) the quantile index regression (QIR) for simplicity, and its flexibility is mainly determined by the explicit form of $Q(\tau, \boldsymbol{\theta})$, as well as the corresponding link functions. Two examples of $Q(\cdot, \cdot)$ below are first introduced to illustrate the new model.

Example 1. Consider the location shift model, $Q(\tau, \theta) = \theta + Q_\Phi(\tau)$, for all $\tau \in [\tau_0, 1)$, where

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0069

$\tau_0 \in (0,1)$ is a fixed level, $\theta \in \mathbb{R}$ is the location index and $Q_\Phi(\tau)$ is the quantile function of standard normality. Under the identity link function, $\theta(\mathbf{X}, \boldsymbol{\beta}) = \mathbf{X}'\boldsymbol{\beta}$, we can construct a linear quantile regression model at τ_0, and then a prediction can be made at any level of $\tau \in (\tau_0, 1)$.

Example 2. Consider a more general form of $Q(\tau, \boldsymbol{\theta}) = \theta_1 + \theta_2 S(\tau, \theta_3)$, and it corresponds to the Tukey lambda, generalized extreme value, or generalized Pareto distribution when $S(\tau, \theta_3)$ has the form of

$$\frac{\tau^{\theta_3} - (1-\tau)^{\theta_3}}{\theta_3}, \quad \frac{1 - (-\log \tau)^{\theta_3}}{\theta_3} \quad \text{or} \quad \frac{1 - (1-\tau)^{\theta_3}}{\theta_3},$$

respectively, where $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3)'$, and $\theta_1 \in \mathbb{R}$, $\theta_2 > 0$ and $\theta_3 \neq 0$ are the location, scale and tail indices, respectively; see Gilchrist (2000) and de Haan and Ferreira (2006).

The Tukey lambda distribution (Vasicek, 1976) can well approximate many commonly used distributions, such as Weibull, uniform and Cauchy distributions. In the meanwhile, the generalized lambda distribution (Gilchrist, 2000; Fournier et al., 2007) has a form of

$$Q(\tau, \theta) = \theta_1 + \theta_2 \left\{ \frac{\tau^{\theta_3} - 1}{\theta_3} - \frac{(1-\tau)^{\theta_4} - 1}{\theta_4} \right\},$$

where the indices θ_3 and θ_4 control the right and left tails, respectively, and it reduces to the Tukey lambda distribution when $\theta_3 = \theta_4$. As a result, the generalized lambda distribution can be considered if we focus on the quantiles with the full range, i.e. $\mathcal{I} \subseteq (0,1)$, while the Tukey lambda distribution may be a better choice if our interest is on the quantiles at one side only, i.e. $\mathcal{I} \subseteq (0, 0.5)$ or $(0.5, 1)$. Moreover, the generalized extreme value distribution (GEVD) can

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0069

depict the extreme behaviour of properly normalized maxima of independent and identically distributed random variables (Fisher and Tippett, 1928), and it hence can be used for the right tail, i.e. $\mathcal{I} \subseteq (0.5, 1)$. Finally, the generalized Pareto distribution (GPD) can model exceedances over a threshold (Pickands, 1975), and it can be used to model the right tail only.

Remark 1. Model (2.1) has a form of single- or multi-index models (Zhang et al., 2020). This motivates us to consider a special multiple-index quantile regression model with $Q_Y(\tau|\mathbf{X})$ being a function of τ and $\mathbf{X}'\boldsymbol{\beta}_j$ with $1 \leq j \leq d$, i.e., the slopes are independent of τ. Since, for a monotonic link function $g^{-1}(\cdot)$, it holds that $g^{-1}\{g(x)\} = x$ for any $x \in \mathbb{R}$, we have $Q_Y(\tau|\mathbf{X}) = q(\tau, \boldsymbol{\theta}(\mathbf{X}, \boldsymbol{\beta}))$ with $\boldsymbol{\theta}(\mathbf{X}, \boldsymbol{\beta}) = (g_1(\mathbf{X}'\boldsymbol{\beta}_1), \cdots, g_d(\mathbf{X}'\boldsymbol{\beta}_d))'$, where $q(\cdot, \cdot)$ is an unknown function, and the link functions $g_j^{-1}(\cdot)$s with $1 \leq j \leq d$ are the ones in model (2.1). A nonparametric method can then be employed to estimate $q(\cdot, \cdot)$ as in the literature, while it may not be able to support the extrapolation emphasized in this paper.

2.2 Composite quantile regression estimation

Denote by $\{(Y_i, \mathbf{X}_i')', i = 1, ..., n\}$ the observed data, and they are independent and identically distributed $(i.i.d.)$ samples of random vector $(Y, \mathbf{X})$, where Y_i is the response, $\mathbf{X}_i = (X_{i1}, ..., X_{ip})'$ contains p covariates, and n is the number of observations. Let $\tau_1 \leq \tau_2 \leq \ldots \leq \tau_K$ be K fixed quantile levels, where $\tau_k \in \mathcal{I}$ for all $1 \leq k \leq K$. To achieve higher efficiency,

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0069

we consider the composite quantile regression (CQR) estimator below,

$$\widehat{\boldsymbol{\beta}}_n = \arg\min_{\boldsymbol{\beta}} L_n(\boldsymbol{\beta}) \text{ and } L_n(\boldsymbol{\beta}) = \sum_{k=1}^{K}\sum_{i=1}^{n} \rho_{\tau_k}\{Y_i - Q(\tau_k, \boldsymbol{\theta}(\mathbf{X}_i, \boldsymbol{\beta}))\}, \tag{2.2}$$

where $\rho_\tau(x) = x\{\tau - I(x < 0)\}$ is the quantile check function; see Zou and Yuan (2008) and Kai et al. (2010). To study the asymptotic properties of $\widehat{\boldsymbol{\beta}}_n$, we consider

$$\boldsymbol{\beta}_0 = (\boldsymbol{\beta}_{01}', ..., \boldsymbol{\beta}_{0d}')' = \arg\min_{\boldsymbol{\beta}} \bar{L}(\boldsymbol{\beta}) \text{ and } \bar{L}(\boldsymbol{\beta}) = E\left[\sum_{k=1}^{K} \rho_{\tau_k}\{Y - Q(\tau_k, \boldsymbol{\theta}(\mathbf{X}, \boldsymbol{\beta}))\}\right],$$

where $\bar{L}(\boldsymbol{\beta})$ is the population loss function.

We first investigate the identification problem of CQR estimation at (2.2) since $\boldsymbol{\beta}_0$ may not be unique for the QIR model at (2.1). For the sake of better illustration, let us consider the case without covariates $\mathbf{X}$, and it is equivalent to purely estimate $\boldsymbol{\theta}$. We hence requires that two different values of $\boldsymbol{\theta}$ cannot yield the same quantile function $Q(\tau_k, \boldsymbol{\theta})$ across all K levels. In other words, if there exists $\boldsymbol{\theta} \neq \boldsymbol{\theta}^*$ that yield $Q(\tau_k, \boldsymbol{\theta}) = Q(\tau_k, \boldsymbol{\theta}^*)$ for all K quantiles, then $\boldsymbol{\theta}$ and $\boldsymbol{\theta}^*$ are not identifiable. As a result, to guarantee that $\boldsymbol{\beta}_0$ is the unique minimizer of the population loss, we make the following assumption on the quantile function $Q(\tau, \boldsymbol{\theta})$.

Assumption 1. *For any two index vectors $\boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_2$, there exists at least one $1 \leq k \leq K$ such that $Q(\tau_k, \boldsymbol{\theta}_1) \neq Q(\tau_k, \boldsymbol{\theta}_2)$.*

Intuitively, for any quantile function $Q(\tau, \boldsymbol{\theta})$, one can always make Assumption 1 hold by increasing the number of quantile levels K, while it may also depend on the structure of quantile functions and number of indices. It hence is of interest to know the minimum number of K,

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0069

which can guarantee Assumption 1, and the following proposition partially solves the problem by giving an answer to the three distributions in Example 2.

Proposition 1. *(i) For the Tukey lambda distribution in Example 2 with $\theta_3 < 1$, Assumption 1 holds if and only if $K \geq 3$ when $\mathcal{I} \subset (0, 0.5)$ or $(0.5, 1)$, and Assumption 1 holds if $K \geq 4$ when $\mathcal{I} \subset (0, 1)$; (ii) For the generalized extreme value and generalized Pareto distributions in Example 2, Assumption 1 holds if and only if $K \geq 3$.*

Assumption 1, together with an additional condition on covariates $\mathbf{X}$, allows us to show that $\boldsymbol{\beta}_0$ is the unique minimizer of $\bar{L}(\boldsymbol{\beta})$, and hence the identification problem is solved.

Theorem 1. *Suppose that $E(\mathbf{X}\mathbf{X}')$ is a $p \times p$ finite and positive definite matrix. If Assumption 1 holds, then $\boldsymbol{\beta}_0$ is the unique minimizer of $\bar{L}(\boldsymbol{\beta})$.*

Note that model (2.1) partially assumes a parametric form to the conditional quantile function $Q_Y(\tau|\mathbf{X})$, while it may be misspecified. It can be verified that the value of $\boldsymbol{\beta}_0$ satisfies

$$\sum_{k=1}^{K} E\left(\left[F_{Y|\mathbf{X}} \left\{ Q(\tau_k, \boldsymbol{\theta}(\mathbf{X}, \boldsymbol{\beta}_0)) \right\} - F_{Y|\mathbf{X}} \left\{ Q_Y(\tau_k|\mathbf{X}) \right\} \right] \frac{\partial Q(\tau_k, \boldsymbol{\theta}(\mathbf{X}, \boldsymbol{\beta}_0))}{\partial \boldsymbol{\beta}} \right) = 0,$$

where $F_{Y|\mathbf{X}}(\cdot) = F_Y(\cdot|\mathbf{X})$ is the conditional distribution function, and it holds that $Q_Y(\tau|\mathbf{X}) = q(\tau, \boldsymbol{\theta}(\mathbf{X}, \boldsymbol{\beta}^*))$ for the example in Remark 1, where $\boldsymbol{\beta}^*$ is the true parameters. As a result, for the case with misspecification, $\boldsymbol{\beta}_0$ depends on quantile levels of τ_k's, including the total number and their placement, and it is chosen such that $Q(\tau, \boldsymbol{\theta}(\mathbf{X}, \boldsymbol{\beta}_0))$ can well approximate $Q_Y(\tau|\mathbf{X})$ at the K levels. However, such approximation has no guarantee for the case with extrapolation in this paper, and hence we may try to place these τ_k's near to the target levels in real applications.

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0069

Remark 2. This paper employs the CQR mainly for two reasons. First, to solve the identification problem, we need at least three quantile levels for the three distributions in Proposition 1. In the meanwhile, the indices $\boldsymbol{\theta}(\boldsymbol{X}, \boldsymbol{\beta})$ are independent of τ, and hence the CQR can be used to aggregate information across multiple levels to improve the estimation efficiency, especially by incorporating the levels with much richer observations.

2.3 Low-dimensional asymptotic properties

In this and the following subsections, we focus on cases without model misspecification, and assume that the true parameter vector $\boldsymbol{\beta}_0$ satisfies the identification condition in Assumption 1.

We first consider the consistency and asymptotic normality of $\widehat{\boldsymbol{\beta}}_n$ for the case with low-dimensional covariates, i.e., p is fixed. Denote by Θ the parameter space, which is a compact set of $\mathbb{R}^{dp}$, and suppose that the true parameter vector $\boldsymbol{\beta}_0$ is an interior point of Θ. Denote

$$\Omega_0 = \sum_{k'=1}^{K}\sum_{k=1}^{K} \min(\tau_k, \tau_{k'}) \left\{1 - \max(\tau_k, \tau_{k'})\right\} E\left\{ \frac{\partial Q(\tau_k, \boldsymbol{\theta}(\mathbf{X}, \boldsymbol{\beta}_0))}{\partial \boldsymbol{\beta}} \frac{\partial Q(\tau_{k'}, \boldsymbol{\theta}(\mathbf{X}, \boldsymbol{\beta}_0))}{\partial \boldsymbol{\beta}'} \right\}$$

and

$$\Omega_1 = \sum_{k=1}^{K} E\left[f_Y\left\{Q(\tau_k, \boldsymbol{\theta}(\mathbf{X}, \boldsymbol{\beta}_0))|\mathbf{X}\right\} \frac{\partial Q(\tau_k, \boldsymbol{\theta}(\mathbf{X}, \boldsymbol{\beta}_0))}{\partial \boldsymbol{\beta}} \frac{\partial Q(\tau_k, \boldsymbol{\theta}(\mathbf{X}, \boldsymbol{\beta}_0))}{\partial \boldsymbol{\beta}'} \right].$$

Assumption 2. *For all* $1 \le k \le K$, *it holds that*

$$E\left\| \frac{\partial Q(\tau_k, \boldsymbol{\theta}(\mathbf{X}, \boldsymbol{\beta}_0))}{\partial \boldsymbol{\beta}} \right\|^3 < \infty \quad \textit{and} \quad E \sup_{\boldsymbol{\beta}\in\Theta} \left\| \frac{\partial^2 Q(\tau_k, \boldsymbol{\theta}(\mathbf{X}, \boldsymbol{\beta}))}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} \right\|_{\mathrm{F}}^2 < \infty.$$

Assumption 3. *The conditional density* $f_Y(y|\mathbf{X})$ *is bounded and continuous uniformly for all* $\mathbf{X}$.

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0069

Theorem 2. *Suppose that* $E\{\max_{1\le k\le K}\sup_{\boldsymbol{\beta}\in\Theta}\|\partial Q(\tau_k,\boldsymbol{\theta}(\mathbf{X},\boldsymbol{\beta}))/\partial\boldsymbol{\beta}\|\}<\infty$. *If the conditions of Theorem 1 hold, then* $\widehat{\boldsymbol{\beta}}_n\to\boldsymbol{\beta}_0$ *in probability as* $n\to\infty$.

Theorem 3. *Suppose that Assumptions 2 and 3 hold, and* Ω_1 *is positive definite. If the conditions of Theorem 2 hold, then* $\sqrt{n}(\widehat{\boldsymbol{\beta}}_n-\boldsymbol{\beta}_0)\to N(\mathbf{0},\Omega_1^{-1}\Omega_0\Omega_1^{-1})$ *in distribution as* $n\to\infty$.

The moment condition in Theorem 2 allows us to adopt the uniform consistency result of Newey and McFadden (1994) to prove the consistency. Assumption 2 is used to establish the root-n consistency in the technical proof of Theorem 3 (Zhu and Ling, 2011). Assumption 3 is commonly used in the literature of quantile regression (Koenker, 2005; Belloni et al., 2019), and it can be relaxed by providing more complicated and lengthy technical details (Kato et al., 2012; Chernozhukov et al., 2015; Galvao and Kato, 2016). Moreover, the objective function $L_n(\boldsymbol{\beta})$ is non-convex and non-differentiable, and this makes it challenging to establish the asymptotic normality of $\widehat{\boldsymbol{\beta}}_n$. We overcome the difficulty by using the bracketing method in Pollard (1985).

We may choose the optimal τ_k's by minimizing the asymptotic variance in Theorem 3. However, it has a complicated form, and we even cannot further simplify it under the model setting at (2.1). This paper simply places τ_k's with equal distance; see Section 3 for details. Moreover, to estimate the asymptotic variance matrix in Theorem 3, we first apply the nonparametric method in Hendricks and Koenker (1991) to estimate the quantities of $f_Y\{Q(\tau_k,\boldsymbol{\theta}(\mathbf{X},\boldsymbol{\beta}_0))|\mathbf{X}\}$ with $1\le k\le K$, and then matrix Ω_1 can be approximated by plugging-in the estimated conditional density and then the sample averaging with $\boldsymbol{\beta}_0$ replaced by $\widehat{\boldsymbol{\beta}}_n$. Moreover, matrix Ω_0 can be estimated by the sample averaging, and hence the asymptotic variance matrix.

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0069

After obtaining the estimator $\widehat{\boldsymbol{\beta}}_n$, one can use the quantity of $Q(\tau, \theta(\mathbf{X}, \widehat{\boldsymbol{\beta}}_n))$ to predict the quantile structure at any level $\tau \in \mathcal{I}$, which can be an extreme quantile level if it is included in $\mathcal{I}$. Note that, for each fixed θ, the assumed parametric form $Q(\tau, \boldsymbol{\theta})$ is increasing with respect to τ. As a result, for a new observation of covariates $\mathbf{X}$, $Q(\tau, \boldsymbol{\theta}(\mathbf{X}, \widehat{\boldsymbol{\beta}}_n))$ is also increasing with respect to τ, and hence it has the non-crossing property. The corresponding theoretical justification can be established by directly applying the delta-method (van der Vaart, 1998, Chapter 3).

Corollary 1. *Suppose that the conditions of Theorem 3 are satisfied. Then, for any $\tau^* \in \mathcal{I}$,*

$$\sqrt{n}\{Q(\tau^*, \boldsymbol{\theta}(\mathbf{X}, \widehat{\boldsymbol{\beta}}_n)) - Q(\tau^*, \boldsymbol{\theta}(\mathbf{X}, \boldsymbol{\beta}_0))\} \to N(\mathbf{0}, \boldsymbol{\delta}'\Omega_1^{-1}\Omega_0\Omega_1^{-1}\boldsymbol{\delta})$$

in distribution as $n \to \infty$, where $\boldsymbol{\delta} = E\{\partial Q(\tau^*, \boldsymbol{\theta}(\mathbf{X}, \boldsymbol{\beta}_0))/\partial \boldsymbol{\beta}\} \in \mathbb{R}^{dp}$.

2.4 High-dimensional regularized estimation

This subsection considers the case with high-dimensional covariates, i.e., $p \gg n$, and the true parameter vector $\boldsymbol{\beta}_0$ is assumed to be s-sparse, i.e., the number of nonzero elements in $\boldsymbol{\beta}_0$ is no more than $s > 0$. A regularized CQR estimation can then be introduced,

$$\widetilde{\boldsymbol{\beta}}_n = \arg\min_{\boldsymbol{\beta} \in \Theta} n^{-1} L_n(\boldsymbol{\beta}) + \sum_{j=1}^{d} p_\lambda(\boldsymbol{\beta}_j), \tag{2.3}$$

where Θ is given in Theorem 4, p_λ is a penalty function, and it depends on a tuning (regularization) parameter $\lambda \in \mathbb{R}^+$ with $\mathbb{R}^+ = (0, \infty)$.

Consider the loss function $L_n(\boldsymbol{\beta}) = \sum_{k=1}^{K} \sum_{i=1}^{n} \rho_{\tau_k}\{Y_i - Q(\tau_k, \boldsymbol{\theta}(\mathbf{X}_i, \boldsymbol{\beta}))\}$ defined in (2.2), and $Q(\tau_k, \boldsymbol{\theta}(\mathbf{X}_i, \boldsymbol{\beta}))$ usually is nonconvex with respect to $\boldsymbol{\beta}$. As a result, $L_n(\boldsymbol{\beta})$ will be non-

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0069

convex although the check loss $\rho_\tau(\cdot)$ is convex, and there is no more harm to use nonconvex penalty functions. Specifically, we consider the component-wise penalization,

$$\sum_{j=1}^{d} p_\lambda(\boldsymbol{\beta}_j) = \sum_{j=1}^{d}\sum_{l=1}^{p} p_\lambda(\beta_{jl}),$$

where $p_\lambda(\cdot)$ is possibly nonconvex and satisfies the following assumption.

Assumption 4. *The univariate function $p_\lambda(\cdot)$ satisfies the following conditions: (i) it is symmetric around zero with $p_\lambda(0) = 0$; (ii) it is nondecreasing on the nonnegative real line; (iii) the function $p_\lambda(t)/t$ is nonincreasing with respect to $t \in \mathbb{R}^+$; (iv) it is differentiable for all $t \neq 0$ and subdifferentiable at $t = 0$, with $\lim_{t\to 0^+} p'_\lambda(t) = \lambda L$ and L being a constant; (v) there exists $\mu > 0$ such that $p_{\lambda,\mu} = p_\lambda(t) + \frac{\mu^2}{2}t^2$ is convex.*

The above is the μ-amenable assumption given in Loh and Wainwright (2015) and Loh (2017), and the penalty function is required not too far from the convexity. Note that the popular penalty functions, including SCAD (Fan and Li, 2001) and MCP (Zhang, 2010), satisfy the above properties.

In the literature of nonconvex penalized quantile regression, Jiang et al. (2012) studied nonlinear quantile regressions with SCAD regularizer from the asymptotic viewpoint, while it can only handle the case with $p = o(n^{1/3})$. Wang et al. (2012) and Sherwood et al. (2016) considered the case that p grows exponentially with n, and their proving techniques heavily depend on the condition that the loss function should be represented as a difference of the two convex functions. However, $L_n(\boldsymbol{\beta})$ does not meet this requirement since quantile function

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0069

$Q(\tau, \boldsymbol{\theta})$ can be nonconvex.

On the other hand, non-asymptotic properties recently have attracted considerable attention in the theories of high-dimensional analysis; see, e.g., Belloni and Chernozhukov (2011); Sivakumar and Banerjee (2017); Pan and Zhou (2021). This subsection attempts to study them for our proposed quantile estimators, while it is a nontrivial task since existing results only focused on linear quantile regression. Loh and Wainwright (2015) and Loh (2017) studied the non-asymptotic properties for M-estimators with both nonconvex loss and regularizers, while they required the loss function to be twice differentiable. The technical proofs in the Supplementary Material follow the framework in Loh and Wainwright (2015) and Loh (2017), and some new techniques are developed to tackle the nondifferentiability of the quantile check function.

Let $\boldsymbol{\theta}(\boldsymbol{\gamma}) = (g_1(\gamma_1), \ldots, g_d(\gamma_d))'$ with $g_j^{-1}(\cdot)$s being link functions and $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_d)'$, and we can then denote $Q(\tau, \boldsymbol{\gamma}) := Q(\tau, \boldsymbol{\theta}(\boldsymbol{\gamma}))$. Moreover, by letting $\gamma_j(\mathbf{X}, \boldsymbol{\beta}) = \mathbf{X}\boldsymbol{\beta}_j$ for $1 \le j \le d$ and $\boldsymbol{\gamma}(\mathbf{X}, \boldsymbol{\beta}) = (\gamma_1(\mathbf{X}, \boldsymbol{\beta}), \ldots, \gamma_d(\mathbf{X}, \boldsymbol{\beta}))'$, we can further denote $Q(\tau, \boldsymbol{\gamma}(\mathbf{X}, \boldsymbol{\beta})) := Q(\tau, \boldsymbol{\theta}(\mathbf{X}, \boldsymbol{\beta}))$.

Assumption 5. *Quantile function $Q(\tau, \boldsymbol{\gamma})$ is differentiable with respect to $\boldsymbol{\gamma}$, and there exist two positive constants L_Q and C_X such that $\max_{1 \le k \le K} \|\partial Q(\tau_k, \boldsymbol{\gamma})/\partial \boldsymbol{\gamma}\| \le L_Q$ and $\|\mathbf{X}\|_\infty \le C_X$.*

The differentiable assumption of quantile functions allows us to use the Lipschitz property and multivariate contraction theorem. The boundedness of covariates is to assure that the bounded difference inequality can be used, and it can be relaxed with more complicated and lengthy technical details (Wang and He, 2022).

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0069

Denote by $\mathcal{B}_R(\boldsymbol{\beta}_0) = \{\boldsymbol{\beta} \in \mathbb{R}^{dp} : \|\boldsymbol{\beta} - \boldsymbol{\beta}_0\| \leq R\}$ the Euclidean ball centered at $\boldsymbol{\beta}_0$ with radius $R > 0$, and let $\lambda_{\min}(\boldsymbol{\beta})$ be the smallest eigenvalue of matrix

$$\Omega_2(\boldsymbol{\beta}) = \sum_{k=1}^{K} E\left\{ \frac{\partial Q(\tau_k, \boldsymbol{\theta}(\mathbf{X}, \boldsymbol{\beta}))}{\partial \boldsymbol{\beta}} \frac{\partial Q(\tau_k, \boldsymbol{\theta}(\mathbf{X}, \boldsymbol{\beta}))}{\partial \boldsymbol{\beta}'} \right\}.$$

Assumption 6. *There exists a fixed $R > 0$ such that $\lambda_{\min}^0 = \inf_{\boldsymbol{\beta} \in \mathcal{B}_R(\boldsymbol{\beta}_0)} \lambda_{\min}(\boldsymbol{\beta}) > 0$. Moreover, $f_{\min} = \min_{1 \leq k \leq K} \inf_{\boldsymbol{\beta} \in \mathcal{B}_R(\boldsymbol{\beta}_0)} f_Y \{Q(\tau_k, \boldsymbol{\theta}(\mathbf{X}, \boldsymbol{\beta})) | \mathbf{X}\} > 0$, and $\alpha = 0.5 f_{\min} \lambda_{\min}^0 > \mu/4$.*

The above assumption guarantees that the population loss $\bar{L}(\boldsymbol{\beta}) = E\{n^{-1} L_n(\boldsymbol{\beta})\}$ is strongly convex around the true parameter vector $\boldsymbol{\beta}_0$. Specifically, let $\bar{\mathcal{E}}(\Delta) = \bar{L}(\boldsymbol{\beta}_0 + \Delta) - \bar{L}(\boldsymbol{\beta}_0) - \Delta' \partial \bar{L}(\boldsymbol{\beta}_0)/\partial \boldsymbol{\beta}$ be the first-order Taylor expansion. Then, by Assumption 6, we have that $\bar{\mathcal{E}}(\Delta) \geq 0.5 f_{\min} \lambda_{\min}^0 \|\Delta\|^2$ for all Δ such that $\|\Delta\| \leq R$; see Lemma S5 in the Supplementary Material for details. We next obtain the non-asymptotic estimation bound of $\widetilde{\boldsymbol{\beta}}_n$.

Theorem 4. *Suppose that Assumptions 1 and 4-6 hold, $n \gtrsim \log p$ and $\lambda \gtrsim \sqrt{\log p / n}$. Then the minimizer $\widetilde{\boldsymbol{\beta}}_n$ of (2.3) with $\Theta = \mathcal{B}_R(\boldsymbol{\beta}_0)$ satisfies the error bounds of*

$$\|\widetilde{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0\| \leq \frac{6L\sqrt{s}\lambda}{4\alpha - \mu} \quad \text{and} \quad \|\widetilde{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0\|_1 \leq \frac{24Ls\lambda}{4\alpha - \mu},$$

with probability at least $1 - c_1 p^{-c_2} - K \max(\log p, \log n) p^{-c^2}$ for any $c > 1$, where α is defined in Assumption 6, μ and L are defined in Assumption 4, s is the number of nonzero elements in $\boldsymbol{\beta}_0$, and the constants c_1 and $c_2 > 0$ are given in Lemma S4 of the Supplementary Material.

In practice, we can choose $\lambda \asymp \sqrt{\log p / n}$, and it then holds that $\|\widetilde{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0\| \lesssim \sqrt{s \log p / n}$, which has the standard rate of error bounds; see, e.g., Loh (2017). Moreover, for the predicted

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0069

conditional quantile of $Q(\tau^*, \boldsymbol{\theta}(\mathbf{X}, \widetilde{\boldsymbol{\beta}}_n))$ at any level $\tau^* \in \mathcal{I}$, it can be readily verified that $|Q(\tau^*, \boldsymbol{\theta}(\mathbf{X}, \widetilde{\boldsymbol{\beta}}_n)) - Q(\tau^*, \boldsymbol{\theta}(\mathbf{X}, \boldsymbol{\beta}_0))|$ has the same convergence rate as $\|\widetilde{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0\|$. Finally, the above theorem requires the minimization (2.3) to be conducted in $\Theta = \mathcal{B}_R(\boldsymbol{\beta}_0)$, which is unknown but fixed. This enables us to solve the problem by conducting a random initialization in optimizing algorithms.

3. Implementation issues

3.1 Optimizing algorithms

This subsection provides algorithms to search for the CQR estimator at (2.2) and regularized estimator at (2.3).

For the CQR estimation without penalty at (2.2), we employ the commonly used gradient descent algorithm to search for estimators, and the $(r+1)$th update is given by

$$\boldsymbol{\beta}^{(r+1)} = \boldsymbol{\beta}^{(r)} - \eta^{(r)} \nabla L_n(\boldsymbol{\beta}^{(r)}),$$

where $\boldsymbol{\beta}_n^{(r)}$ is from the rth iteration, and $\eta^{(r)}$ is the step size. Note that the quantile check loss is nondifferentiable at zero, and $\nabla L_n(\boldsymbol{\beta}^{(r)})$ in the above refers to the subgradient (Moon et al., 2021) instead. In practice, too small step size will cause the algorithm to converge slowly, while too large step size may cause the algorithm to diverge. We choose the step size by the backtracking line search (BLS) method, which is shown to be simple and effective; see Bertsekas (2016). Specifically, the algorithm starts with a large step size and, at $(r+1)$th update, it is reduced

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0069

by keeping multiplying a fraction of b until $L_n(\boldsymbol{\beta}^{(r+1)}) - L_n(\boldsymbol{\beta}^{(r)}) < -a\eta^{(r)}\|\nabla L_n(\boldsymbol{\beta}^{(r)})\|_2^2$, where a is another hyper-parameter. The simulation experiments in Section 4 work well with the setting of $(a, b) = (0.3, 0.5K^{-1})$.

For the regularized estimation at (2.3), we adopt the composite gradient descent algorithm (Loh and Wainwright, 2015), which is designed for a nonconvex problem and fits our objective functions well. Consider the SCAD penalty, which satisfies Assumption 4 with $L = 1$ and $\mu = 1/(\alpha - 1)$. We then can rewrite the optimization problem at (2.3) into

$$\widetilde{\boldsymbol{\beta}}_n = \arg\min_{\boldsymbol{\beta}\in\Theta} \underbrace{\{n^{-1}L_n(\boldsymbol{\beta}) - \mu\|\boldsymbol{\beta}\|_2^2/2\}}_{\widetilde{L}_n(\boldsymbol{\beta})} + \lambda g(\boldsymbol{\beta}),$$

where, from Assumption 4, $g(\boldsymbol{\beta}) = \{\sum_{j=1}^d p_\lambda(\boldsymbol{\beta}_j) + \mu\|\boldsymbol{\beta}\|_2^2/2\}/\lambda$ is convex. As a result, similar to the composite gradient descent algorithm in Loh and Wainwright (2015), the $(r+1)$th update can be calculated by

$$\boldsymbol{\beta}^{(r+1)} = \arg\min\left[\|\boldsymbol{\beta} - \{\boldsymbol{\beta}^{(r)} - \eta\nabla\widetilde{L}_n(\boldsymbol{\beta})\}\|_2^2/2 + \lambda\eta g(\boldsymbol{\beta})\right],$$

which has a closed-form solution of

$$\boldsymbol{\beta}^{(r+1)} = \begin{cases} 0, & 0 \le |z| \le \nu\lambda \\ z - \text{sign}(z)\cdot\nu\lambda, & \nu\lambda \le |z| \le (\nu+1)\lambda \\ \{z - \text{sign}(z)\cdot\dfrac{\alpha\nu\lambda}{\alpha-1}\}/\{1 - \dfrac{\nu}{\alpha-1}\}, & (\nu+1)\lambda \le |z| \le \alpha\lambda \\ z, & |z| \ge \alpha\lambda \end{cases}$$

with $z = \{\boldsymbol{\beta}^{(r)} - \eta\nabla\widetilde{L}_n(\boldsymbol{\beta}^{(r)})\}/(1+\mu\eta)$ and $\nu = \eta/(1+\mu\eta)$, where the step size η is chosen by the BLS method.

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0069

3.2 Hyper-parameter selection

There are two types of hyper-parameters in the penalized estimation at (2.3): the tuning parameter λ and quantile levels of τ_k with $1 \leq k \leq K$. We can employ validation methods to select the tuning parameter λ such that the composite quantile check loss is minimized.

The selection of τ_k's is another important task since it will affect the efficiency of resulting estimators. Suppose that we are interested in some high quantiles of τ_m^* with $1 \leq m \leq M$, and then the QIR model can be assumed to the interval of $\mathcal{I} = [\tau_0, 1)$, which contains all τ_m^*'s. We may further choose a suitable interval of $[\tau_L, \tau_U] \subset \mathcal{I}$ such that τ_k's can be equally spaced on it, i.e. $\tau_k = \tau_L + k(\tau_U - \tau_L)/(K+1)$ for $1 \leq k \leq K$, where it can be set to $\tau_0 = \tau_L$. As a result, the selection of τ_k's is equivalent to that of $[\tau_L, \tau_U]$.

We may choose τ_U such that it is close to τ_m^*'s, while a reliable estimation can be afforded at this level. The selection of τ_L is a trade-off between estimation efficiency and model misspecification; see Wang et al. (2012); Wang and Tsai (2009). On one hand, to improve estimation efficiency, we may choose τ_L close to 0.5 since the richest observations will appear at the middle for most real data. On the other hand, we have to assume the parametric structure over the whole interval of $[\tau_L, 1)$, i.e. more limitations will be added to the real example. The criterion of prediction errors (PEs) is hence introduced,

$$PE = \frac{1}{M}\sum_{m=1}^{M} \frac{1}{\sqrt{\tau_m^*(1-\tau_m^*)}} \cdot \sqrt{n} \left| \frac{1}{n}\sum_{i=1}^{n} I\{y_i < \widehat{Q}_Y(\tau_m^*|\mathbf{X}_i)\} - \tau_m^* \right|,$$

where we will choose τ_L with the minimum value of PEs; see also Wang et al. (2012). Moreover,

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0069

for a given interval $[\tau_L, \tau_U]$, the CQR estimator is not sensitive to the number of levels K once it reaches the minimum number for identification; see Section S5.2 of the Supplementary Material for details. In practice, we may simply choose $K = 5$ or 10.

In practice, the cross validation method can be used to select λ and τ_L simultaneously. Specifically, the composite quantile check loss and PEs are both evaluated at validation sets. For each candidate interval of $[\tau_L, \tau_U]$, the tuning parameter λ is selected according to the composite quantile check loss, and the corresponding value of PE is also recorded. We then will choose the value of τ_L, which corresponds to the minimum value of PEs among all candidate intervals.

4. Simulation Studies

4.1 Composite quantile regression estimation

This subsection conducts simulation experiments to evaluate the finite-sample performance of the low-dimensional composite quantile regression (CQR) estimation at (2.2).

The Tukey Lambda distribution in Example 2 is used to generate the $i.i.d.$ sample,

$$Y_i = Q_Y(U_i, \boldsymbol{\theta}(\mathbf{X}_i, \boldsymbol{\beta})) = \theta_1(\mathbf{X}_i, \boldsymbol{\beta}) + \theta_2(\mathbf{X}_i, \boldsymbol{\beta}) \frac{U_i^{\theta_3(\mathbf{X}_i, \boldsymbol{\beta})} - (1 - U_i)^{\theta_3(\mathbf{X}_i, \boldsymbol{\beta})}}{\theta_3(\mathbf{X}_i, \boldsymbol{\beta})} \tag{4.1}$$
$$\theta_1(\mathbf{X}_i, \boldsymbol{\beta}) = g_1(\mathbf{X}_i' \boldsymbol{\beta}_1),\ \theta_2(\mathbf{X}_i, \boldsymbol{\beta}) = g_2(\mathbf{X}_i' \boldsymbol{\beta}_2),\ \theta_3(\mathbf{X}_i, \boldsymbol{\beta}) = g_3(\mathbf{X}_i' \boldsymbol{\beta}_3),$$

where $\{U_i\}$ are independent and follow Uniform$(0, 1)$, $\mathbf{X}_i = (1, X_{i1}, X_{i2})'$, $\{(X_{i1}, X_{i2})'\}$ is an $i.i.d.$ sequence with bivariate standard normality. The true parameter vector is $\boldsymbol{\beta}_0 = (\boldsymbol{\beta}_{01}', \boldsymbol{\beta}_{02}', \boldsymbol{\beta}_{03}')'$, and we set the location parameters $\boldsymbol{\beta}_{01} = (1, 0.5, -1)'$, the scale parameters $\boldsymbol{\beta}_{02} = (1, 0.5, -1)'$ and the tail parameters $\boldsymbol{\beta}_{03} = (1, -1, 1)'$. For the tail index $\theta_3(\mathbf{X}_i, \boldsymbol{\beta})$,

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0069

before generating the data, we first scale each covariate into the range of $[-0.5, 0.5]$ such that a relatively stable sample can be generated. In addition, g_1, g_2 and g_3 are the inverse of link functions. We choose identity link for the location index and softplus-related link for the scale and tail indices, i.e., $g_1(x) = x, g_2(x) = \text{softplus}(x)$ and $g_3(x) = 1 - \text{softplus}(x)$, where $\text{softplus}(x) = \log\{1 + \exp(x)\}$ is a smoothed version of $x_+ = \max(0, x)$ and hence the name. Note that $g_2(x) > 0$ and $g_3(x) < 1$. We consider three sample sizes of $n = 500$, 1000 and 2000, and there are 500 replications for each sample size.

The algorithm for CQR estimation in Section 3 is applied with $K = 10$ and τ_k's being equally spaced over $[\tau_L, \tau_U]$. We consider three quantile ranges of $(\tau_L, \tau_U) = (0.5, 0.99)$, $(0.7, 0.99)$ and $(0.9, 0.99)$, and the estimation efficiency is first evaluated. Figure 1 gives the boxplots of three fitted location parameters $\widehat{\boldsymbol{\beta}}_{1n} = (\widehat{\beta}_{1,1}, \widehat{\beta}_{1,2}, \widehat{\beta}_{1,3})'$. It can be seen that both bias and standard deviation decrease as the sample size increase. Moreover, when τ_L decreases, the quantile levels with richer observations will be used for the estimation and, as expected, both bias and standard deviation will decrease. Boxplots for fitted scale and tail parameters show a similar pattern and hence are omitted to save the space.

We next evaluate the prediction performance of $Q(\tau^*, \boldsymbol{\theta}(\mathbf{X}, \widehat{\boldsymbol{\beta}}_n))$ at two interesting quantile levels of $\tau^* = 0.991$ and 0.995. Consider two values of covariates, $\boldsymbol{X} = (1, 0.1, -0.2)'$ and $(1, 0, 0)'$, and the corresponding tail indices are $\theta_3(\mathbf{X}, \boldsymbol{\beta}_0) = -0.1032$ and -0.3133, respectively. Note that the Tukey lambda distribution can provide a good approximation to Cauchy and normal distributions when the tail indices are -1 and 0.14, respectively, and it becomes

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0069

more heavy-tailed when the tail index decreases (Freimer et al., 1988). The prediction error in terms of squared loss (PES), $\{Q(\tau^*, \boldsymbol{\theta}(\mathbf{X}, \widehat{\boldsymbol{\beta}}_n)) - Q(\tau^*, \boldsymbol{\theta}(\mathbf{X}, \boldsymbol{\beta}_0))\}^2$, is calculated for each replication, and the corresponding sample mean refers to the commonly used mean square error. Table 1 presents both the sample mean and standard deviation of PESs across 500 replications. A clear trend of improvement can be observed as the sample size becomes larger, and the prediction is more accurate at the 99.1-th level for almost all cases.

We have also conducted three experiments to evaluate the performance with quantile functions being GEVD and GPD, to check the sensitivity of K and link functions, and to compare the proposed method with existing ones in the literature, respectively. The results are relegated to the Supplementary Material, and we briefly state the findings below. First, the proposed CQR performs similarly for the DGPs with three different quantile functions, and it is not sensitive to the selection of K when the model is correctly specified. Second, when link functions are wrongly specified, our methodology will be affected dramatically by the misspecification of tail index, while it is not that sensitive to the misspecification of location index. Finally, our QIR has better performance than existing methods especially when data exists heterogeneity in tail.

4.2 High-dimensional regularized estimation

This subsection conducts simulation experiments to evaluate the finite-sample performance of the high-dimensional regularized estimation at (2.3).

For the DGP at (4.1), we consider three dimensions of $p = 50$, 100 and 150, and the

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0069

true parameter vectors are extended from those in Section 4.1 by adding zeros, i.e. $\boldsymbol{\beta}_{01} = (1, 0.5, -1, 0, ..., 0)'$, $\boldsymbol{\beta}_{02} = (1, 0.5, -1, 0, ..., 0)'$ and $\boldsymbol{\beta}_{03} = (1, -1, 1, 0, ..., 0)'$, which are vectors of length p with 3 non-zero entries. As a result, all true parameters $\boldsymbol{\beta}_0 = (\boldsymbol{\beta}'_{01}, \boldsymbol{\beta}'_{02}, \boldsymbol{\beta}'_{03})'$ make a vector of length $3p$ with $s = 9$ non-zero entries. The sample size is chosen such that $n = \lfloor cs \log p \rfloor$ with $c = 5$, 10, 20, 30, 40 and 50, where $\lfloor x \rfloor$ refers to the largest integer smaller than or equal to x. All other settings are the same as in the previous subsection.

The algorithm for regularized estimation in Section 3 is used to search for the estimators, and we generate an independent validation set of size $5n$ to select tuning parameter λ by minimizing the composite quantile check loss; see also Wang et al. (2012). Figure 2 gives the estimation errors of $\|\widetilde{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0\|$. It can be seen that $\|\widetilde{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0\|$ is roughly proportional to the quantity of $\sqrt{s \log p/n}$, and this confirms the convergence rate in Theorem 4. Moreover, the estimation errors approach zero as the sample size n increases, and we can then conclude the consistency of $\widetilde{\boldsymbol{\beta}}_n$. Finally, when τ_L increases, the quantile levels with less observations will be used in the estimating procedure, and hence larger estimation errors can be observed.

We next evaluate the prediction performance at quantile levels $\tau^* = 0.991$ and 0.995, and covariates $\mathbf{X}$ take values of $(1, 0.1, -0.2, 0, \cdots, 0)'$ and $(1, 0, 0, 0, \cdots, 0)'$, similar to those in the previous subsection. Table 2 gives mean square errors of the predicted conditional quantiles $Q(\tau^*, \boldsymbol{\theta}(\mathbf{X}, \widetilde{\boldsymbol{\beta}}_n))$, as well as the sample standard deviations of prediction errors in squared loss, with $p = 50$. It can be seen that larger sample size leads to much smaller mean square errors. Moreover, when τ_L is larger, the prediction also becomes worse, and it may be due to the

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0069

lower estimation efficiency. Finally, similar to the experiments in the previous subsection, the prediction at $\tau^* = 0.991$ is more accurate for almost all cases. The results for the cases with $p = 100$ and 150 are similar and hence omitted.

Finally, we consider the following criteria to evaluate the performance of variable selection: average number of selected active coefficients (size), percentage of active and inactive coefficients both correctly selected simultaneously (P_{AI}), percentage of active coefficients correctly selected (P_A), percentage of inactive coefficients correctly selected (P_I), false positive rate (FP), and false negative rate (FN). Table 3 reports the selecting results with $p = 50$ and $c = 10$, 30 and 50. When τ_L is larger, both P_{AI} and P_A decrease, and it indicates the increasing of selection accuracy. In addition, performance improves when sample size gets larger. The results for $p = 100$ and 150 are similar and hence omitted.

In addition, we have also conducted experiments to evaluate the performance with quantile functions being GEVD or GPD, receptively, and the similar findings can be observed; see the Supplementary Material for details.

5. Application to Childhood Malnutrition

Childhood malnutrition is well known to be one of the most urgent problems in developing countries. The Demographic and Health Surveys (DHS) has conducted nationally representative surveys on child and maternal health, family planning and child survival, etc., and this results in many datasets for research purposes. The dataset for India was first analyzed by Koenker (2011),

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0069

and can be downloaded from the website at `http://www.econ.uiuc.edu/roger/research/bandaids/bandaids.html`. It has also been studied by many researchers (Fenske et al., 2011; Belloni et al., 2019) for childhood malnutrition problem in India, and quantile regression with low- or high-dimensional covariates was conducted at the levels of $\tau = 0.1$ and 0.05. The proposed model enables us to consider much lower quantiles, corresponding to more severe childhood malnutrition problem.

The child's height is chosen as the indicator for malnutrition as in Belloni et al. (2019). Specifically, the response is set to $Y = -100 \log$(child's height in centimeters), and we then consider high quantiles to study the childhood malnutrition problem such that it is consistent with previous sections. Other variables include seven continuous and 13 categorical ones, and they contain both biological factors and socioeconomic factors that are possibly related to childhood malnutrition. Examples of biological factors include the child's age, gender, duration of breastfeeding in months, the mother's age and body-mass index (BMI), and socioeconomic factors contain the mother's employment status, religion, residence, and the availability of electricity. All seven continuous variables are standardized to have mean zero and variance one, and two-way interactions between all variables are also included. Moreover, we concentrate on the samples from pool families. As a result, there are $p = 328$ covariates in total after removing variables with all elements being zero, and the sample size is $n = 6858$. Denote the full model size by $(328, 328, 328)$, which correspond to the sizes of location, scale and tail, respectively. Furthermore, as in the simulation experiments, covariates are further rescaled to the

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0069

range $[-0.5, 0.5]$ for the tail index.

We aim at two high quantiles of $\tau^* = 0.991$ and 0.995, and the algorithm for high-dimensional regularized estimation in Section 3 is first applied to select the interval of $[\tau_L, \tau_U]$. Specifically, the value of τ_U is fixed to 0.99, and that of τ_L is selected among $\tau_L = 0.9 + 0.01j$ with $1 \leq j \leq 8$. The value of K is set to 10, and the τ_k's with $1 \leq k \leq K$ are equally spaced over $[\tau_L, \tau_U]$. For each τ_L, the whole samples are randomly split into five parts with equal size, except that one part is short of two observations, and the 5-fold cross validation is used to select the tuning parameter λ. To stabilize the process, we conduct the random splitting five times and choose the value of λ minimizing the composite check loss over all five splittings. The averaged value of PEs is also calculated over all five splittings, and the corresponding plot is presented in Figure 3. As a result, we choose $\tau_L = 0.96$ since it corresponds to the minimum value of PEs.

We next apply the QIR model to the whole dataset with $[\tau_L, \tau_U] = [0.96, 0.99]$, and the tuning parameter λ is scaled by $\sqrt{4/5}$ since the sample size changes from $4n/5$ to n. The fitted model is of size $(14, 16, 19)$, and we can predict the conditional quantile structure at any level $\tau^* \in (0.96, 1)$. For example, consider the variable of child's age, and we are interested in children with ages of 20, 30 and 40 months. The duration of breastfeeding is set to be the same as child's age, since the age is always larger than the duration of breastfeeding, and we set the values of all other variables in $\boldsymbol{X}$ to be the same as the 460th observation, which has the response value being the sample median. Figure 4 plots the predicted quantile curves for three different ages. It can be seen that younger children may have extremely lower heights, and we

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0069

may conclude that it may be easier for younger children to be affected by malnutrition.

Figure 4 also draws quantile curves for mother's education, child's gender and mother's unemployment condition, and the values of variables at the 460th observation are also used for non-focal covariates in the prediction. For child's gender, the baby boy is usually higher than baby girls as observed in Koenker (2011), while the difference varnishes for much larger quantiles. In addition, the quantile curves for mother's education are almost parallel, while those for mother's unemployment condition are crossed. More importantly, all these new insights are at very high quantiles, and this confirms the necessity of the proposed model.

Finally, we compare the proposed QIR model with two commonly used ones in the literature and a special case of QIR: (i.) linear quantile regression (LQR) at the level of τ^* with ℓ_1 penalty in Belloni et al. (2019), (ii.) extremal quantile regression (EQR) in (Wang et al., 2012) adapted to high-dimensional data, and (iii.) degenerated QIR (dQIR) with identity link functions for location and scale indices and a constant tail index. The prediction performance at $\tau^* = 0.991$ and 0.995 is considered for the comparison, and we fix $[\tau_L, \tau_U] = [0.96, 0.99]$. For Method (i.), the τ^*th conditional quantile prediction is $\check{Q}_Y(\tau^* \mid \boldsymbol{X}) = \check{\alpha}(\tau^*) + \mathbf{X}'\check{\boldsymbol{\beta}}(\tau^*)$, where $\{\check{\alpha}(\tau^*), \check{\boldsymbol{\beta}}(\tau^*)\} = \arg\min_{\alpha,\boldsymbol{\beta}} n^{-1} \sum_{i=1}^n \rho_{\tau^*}(Y_i - \alpha - \mathbf{X}_i'\boldsymbol{\beta}) + \lambda \sum_{j=1}^d |\beta_j|$ is the Lasso-penalized estimator. For Method (ii.), we first estimate the intermediate conditional quantiles using Lasso-penalized LQR, and then extrapolates these estimates to the high tails based on the estimated tail index. Specifically, we consider $\widetilde{K} = \lfloor 4.5 n_{\text{train}}^{1/3} \rfloor = 38$ quantile levels, equally spaced over $[0.96, 0.99]$, and the LQR with ℓ_1 penalty is conducted at each level. As in Wang

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0069

et al. (2012), we can estimate the extreme value index based on these estimated intermediate conditional quantiles, and hence the fitted structures can be extrapolated to the level of τ^*; see S6 of the Supplementary Material for further details. Note that there is no theoretical justification for Method (ii.) in the literature. Moreover, the dQIR has a comparable structure to that of EQR, while they differ in estimation. As in simulation experiments, the tuning parameter λ in the above four methods is selected by minimizing the composite check loss in the testing set. We randomly split the data 100 times, and one value of PE can be obtained from each splitting. Figure 3 gives the boxplots of PEs from our model, the degenerated QIR and two competing methods, and the advantages of our model can be observed at both target levels of $\tau^* = 0.991$ and 0.995.

6. Conclusions and discussions

This paper proposes a reliable method for the inference at extreme quantiles with both low- and high-dimensional covariates. The main idea is first to conduct a composite quantile regression at fixed quantile levels, and we then can extrapolate the estimated results to extreme quantiles by assuming a parametric structure at tails. The Tukey lambda structure can be used due to its flexibility and the explicit form of its quantile functions, and the success of the proposed methodology has been demonstrated by extensive numeral studies.

This paper can be extended in the following two directions. On one hand, in the proposed model, a parametric structure is assumed over the interval of $[\tau_0, 1)$. Although the criterion of

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0069

PE is suggested in Section 3 to balance the estimation efficiency and model misspecification, it should be interesting to provide a statistical tool for the goodness-of-fit. Dong et al. (2019) introduced a goodness-of-fit test for parametric quantile regression at a fixed quantile level, and it can be used for our problem by extending the test statistic from a fixed level to the interval of $[\tau_0, 1)$. We leave it for the future research. On the other hand, the idea in this paper is general and can be applied to many other scenarios. For example, for conditional heteroscedastic time series models, it is usually difficult to conduct the quantile estimation at both median and extreme quantiles. The difficulty at extreme quantiles is due to the sparse data at tails, while that at median is due to the tiny values of fitted parameters (Zhu et al., 2018; Zhu and Li, 2022). Our idea certainly can be used to solve this problem to some extent.

Acknowledgments

We are deeply grateful to the Co-Editor, the Associate Editor and two anonymous referees for their valuable comments that led to the substantial improvement in the quality of this paper. Zhang's research was supported by the National Natural Science Foundation of China Grants 12471280 and 12101241. Zhu's research was supported by the National Natural Science Foundation of China Grants 72373087 and 12271330. Li's research was supported by the Hong Kong Research Grant Council Grants 17313722 and 17309625, and the National Natural Science Foundation of China Grant 72033002.

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0069

Supplementary Material

The online supplementary materials include the proofs of Proposition 1 and all theorems, and additional numerical results for simulation and real data analysis.

References

Abrevaya, J. (2001). The effect of demographics and maternal behavior on the distribution of birth outcomes. *Empirical Economics 26*, 247–259.

Beirlant, J. and Y. Goegebeur (2004). Local polynomial maximum likelihood estimation for Pareto-type distributions. *Journal of Multivariate Analysis 89*, 97–118.

Belloni, A. and V. Chernozhukov (2011). l_1-penalized quantile regression in high-dimensional sparse models. *The Annals of Statistics 39*, 82–130.

Belloni, A., V. Chernozhukov, D. Chetverikov, and I. Fernandez-Val (2019). Conditional quantile processes based on series or many regressors. *Journal of Econometrics 213*, 4–29.

Belloni, A., V. Chernozhukov, and K. Kato (2019). Valid post-selection inference in high-dimensional approximately sparse quantile regression models. *Journal of the American Statistical Association 114*, 749–758.

Bertsekas, D. P. (2016). *Nonlinear Programming*. Athena Scientific.

Cade, B. S. and B. R. Noon (2003). A gentle introduction to quantile regression for ecologists. *Frontiers in Ecology and the Environment 1*, 412–420.

Chernozhukov, V. (2005). Extremal quantile regression. *The Annals of Statistics 33*, 806–839.

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0069

REFERENCES

Chernozhukov, V., I. Fernández-Val, and A. E. Kowalski (2015). Quantile regression with censoring and endogeneity. *Journal of Econometrics 186*, 201–221.

Chernozhukov, V., I. Fernández-Val, and B. Melly (2022). Fast algorithms for the quantile regression process. *Empirical economics 62*, 1–27.

Davino, C., M. Furno, and D. Vistocco (2014). *Quantile Regression: Theory and Applications*. New York: Wiley.

de Haan, L. and A. Ferreira (2006). *Extreme Value Theory: An Introduction*. Springer.

Dong, C., G. Li, and X. Feng (2019). Lack-of-fit tests for quantile regression models. *Journal of the Royal Statistical Society, Series B 81*, 629–648.

Elsner, J. B., J. P. Kossin, and T. H. Jagger (2008). The increasing intensity of the strongest tropical cyclones. *Nature 455*, 92–95.

Fan, J. and R. Li (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association 96*, 1348–1360.

Fenske, N., T. Kneib, and T. Hothorn (2011). Identifying risk factors for severe childhood malnutrition by boosting additive quantile regression. *Journal of the American Statistical Association 106*, 494–510.

Fisher, R. A. and L. H. C. Tippett (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Mathematical Proceedings of the Cambridge Philosophical Society 24*, 180–190.

Fournier, B., N. Rupin, M. Bigerelle, D. Najjar, A. Iost, and R. Wilcox (2007). Estimating the parameters of a generalized lambda distribution. *Computational Statistics & Data Analysis 51*, 2813–2835.

Freimer, M., G. Kollia, G. S. Mudholkar, and C. T. Lin (1988). A study of the generalized tukey lambda family. *Communications*

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0069

in Statistics-Theory and Methods 17, 3547–3567.

Galvao, A. F. and K. Kato (2016). Smoothed quantile regression for panel data. *Journal of Econometrics 193*, 92–112.

Gilchrist, W. G. (2000). *Statistical modelling with quantile functions*. London: Chapman & Hall/CRC.

Hendricks, W. and R. Koenker (1991). Hierarchical spline models for conditional quantiles and the demand for electricity. *Journal of the American Statistical Association 87*, 58–68.

Jagger, T. H. and J. B. Elsner (2008). Modeling tropical cyclone intensity with quantile regression. *International Journal of Climatology 29*, 1351–1361.

Jiang, X., J. Jiang, and X. Song (2012). Oracle model selection for nonlinear models based on weighted composite quantile regression. *Statistica Sinica 22*, 1479–1506.

Kai, B., R. Li, and H. Zou (2010). Local composite quantile regression smoothing: An efficient and safe alternative to local polynomial regression. *Journal of the Royal Statistical Society, Series B 72*, 49–69.

Kato, K., A. F. Galvao, and G. V. Montes-Rojas (2012). Asymptotics for panel quantile regression models with individual effects. *Journal of Econometrics 170*, 76–91.

Koenker, R. (2005). *Quantile regression*. Cambridge: Cambridge University Press.

Koenker, R. (2011). Additive models for quantile regression: Model selection and confidence bandaids. *Brazilian Journal of Probability and Statistics 25*, 239–262.

Koenker, R. (2017). Quantile regression: 40 years on. *Annual Review of Economics 9*, 155–176.

Koenker, R. and G. Bassett (1978). Regression quantiles. *Econometrica 46*, 33–50.

Kuester, K., S. Mittnik, and M. S. Paolella (2006). Value-at-Risk prediction: A comparison of alternative strategies. *Journal of*

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0069

Financial Econometrics 4, 53–89.

Linton, O. and Z. Xiao (2017). Quantile regression applications in finance. In *Handbook of Quantile Regression*, pp. 381–407. Chapman and Hall/CRC.

Lobeto, H., M. Menendez, and I. J. Losada (2021). Future behavior of wind wave extremes due to climate change. *Scientific reports 11*, 1–12.

Loh, P. (2017). Statistical consistency and asymptotic normality for high-dimensional robust M-estimators. *The Annals of Statistics 45*, 866–896.

Loh, P. and M. J. Wainwright (2015). Regularized M-estimators with nonconvexity: Statistical and algorithmic theory for local optima. *Journal of Machine Learning Research 16*, 559–616.

Meinshausen, N. and G. Ridgeway (2006). Quantile regression forests. *Journal of Machine Learning Research 7*, 983–999.

Moon, S. J., J.-J. Jeon, J. S. H. Lee, and Y. Kim (2021). Learning multiple quantiles with neural networks. *Journal of Computational and Graphical Statistics 30*, 1238–1248.

Newey, K. and D. McFadden (1994). Large sample estimation and hypothesis testing. *Handbook of Econometrics 4*, 2112–2245.

Noufaily, A. and M. C. Jones (2013). Parametric quantile regression based on the generalized gamma distribution. *Journal of the Royal Statistical Society, Series C 62*, 723–740.

Pan, X. and W.-X. Zhou (2021). Multiplier bootstrap for quantile regression: Non-asymptotic theory under random design. *Information and Inference 3*, 813–861.

Pickands, J. (1975). Statistical Inference Using Extreme Order Statistics. *The Annals of Statistics 3*, 119–131.

Pollard, D. (1985). New ways to prove central limit theorems. *Econometric Theory 1*, 295–313.

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0069

REFERENCES

Racine, J. S. and K. Li (2017). Nonparametric conditional quantile estimation: a locally weighted quantile kernel approach. *Journal of Econometrics 201*, 72–94.

Sherwood, B., L. Wang, et al. (2016). Partially linear additive quantile regression in ultra-high dimension. *The Annals of Statistics 44*, 288–317.

Sivakumar, V. and A. Banerjee (2017). High-dimensional structured quantile regression. In *International Conference on Machine Learning*, pp. 3220–3229.

van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge: Cambridge University Press.

Vasicek, O. (1976). A test for normality based on sample entropy. *Journal of the Royal Statistical Society, Series B 38*, 54–59.

Wang, H. and C.-L. Tsai (2009). Tail index regression. *Journal of the American Statistical Association 104*, 1233–1240.

Wang, H. J. and D. Li (2013). Estimation of extreme conditional quantiles through power transformation. *Journal of the American Statistical Association 108*, 1062–1074.

Wang, H. J., D. Li, and X. He (2012). Estimation of high conditional quantiles for heavy-tailed distributions. *Journal of the American Statistical Association 107*, 1453–1464.

Wang, L. and X. He (2022). Analysis of global and local optima of regularized quantile regression in high dimensions: A subgradient approach. *Econometric Theory 40*, 233–277.

Wang, L., Y. Wu, and R. Li (2012). Quantile regression for analyzing heterogeneity in ultra-high dimension. *Journal of the American Statistical Association 107*, 214–222.

Yu, K., Z. Lu, and J. Stander (2003). Quantile regression: applications and current research areas. *Journal of the Royal Statistical Society: Series D 52*, 331–350.

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0069

REFERENCES

Zhang, C.-H. (2010). Nearly unbiased variable selection under minimax concave penalty. *The Annals of Statistics 38*, 894–942.

Zhang, Y., H. Lian, and Y. Yu (2020). Ultra-high dimensional single-index quantile regression. *Journal of Machine Learning Research 21*, 1–25.

Zheng, Q., L. Peng, and X. He (2015). Globally adaptive quantile regression with ultra-high dimensional data. *The Annals of Statistics 43*, 2225–2258.

Zheng, Y., Q. Zhu, G. Li, and Z. Xiao (2018). Hybrid quantile regression estimation for time series models with conditional heteroscedasticity. *Journal of the Royal Statistical Society, Series B 80*, 975–993.

Zhu, K. and S. Ling (2011). Global self-weighted and local quasi-maximum exponential likelihood estimators for arma–garch/igarch models. *The Annals of Statistics 39*, 2131–2163.

Zhu, Q. and G. Li (2022). Quantile double autoregression. *Econometric Theory 38*, 793–839.

Zhu, Q., Y. Zheng, and G. Li (2018). Linear double autoregression. *Journal of Econometrics 207*, 162–174.

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

Yingying Zhang, East China Normal University, KLATASDS - MOE, School of Statistics, Academy of Statistics and Interdisciplinary Sciences

E-mail: yyzhang@fem.ecnu.edu.cn

Qianqian Zhu, Shanghai University of Finance and Economics, School of Statistics and Data Science, Institute of Big Data Research

E-mail: zhu.qianqian@mail.shufe.edu.cn

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0069

Table 1: Mean square errors of the predicted conditonal quantile $Q(\tau^*, \boldsymbol{\theta}(\mathbf{X}, \widehat{\boldsymbol{\beta}}_n))$ at the level of $\tau^* = 0.991$ or 0.995. The values in bracket refer to the corresponding sample standard deviations of prediction errors in squared loss.

			$\boldsymbol{X} = (1, 0.1, -0.2)^T$		$\boldsymbol{X} = (1, 0, 0)^T$	
n	$[\tau_L, \tau_U]$		0.991	0.995	0.991	0.995
		True	10.34	11.83	15.13	18.84
500	$[0.5, 0.99]$		1.32(2.17)	2.35(4.11)	5.41(11.41)	12.13(28.19)
	$[0.7, 0.99]$		1.42(2.20)	2.55(4.07)	5.42(10.95)	12.12(26.73)
	$[0.9, 0.99]$		2.00(3.92)	3.64(7.56)	6.18(12.86)	14.10(32.78)
1000	$[0.5, 0.99]$		0.77(1.68)	1.39(3.29)	2.67(5.28)	5.93(12.61)
	$[0.7, 0.99]$		0.80(1.39)	1.44(2.64)	2.62(4.27)	5.75(9.75)
	$[0.9, 0.99]$		1.31(2.53)	2.44(5.07)	3.22(5.08)	7.23(11.78)
2000	$[0.5, 0.99]$		0.32(0.47)	0.57(0.85)	1.03(1.56)	2.25(3.49)
	$[0.7, 0.99]$		0.36(0.49)	0.64(0.90)	1.05(1.47)	2.31(3.24)
	$[0.9, 0.99]$		0.70(1.34)	1.30(2.44)	1.34(1.75)	3.05(4.06)

Yuefeng Si, The University of Hong Kong, Department of Statistics and Actuarial Science

E-mail: u3006932@connect.hku.hk

Guodong Li, The University of Hong Kong, Department of Statistics and Actuarial Science

E-mail: gdli@hku.hk

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0069

Table 2: Mean square errors of the predicted conditonal quantile $Q(\tau^*, \boldsymbol{\theta}(\mathbf{X}, \widetilde{\boldsymbol{\beta}}_n))$ at the level of $\tau^* = 0.991$ or 0.995 with $p = 50$ and $n = \lfloor ck \log p \rfloor$. The values in bracket refer to the corresponding sample standard deviations of prediction errors in squared loss.

			$\boldsymbol{X} = (1, 0.1, -0.2, 0, \cdots, 0)^T$		$\boldsymbol{X} = (1, 0, 0, 0, \cdots, 0)^T$	
c	$[\tau_L, \tau_U]$		0.991	0.995	0.991	0.995
		True	10.34	11.83	15.13	18.84
10	$[0.5, 0.99]$		1.82(2.61)	3.23(4.82)	6.75(9.47)	14.83(21.98)
	$[0.7, 0.99]$		2.05(6.00)	3.78(13.10)	6.11(8.22)	13.32(18.56)
	$[0.9, 0.99]$		2.92(7.19)	5.44(15.60)	7.24(10.36)	15.91(25.05)
30	$[0.5, 0.99]$		0.65(1.59)	1.15(3.12)	1.97(3.08)	4.26(6.90)
	$[0.7, 0.99]$		0.65(1.49)	1.18(2.94)	1.96(2.85)	4.26(6.31)
	$[0.9, 0.99]$		0.92(2.15)	1.66(4.08)	2.22(2.95)	4.86(6.40)
50	$[0.5, 0.99]$		0.33(0.49)	0.58(0.84)	1.17(1.77)	2.52(3.88)
	$[0.7, 0.99]$		0.39(0.57)	0.69(1.01)	1.28(1.93)	2.78(4.26)
	$[0.9, 0.99]$		0.54(0.86)	0.99(1.58)	1.55(2.39)	3.51(5.57)

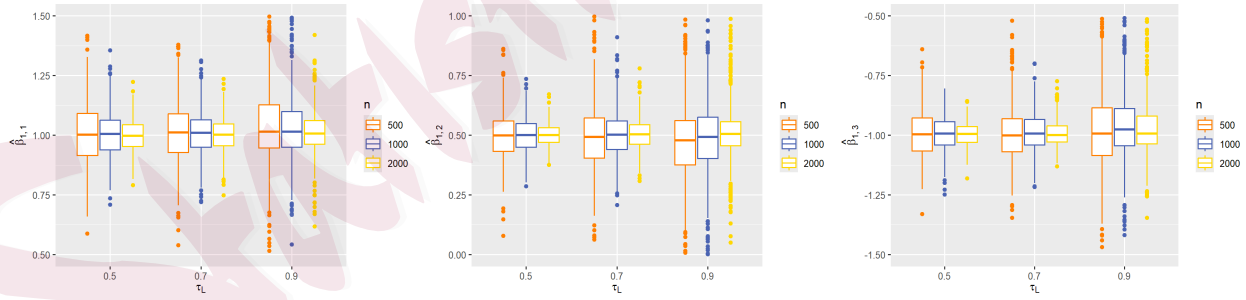

Figure 1: Boxplots for fitted location parameters of $\widehat{\boldsymbol{\beta}}_{1,1}$ (left panel), $\widehat{\boldsymbol{\beta}}_{1,2}$ (middle panel), and $\widehat{\boldsymbol{\beta}}_{1,3}$ (right panel). Sample size is $n = 500$, 1000 or 2000, and the lower bound of quantile range $[\tau_L, \tau_U]$ is $\tau_L = 0.5$, 0.7 or 0.9.

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0069

Table 3: Selection results for regularized estimation with $p = 50$ and $n = \lfloor ck \log p \rfloor$. The values in brackets are the corresponding standard deviations.

$[\tau_L, \tau_U]$	c	size	P_{AI}	P_A	P_I	FP	FN
$[0.5, 0.99]$	10	9.04(0.99)	91.6	96	95.6	0.06(0.68)	0.47(2.34)
	30	9.00(0.00)	100	100	100	0.00(0.00)	0.00(0.00)
	50	9.00(0.00)	100	100	100	0.00(0.00)	0.00(0.00)
$[0.7, 0.99]$	10	8.91(1.25)	79	82.6	95.4	0.07(0.84)	2.02(4.52)
	30	9.00(0.08)	99.4	99.6	99.8	0.00(0.03)	0.04(0.70)
	50	9.00(0.00)	100	100	100	0.00(0.00)	0.00(0.00)
$[0.9, 0.99]$	10	8.56(0.99)	48.4	54.6	90.4	0.10(0.41)	6.51(8.43)
	30	8.88(0.38)	87.8	88.4	99.2	0.01(0.06)	1.42(4.10)
	50	8.96(0.23)	96.4	96.6	99.8	0.00(0.03)	0.44(2.50)

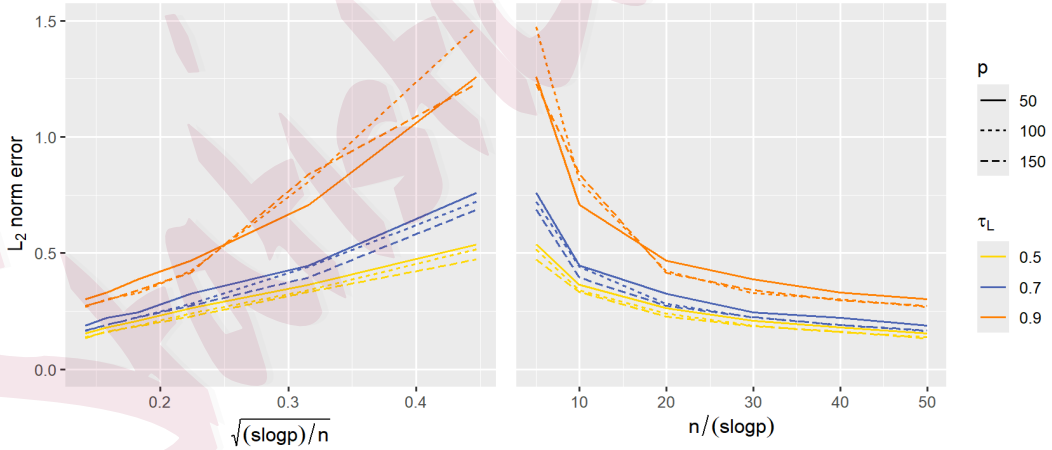

Figure 2: Estimation errors of $\|\widetilde{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0\|$ against the quantities of $\sqrt{(s \log p)/n}$ (left panel) and $n/(s \log p)$ (right panel), respectively.

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0069

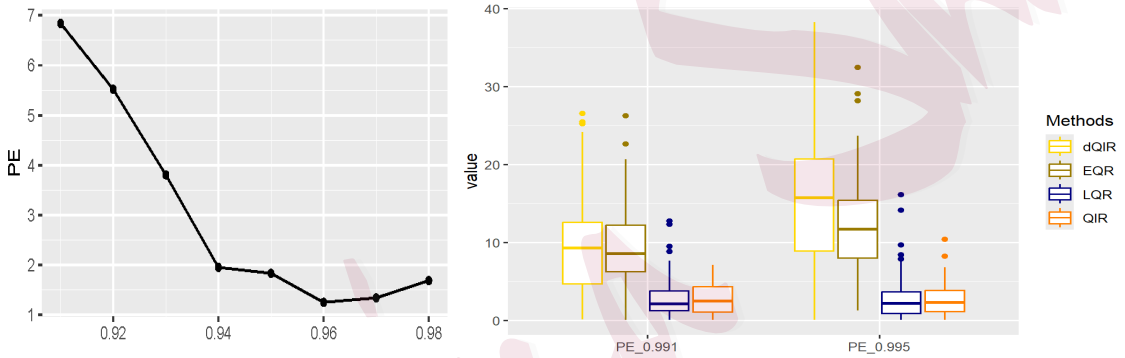

Figure 3: Plot of PEs against τ_L (left panel) and boxplots of PEs from the degenerated QIR (dQIR), extreme quantile regression (EQR), linear quantile regression (LQR) and QIR models at two target levels of $\tau^* = 0.991$ and 0.995 (right panel).

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0069

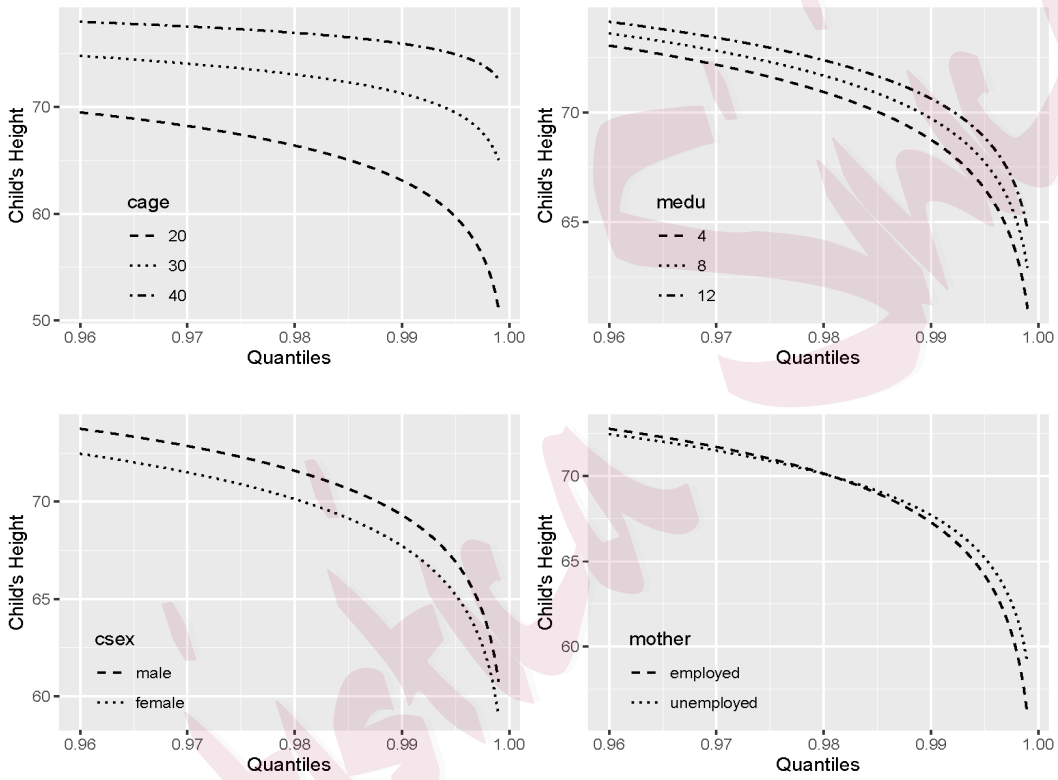

Figure 4: Quantile curves for child's age in months (top left) and mother's education in years (top right) on the three target quantiles. Effects of child's sex (bottom left) and mother's unemployment condition (bottom right).