Biometrika (2021), **103**, 1, *pp*. 1–22
Printed in Great Britain Advance Access publication on 31 July 2018

Quantile index regression

BY YINGYING ZHANG

Academy of Statistics and Interdisciplinary Sciences, East China Normal University, Shanghai, China

yyzhang@fem.ecnu.edu.cn

YUEFENG SI, GUODONG LI

Department of Statistics and Actuarial Science, The University of Hong Kong, Pokfulam Road, Hong Kong, China

u3006932@connect.hku.hk, gdli@hku.hk

AND CHIL-LING TSAI

Graduate School of Management, University of California at Davis, California, U.S.A.

cltsai@ucdavis.edu

SUMMARY

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.

Some key words: Asymptotic normality; High-dimensional analysis; Non-asymptotic property; Partially parametric model; Quantile regression.

1. INTRODUCTION

Quantile regression proposed by Koenker & Bassett (1978) has been widely used across various fields such as biological science, ecology, economics, finance, and machine learning, etc.; see, e.g., Cade & Noon (2003), Yu et al. (2003), Meinshausen & Ridgeway (2006), Linton & 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 & Chernozhukov (2011), Wang et al. (2012b) 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 & Elsner, 2008; Lobeto et al., 2021), and low infant birth weights in medicine (Abrevaya, 2001; Chernozhukov et al., 2020). 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 & Goegebeur (2004) and Wang & Tsai (2009). While this method cannot provide conditional quantiles in explicit forms. Later, Noufaily & Jones (2013) considered a full parametric form, the generalized gamma distribution, to the CDF and then inverted the fitted distribution into a conditional quantile distribution. However, as indicated in Racine & 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. (2012a) and Wang & 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 around these levels. 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 distribution or 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 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:

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

where $F_Y(\cdot|\boldsymbol{X})$ is the distribution of Y conditional on $\mathbf{X}$, and $\boldsymbol{\theta}(\mathbf{X})$ is a d-dimensional parametric function. Note that $\boldsymbol{\theta}(\mathbf{X})$ can be referred to d indices, and model (1.1) can then be called the quantile index regression (QIR) for simplicity. 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 composite 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 not only estimate quantile regression functions effectively, but also forecast extreme quantiles directly. Finally, the QIR model naturally yields non-crossing quantile regression estimators 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 has been carefully studied for the flexible Tukey lambda distribution in Section 2.2.

(b) Section 2.2 also 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 2.3 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 Appendix.

For the sake of convenience, this paper denotes vectors and matrices by boldface letters, 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 $\boldsymbol{X} = (X_1, ..., X_p)'$. We then rewrite the quantile function of Y conditional on $\boldsymbol{X}$ at (1.1) with an explicit form of $\boldsymbol{\theta}(\boldsymbol{X}, \boldsymbol{\beta})$ below,

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

where $\mathcal{I} \subset [0, 1]$ is an interval or the union of multiple disjoint intervals, $\boldsymbol{\theta}(\boldsymbol{X}, \boldsymbol{\beta}) = (\theta_1(\boldsymbol{X}, \boldsymbol{\beta}), \cdots, \theta_d(\boldsymbol{X}, \boldsymbol{\beta}))'$, $\boldsymbol{\beta} = (\boldsymbol{\beta}_1', ..., \boldsymbol{\beta}_d')'$, $\boldsymbol{\beta}_j = (\beta_{j1}, \cdots, \beta_{jp})'$, $\theta_j(\boldsymbol{X}, \boldsymbol{\beta}) = g_j(\boldsymbol{X}'\boldsymbol{\beta}_j)$, the link functions $g_j^{-1}(\cdot)$s are monotonic for $1 \leq j \leq d$, 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 the 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 $\tau_0 \in (0, 1)$ 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. Then, after estimation, we can make a prediction at any level of $\tau \in (\tau_0, 1)$.

Example 2. Consider the Tukey lambda distribution (Vasicek, 1976) that is defined by its quantile function,

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

where $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3)'$, and $\theta_1 \in \mathbb{R}$, $\theta_2 > 0$ and $\theta_3 \leq 1$ are the location, scale and tail indices, respectively. Due to its flexibility, the Tukey lambda distribution can provide an accurate approximation to many commonly used distributions, such as normal, logistic, Weibull, uniform, and Cauchy distributions, etc.; see Gilchrist (2000). It is then expected to have a better performance when model (2.1) is combined with this distribution.

In the literature, there exist many quantile functions, which have explicit forms, such as the generalized lambda and Burr XII distributions (Gilchrist, 2000; Fournier et al., 2007). For example, the generalized lambda distribution has the form of

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

where the indices θ_3 and θ_4 control the right and left tails, respectively. Note that it reduces to the Tukey lambda distribution when $\theta_3 = \theta_4$. This indicates that the generalized lambda distribution can be considered if we focus on the quantiles with the full range, i.e. $\mathcal{I} = [0, 1]$. On the other hand, the Tukey lambda may be a better choice if our interest is on the quantiles at one side only, i.e. $\mathcal{I} \subset [0, 0.5)$ or $\mathcal{I} \subset (0.5, 1]$.

2.2. *Low-dimensional composite quantile regression estimation*

Denote the observed data by $\{(Y_i, \boldsymbol{X}_i')', i = 1, ..., n\}$, and they are independent and identically distributed $(i.i.d.)$ samples of random vector $(Y, \boldsymbol{X})$, where Y_i is the response, $\boldsymbol{X}_i = (X_{i1}, ..., X_{ip})'$ 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, 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 & Yuan (2008) and Kai et al. (2010). Note that $Q(\tau, \boldsymbol{\theta}(\mathbf{X}, \widehat{\boldsymbol{\beta}}_n))$ has the non-crossing property with respect to τ since, for each $\boldsymbol{\theta}$, $Q(\tau, \boldsymbol{\theta})$ is a non-decreasing quantile function.

To study the asymptotic properties of $\widehat{\boldsymbol{\beta}}_n$, we first consider $\boldsymbol{\beta}_0 = (\boldsymbol{\beta}_{01}', ..., \boldsymbol{\beta}_{0d}')' = \arg\min_{\boldsymbol{\beta}} \bar{L}(\boldsymbol{\beta})$, where $\bar{L}(\boldsymbol{\beta}) = E[\sum_{k=1}^{K} \rho_{\tau_k}\{Y - Q(\tau_k, \boldsymbol{\theta}(\mathbf{X}, \boldsymbol{\beta}))\}]$ is the population loss function, and it may not be unique, i.e. the CQR estimation at (2.2) may suffer from the identification problem. For the sake of illustration, let us consider the case without including covariates $\mathbf{X}$, i.e. we estimate $\boldsymbol{\theta}$ with a sequence of quantile levels $\tau_k \in \mathcal{I}$ and $1 \le k \le K$. To this end, it requires that two different values of $\boldsymbol{\theta}$ can not 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. In sum, 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 $1 \le k \le K$ 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 increase the number of quantile levels K to make Assumption 1 hold. However, it may also depend on the structures of quantile functions and the number of indices. For the sake of illustration, we state the sufficient and necessary condition for the Tukey lambda distribution that satisfies Assumption 1.

LEMMA 1. *For the Tukey lambda distribution defined in Example 2, we have that (i) for $\tau_k \in (0, 1)$ with $1 \le k \le K$, Assumption 1 holds if $K \ge 4$; (ii) for $\tau_k \in (0.5, 1)$ (or $\tau_k \in (0, 0.5)$) with $1 \le k \le K$, Assumption 1 holds if and only if $K \ge 3$.*

Assumption 1, together with an additional assumption on covariates $\mathbf{X}$, allows us to show that $\boldsymbol{\beta}_0$ is the unique minimizer of $\bar{L}(\boldsymbol{\beta})$; see the following theorem, which is critical to establish the asymptotic properties of $\widehat{\boldsymbol{\beta}}_n$.

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

To demonstrate the consistency of $\widehat{\boldsymbol{\beta}}_n$ given below, we assume that the parameter space $\Theta \subset \mathbb{R}^{dp}$ is compact, and the true parameter vector $\boldsymbol{\beta}_0$ is an interior point of Θ.

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 $n \to \infty$.*

The moment condition assumed in the above theorem allows us to adopt the uniform consistency theorem of Andrews (1987) in our technical proofs. To show the asymptotic distribution of $\widehat{\boldsymbol{\beta}}_n$, we introduce two additional assumptions given below.

Assumption 2. For all $1 \leq k \leq K$,

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

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

Assumption 2 is used to prove Lemma S2 in the Appendix; see also Zhu & Ling (2011). Assumption 3 is commonly used in the literature of quantile regression (Koenker, 2005; Belloni et al., 2019a), and it can be relaxed by providing more complicated and lengthy technical details (Kato et al., 2012; Chernozhukov et al., 2015; Galvao & Kato, 2016).

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

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

THEOREM 3. *Suppose that Assumptions 2 and 3 hold, and Ω_1 is positive definite. If the conditions of Theorem 2 are satisfied, 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 $n \to \infty$.*

Note that 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 making use of the bracketing method in Pollard (1985). Moreover, to estimate the asymptotic variance in Theorem 3, we first apply the nonparametric method in Hendricks & Koenker (1991) to estimate the quantities of $f_Y\left\{Q(\tau_k, \boldsymbol{\theta}(\mathbf{X}, \boldsymbol{\beta}_0))|\mathbf{X}\right\}$ with $1 \leq k \leq K$, and then matrices Ω_0 and Ω_1 can be approximated by the sample averaging.

In addition, based on the estimator $\widehat{\boldsymbol{\beta}}_n$, one can use $Q(\tau^*, \boldsymbol{\theta}(\mathbf{X}, \widehat{\boldsymbol{\beta}}_n))$ to further predict the conditional quantile at any level $\tau^* \in \mathcal{I}$, and 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* $\boldsymbol{\delta} = E[\partial Q(\tau^*, \boldsymbol{\theta}(\mathbf{X}, \boldsymbol{\beta}_0))/\partial\boldsymbol{\beta}] \in \mathbb{R}^{dp}$.

2.3. *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 nonconvex 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 & 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 & 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. (2012b) and Sherwood et al. (2016) consid-

ered 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 $Q(\tau, \boldsymbol{\theta})$ 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 & Chernozhukov (2011); Sivakumar & Banerjee (2017); Pan & 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 & 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 Appendix follow the framework in Loh & 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 & He, 2021).

Denote by $\mathcal{B}_R(\boldsymbol{\beta}_0) = \{\boldsymbol{\beta} \in \mathbb{R}^{dp} : \|\boldsymbol{\beta} - \boldsymbol{\beta}_0\| \le R\}$ 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 $\lambda_{\min}^0 = \inf_{\boldsymbol{\beta} \in \mathcal{B}_R(\boldsymbol{\beta}_0)} \lambda_{\min}(\boldsymbol{\beta}) > 0$, and assume that $f_{\min} = \min_{1 \le k \le K} \inf_{\boldsymbol{\beta} \in \mathcal{B}_R(\boldsymbol{\beta}_0)} f_Y\{Q(\tau_k, \boldsymbol{\theta}(\mathbf{X}, \boldsymbol{\beta}))|\mathbf{X}\} > 0$.

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 $\bar{\mathcal{E}}(\Delta) \geq 0.5 f_{\min}\lambda^0_{\min}\|\Delta\|^2$ for all Δ such that $\|\Delta\| \leq R$; see Lemma S5 in the Appendix 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,* $2f_{\min}\lambda^0_{\min} > \mu$, $n \gtrsim \log p$ *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*

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

with probability at least $1 - c_1 p^{-c_2} - K\max\{\log p, \log n\}p^{-c^2}$ *for any* $c > 1$, *where* $\alpha = 0.5 f_{\min}\lambda^0_{\min}$, 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 Appendix.*

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 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 $\widehat{\boldsymbol{\beta}}^{(r)}_n$ 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 backtrack-

ing 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 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 $(a, b) = (0.3, 0.5)$.

For the regularized estimation at (2.3), we adopt the composite gradient descent algorithm (Loh & 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 & 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.

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 $1 \le k \le K$. 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 $1 \le m \le M$, 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 $1 \le k \le K$, 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. (2012a); Wang & 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. (2012a).

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,

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

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 2-dimensional 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 $\theta_3(\mathbf{X}_i, \boldsymbol{\beta})$, 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 $\theta_3(\mathbf{X}, \boldsymbol{\beta}_0) = -0.1139$ and -0.3132, 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 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 as in Wang et al. (2012a). 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 quantile level for almost all cases.

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 data generating process at (4.1), we consider three dimensions of $p = 50$, 100 and 150, and the 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. (2012b). 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 increase, 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 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.

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), and can be downloaded from `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., 2019b) 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. (2019b). Specifically, the response is set to $Y = -100\log(\text{child's height in centimeters})$, 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 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 $1 \le j \le 8$. The value of K is set to 10, and the τ_k's with $1 \le k \le K$ 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 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: (i.) linear quantile regression at the level of τ^* with ℓ_1 penalty in Belloni et al. (2019b), and (ii.) extremal quantile regression in (Wang et al., 2012a) adapted to high-dimensional data. 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 (ii.), we consider $K = 4.5n^{1/3}$ quantile levels, equally spaced over $[0.96, 0.99]$, and the linear quantile regression with ℓ_1 penalty is conducted at each level. As in Wang et al. (2012a), we can estimate the extreme value index, and hence the fitted structures can be extrapolated to the level of τ^*. Note that there is no theoretical justification for Method (ii.) in the literature. As in simulation experiments, the tuning parameter λ in the above three 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 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 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 & Li, 2021). Our idea certainly can be used to solve this problem to some extent.

APPENDIX

This appendix gives the proof of Lemma 1 and other technical proofs are given in a supplementary file.

Proof of Lemma 1. The Tukey lambda distribution has the form of

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

where $\theta_1 \in \mathbb{R}$, $\theta_2 > 0$, $\theta_3 \neq 0$. We prove Lemma 1 for $\theta_3 < 1$. Consider four arbitrary quantile levels $0 < \tau_1 < \tau_2 < \tau_3 < \tau_4 < 1$, and two arbitrary index vectors $\widetilde{\boldsymbol{\theta}} = (\widetilde{\theta}_1, ..., \widetilde{\theta}_4)'$ and $\boldsymbol{\theta} = (\theta_1, .., \theta_4)'$ such that

$$Q(\tau_j, \widetilde{\boldsymbol{\theta}}) = Q(\tau_j, \boldsymbol{\theta}) \text{ for all } 1 \leq j \leq 4. \tag{A1}$$

We show that $\widetilde{\boldsymbol{\theta}} = \boldsymbol{\theta}$ in the following.

The first step is to prove $\widetilde{\theta}_3 = \theta_3$ using the proof by contradiction. Suppose that $\widetilde{\theta}_3 \neq \theta_3$ and, without loss of generality, we assume that $\theta_3 < \widetilde{\theta}_3 < 0$. Denote $f_j(\theta_3) = \tau_j^{\theta_3} - (1-\tau_j)^{\theta_3}$ for $1 \leq j \leq 4$. From (A1), we have

$$\frac{f_1(\widetilde{\theta}_3) - f_2(\widetilde{\theta}_3)}{f_3(\widetilde{\theta}_3) - f_2(\widetilde{\theta}_3)} - \frac{f_1(\theta_3) - f_2(\theta_3)}{f_3(\theta_3) - f_2(\theta_3)} = 0, \tag{A2}$$

and

$$\frac{f_4(\widetilde{\theta}_3) - f_2(\widetilde{\theta}_3)}{f_3(\widetilde{\theta}_3) - f_2(\widetilde{\theta}_3)} - \frac{f_4(\theta_3) - f_2(\theta_3)}{f_3(\theta_3) - f_2(\theta_3)} = 0. \tag{A3}$$

Let us fix θ_3, $\widetilde{\theta}_3$, τ_2 and τ_3. As a result, $\kappa_1 = f_2(\widetilde{\theta}_3)$, $\kappa_2 = f_2(\theta_3)$, $\kappa_3 = f_3(\theta_3) - f_2(\theta_3)$ and $\kappa_4 = f_3(\widetilde{\theta}_3) - f_2(\widetilde{\theta}_3)$ are all fixed values. Denote

$$F(\tau) = \kappa_3 \left\{ \tau^{\widetilde{\theta}_3} - (1-\tau)^{\widetilde{\theta}_3} - \kappa_1 \right\} - \kappa_4 \left\{ \tau^{\theta_3} - (1-\tau)^{\theta_3} - \kappa_2 \right\}, \tag{A4}$$

$$\dot{F}(\tau) = \kappa_3 \widetilde{\theta}_3 \{ \tau^{\widetilde{\theta}_3 - 1} + (1-\tau)^{\widetilde{\theta}_3 - 1} \} - \kappa_4 \theta_3 \{ \tau^{\theta_3 - 1} + (1-\tau)^{\theta_3 - 1} \},$$

and

$$G(\tau) = \frac{\tau^{\widetilde{\theta}_3 - 1} + (1-\tau)^{\widetilde{\theta}_3 - 1}}{\tau^{\theta_3 - 1} + (1-\tau)^{\theta_3 - 1}},$$

where $\dot{F}(\cdot)$ is the derivative function of $F(\cdot)$, and $\dot{F}(\tau) = 0$ if and only if $G(\tau) = \kappa_4 \theta_3 / (\kappa_3 \widetilde{\theta}_3)$. Note that equations (A2) and (A3) correspond to $F(\tau_1) = 0$ and $F(\tau_4) = 0$, respectively. Moreover, it can be verified that $F(\tau_2) = 0$ and $F(\tau_3) = 0$, i.e. the equation $F(\tau) = 0$ has at least four different solutions. As a result, the equation $\dot{F}(\tau) = 0$ or $G(\tau) = \kappa_4 \theta_3 / (\kappa_3 \widetilde{\theta}_3)$ has at least three different solutions. While it is implied by Lemma S1 that the equation $G(\tau) = \kappa_4 \theta_3 / (\kappa_3 \widetilde{\theta}_3)$ has at most two different solutions. Due to the contradiction, we prove that $\widetilde{\theta}_3 = \theta_3$, and it is readily to further verify that $(\widetilde{\theta}_1, \widetilde{\theta}_2) = (\theta_1, \theta_2)$. We hence accomplish the proof of Lemma 1(i). The result of Lemma 1(ii) can be proved similarly. □

REFERENCES

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

ANDREWS, D. W. K. (1987). Consistency in nonlinear econometric models: A generic uniform law of large numbers. *Econometrica* **55**, 1465–1471.

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

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

BELLONI, A., CHERNOZHUKOV, V., CHETVERIKOV, D. & FERNANDEZ-VAL, I. (2019a). Conditional quantile processes based on series or many regressors. *Journal of Econometrics* **213**, 4–29.

BELLONI, A., CHERNOZHUKOV, V. & KATO, K. (2019b). 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. & NOON, B. R. (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.

CHERNOZHUKOV, V., FERNÁNDEZ-VAL, I. & KOWALSKI, A. E. (2015). Quantile regression with censoring and endogeneity. *Journal of Econometrics* **186**, 201–221.

CHERNOZHUKOV, V., FERNÁNDEZ-VAL, I. & MELLY, B. (2020). Fast algorithms for the quantile regression process. *Empirical Econometrics* .

DAVINO, C., FURNO, M. & VISTOCCO, D. (2014). *Quantile Regression: Theory and Applications*. New York: Wiley.

DONG, C., LI, G. & FENG, X. (2019). Lack-of-fit tests for quantile regression models. *Journal of the Royal Statistical Society, Series B* **81**, 629–648.

ELSNER, J. B., KOSSIN, J. P. & JAGGER, T. H. (2008). The increasing intensity of the strongest tropical cyclones. *Nature* **455**, 92–95.

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

FENSKE, N., KNEIB, T. & HOTHORN, T. (2011). Identifying risk factors for severe childhood malnutrition by boosting additive quantile regression. *Journal of the American Statistical Association* **106**, 494–510.

FOURNIER, B., RUPIN, N., BIGERELLE, M., NAJJAR, D., IOST, A. & WILCOX, R. (2007). Estimating the parameters of a generalized lambda distribution. *Computational Statistics & Data Analysis* **51**, 2813–2835.

FREIMER, M., KOLLIA, G., MUDHOLKAR, G. S. & LIN, C. T. (1988). A study of the generalized tukey lambda family. *Communications in Statistics-Theory and Methods* **17**, 3547–3567.

GALVAO, A. F. & KATO, K. (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. & KOENKER, R. (1991). Hierarchical spline models for conditional quantiles and the demand for electricity. *Journal of the American Statistical Association* **87**, 58–68.

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

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

KAI, B., LI, R. & ZOU, H. (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., GALVAO, A. F. & MONTES-ROJAS, G. V. (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. & BASSETT, G. (1978). Regression quantiles. *Econometrica* **46**, 33–50.

KUESTER, K., MITTNIK, S. & PAOLELLA, M. S. (2006). Value-at-Risk prediction: A comparison of alternative strategies. *Journal of Financial Econometrics* **4**, 53–89.

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

LOBETO, H., MENENDEZ, M. & LOSADA, I. J. (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. & WAINWRIGHT, M. J. (2015). Regularized M-estimators with nonconvexity: Statistical and algorithmic theory for local optima. *Journal of Machine Learning Research* **16**, 559–616.

MEINSHAUSEN, N. & RIDGEWAY, G. (2006). Quantile regression forests. *Journal of Machine Learning Research* **7**, 983–999.

MOON, S. J., JEON, J.-J., LEE, J. S. H. & KIM, Y. (2021). Learning multiple quantiles with neural networks. *Journal of Computational and Graphical Statistics* **In press**.

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

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

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

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

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

SIVAKUMAR, V. & BANERJEE, A. (2017). High-dimensional structured quantile regression. In *International Conference on Machine Learning*.

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. & TSAI, C.-L. (2009). Tail index regression. *Journal of the American Statistical Association* **104**, 1233–1240.

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

WANG, H. J., LI, D. & HE, X. (2012a). Estimation of high conditional quantiles for heavy-tailed distributions. *Journal of the American Statistical Association* **107**, 1453–1464.

WANG, L. & HE, X. (2021). Analysis of global and local optima of regularized quantile regression in high dimension: a subgradient approach. Tech. rep., Miami Herbert Business School, University of Miami.

WANG, L., WU, Y. & LI, R. (2012b). Quantile regression for analyzing heterogeneity in ultra-high dimension. *Journal of the American Statistical Association* **107**, 214–222.

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

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

ZHENG, Q., PENG, L. & HE, X. (2015). Globally adaptive quantile regression with ultra-high dimensional data. *The Annals of Statistics* **43**, 2225–2258.

ZHENG, Y., ZHU, Q., LI, G. & XIAO, Z. (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. & LING, S. (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. & LI, G. (2021). Quantile double autoregression. *Econometric Theory* **In press**.

ZHU, Q., ZHENG, Y. & LI, G. (2018). Linear double autoregression. *Journal of Econometrics* **207**, 162–174.

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

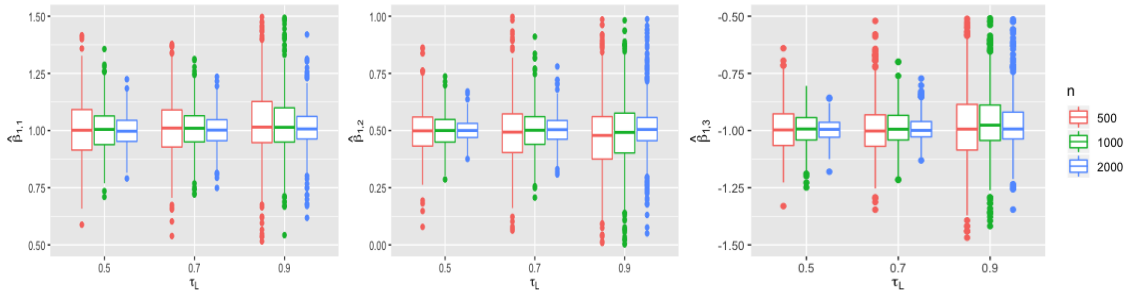

Fig. 1: Boxplots for fitted location parameters of $\widehat{\boldsymbol{\beta}}_{1,1}$ (left panel), $\widehat{\boldsymbol{\beta}}_{1,2}$ (middle panel), and $\widehat{\boldsymbol{\beta}}_{1,3}$ (right panel). Sample size is $n = 500$, 1000 or 2000, and the lower bound of quantile range $[\tau_L, \tau_U]$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

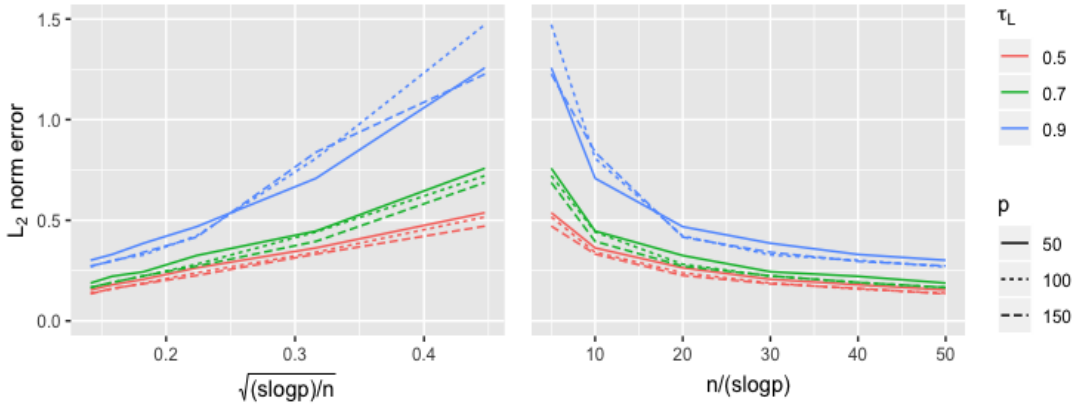

Fig. 2: Estimation errors of $\|\widetilde{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0\|$ against the quantities of $\sqrt{(s \log p)/n}$ (left panel) and $n/(s \log p)$ (right panel), respectively.

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)

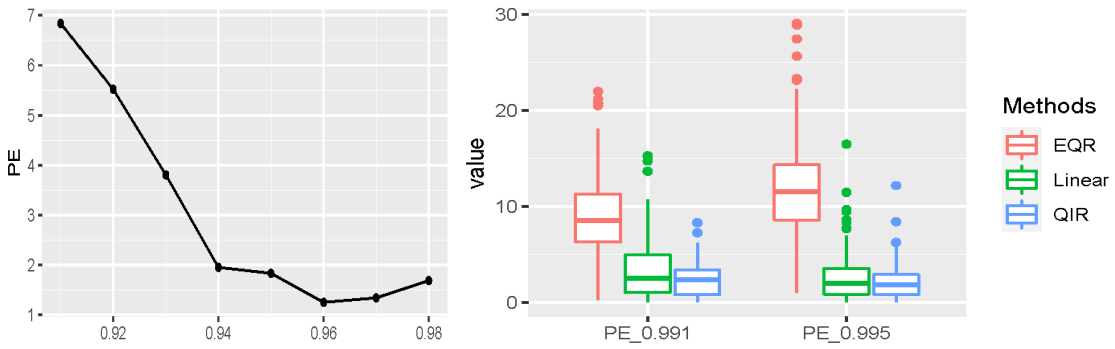

Fig. 3: Plot of PEs against τ_L (left panel) and boxplots of PEs from the extreme quantile regression (EQR), linear quantile regression (Linear) and QIR models at two target levels of $\tau^* = 0.991$ and 0.995 (right panel).

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)

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 errors.

$[\tau_L, \tau_U]$	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)

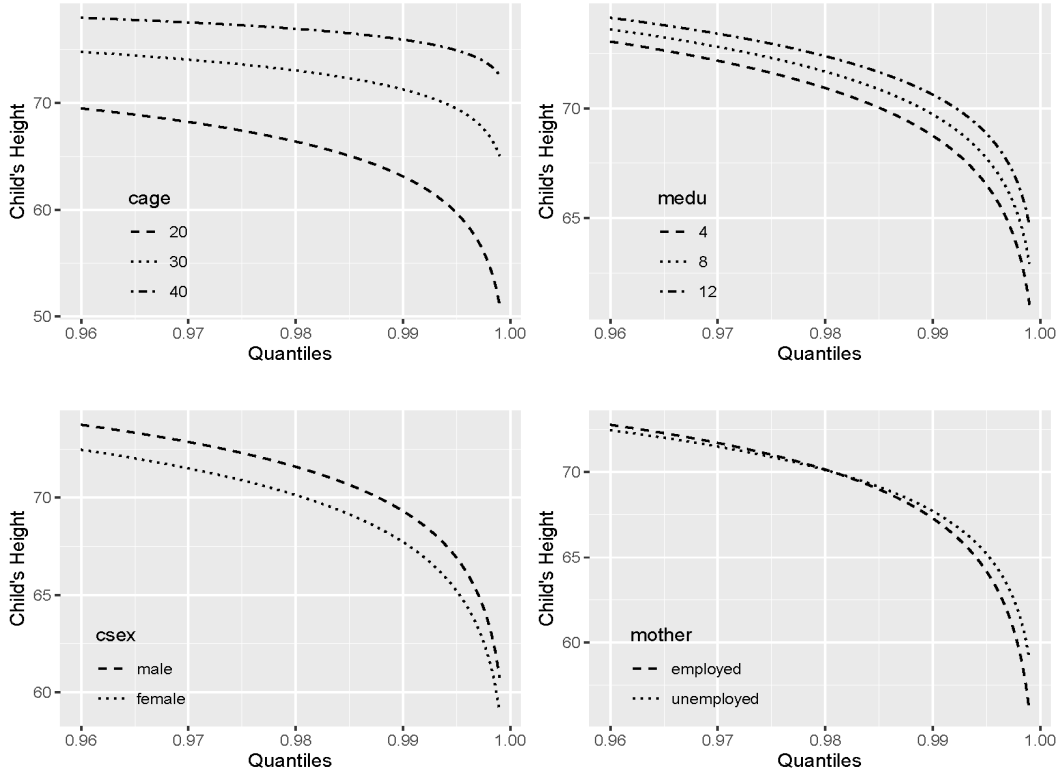

Fig. 4: Quantile curves for child's age in months (top left), mother's education in years (top right) on the three target quantiles. Effects of child's sex (bottom left) and mother's unemployment condition (bottom right).