

## OPTIMALLY COMBINED ESTIMATION FOR TAIL QUANTILE REGRESSION

Kehui Wang and Huixia Judy Wang

*PPD Inc. and George Washington University*

*Abstract:* Quantile regression offers a convenient tool to access the relationship between a response and covariates in a comprehensive way and it is appealing especially in applications where interests are on the tails of the response distribution. However, due to data sparsity, the finite sample estimation at tail quantiles often suffers from high variability. To improve the tail estimation efficiency, we consider modeling multiple quantiles jointly for cases where the quantile slope coefficients tend to be constant at the tails. We propose two estimators, the weighted composite estimator that minimizes the weighted combined quantile objective function across quantiles, and the weighted quantile average estimator that is the weighted average of quantile-specific slope estimators. By using extreme value theory, we establish the asymptotic distributions of the two estimators at the tails, and propose a procedure for estimating optimal weights. We show that the optimally weighted estimators improve the efficiency over equally weighted estimators, and the efficiency gain depends on the heaviness of the tail distribution. The performance of the proposed estimators is assessed through a simulation study and the analysis of precipitation downscaling data.

*Key words and phrases:* Efficiency, extreme value index, information aggregation, joint-quantile regression, optimal weights, regularly varying.

### 1. Introduction

An important problem in many fields is to model rare and extreme phenomena that corresponds to the lower or upper tails of a variable; see Pandey and Nguyen (1999), Friederichs (2010), Wang, Li and He (2012), Abrevaya (2001), Chernozhukov and Du (2008) for various examples. Quantile regression, first proposed by Koenker and Bassett (1978), provides a useful tool for studying the conditional tail distribution of the response. Without loss of generality, we focus on high quantile regression.

The main challenge for tail quantile regression is data sparsity, especially for heavy-tailed distributions where the disturbance of errors remains strong in the far tail. By fitting the regression model at one quantile level at a time, the *conventional (or local) quantile regression* is often unstable at the tails. In applications where the covariate effects have some common features across quantile levels in the tail region, it is desirable to aggregate information across multiple quantiles to improve the estimation efficiency over the conventional quantile-specific

estimator. For instance, Figure 1 in Section 4 plots the conventional quantile regression estimates of slopes at multiple quantiles, where the response is the observed daily precipitation in Chicago area and the predictor is the simulated daily precipitation from the ERA-40 reanalysis model. The estimated quantile slopes appear to be constant at the upper quantiles  $\tau \in [0.990, 0.995]$ . For such data sets, one could utilize the commonality of quantile slopes to improve the estimation efficiency at tails by pooling information across tail quantiles.

Consider a linear regression model with quantile-invariant covariate effects, there exist two plausible ways to combine information across quantiles: combining the local quantile estimators or the criterion functions involved in the estimation procedure at different quantiles. The first strategy leads to the weighted quantile average estimator (WQAE) introduced by Koenker and Bassett (1978), the weighted average of quantile-specific slope estimators. The second strategy leads to the weighted composite regression of quantiles (WCRQ) estimator, first proposed by Hogg (1980), that minimizes the combined quantile objective function across quantiles. For central quantiles, Koenker (1984) studied the asymptotic properties of these two estimators and showed their asymptotic equivalency when optimal weights are used. In recent years, combined quantile regression has been studied in various setups with more work focusing on the second strategy; see for instance Zou and Yuan (2008), Jiang, Jiang and Song (2012), Kai, Li and Zou (2010), Guo, Tian, and Zhu (2012), Jiang et al. (2012), Jiang et al. (2013), Jiang, Qian and Zhou (2012), Tang, Zhou and Wu (2012), and Jiang, Wang and Bondell (2013). Zhao and Xiao (2014) discussed both WQAE and WCRQ methods for linear and nonparametric regression models.

Previous works on combined quantile regression have been restricted to central quantiles with quantile level  $\tau \in [\epsilon, 1 - \epsilon]$ , where  $0 < \epsilon < 1$  is some positive constant. To our knowledge, there exists no discussion about how to optimally combine information across tail quantiles. At the tails, with quantile level  $\tau \rightarrow 1$  as the sample size  $n \rightarrow \infty$ , the convergence rate of quantile regression estimator depends on the heaviness of the tails of the response distribution and is slower than root- $n$ . In this paper, using the tools of extreme value theory, we establish the asymptotic properties of the weighted composite and weighted quantile average estimators for tail quantile regression, and propose a procedure for estimating the optimal weights for both estimators. The technical details are provided in the online supplementary material.

## 2. Proposed Methods

### 2.1. Joint-quantile regression model

Let  $Y$  be the scalar response variable,  $\mathbf{X}$  be the  $p$ -dimensional vector of covariates, and  $\{(y_i, \mathbf{x}_i)\}_1^n$  be a random sample of  $(Y, \mathbf{X})$ . Suppose we are interested

in regression at the upper tails with quantile level  $\tau \in \mathcal{T} = (1 - \epsilon_{1,n}, 1 - \epsilon_{2,n})$ , where  $\epsilon_{1,n} > \epsilon_{2,n} > 0$ ,  $\epsilon_{1,n} \rightarrow 0$  and  $n\epsilon_{2,n} \rightarrow \infty$  as  $n \rightarrow \infty$ . Let  $F_Y(\cdot|\mathbf{x})$  denote the conditional distribution function of  $Y$  given  $\mathbf{x}$ . The linear quantile regression model assumes that

$$Q_Y(\tau|\mathbf{x}) = \alpha_0(\tau) + \mathbf{x}^T \boldsymbol{\beta}_0(\tau), \quad \tau \in \mathcal{T}, \quad (2.1)$$

where  $Q_Y(\tau|\mathbf{x}) = \inf\{y : F_Y(y|\mathbf{x}) \geq \tau\}$  is the  $\tau$ th conditional quantile of  $Y$  given  $\mathbf{X} = \mathbf{x}$ , and  $\alpha_0(\tau) \in \mathbb{R}$  and  $\boldsymbol{\beta}_0(\tau) \in \mathbb{R}^p$  are the unknown quantile coefficients associated with the  $\tau$ th quantile level. At a given quantile level  $\tau$ , the conventional quantile regression estimator of  $(\alpha(\tau), \boldsymbol{\beta}(\tau))$  is defined as

$$\left(\widehat{\alpha}(\tau), \widehat{\boldsymbol{\beta}}(\tau)\right) = \underset{(\alpha, \boldsymbol{\beta}) \in \mathbb{R}^{p+1}}{\operatorname{argmin}} \sum_{i=1}^n \rho_\tau(y_i - \alpha - \mathbf{x}_i^T \boldsymbol{\beta}), \quad (2.2)$$

where  $\rho_\tau(t) = t\{\tau - I(t \leq 0)\}$  is the quantile loss function and  $I(\cdot)$  is the indicator function (Koenker (2005)). The conventional quantile regression method estimates the quantile coefficient at each quantile level of interest separately, and the resulting local slope estimator  $\widehat{\boldsymbol{\beta}}(\tau)$  can vary freely in  $\tau$ . However, in data-sparse area such as the extreme tails, the variability of local estimates is often overly high. In some applications it might be reasonable to assume the slope coefficient  $\boldsymbol{\beta}(\tau)$  to share some common features, for instance, to be constant or locally linear, within a certain region of quantiles. By utilizing this commonality, we can aggregate information across quantiles to improve the estimation efficiency. We focus on linear quantile regression with constant slopes at the upper tails, but the proposed method can be adapted to accommodate such other common features as local linearity, and cases where only a subset of the components of  $\boldsymbol{\beta}(\tau)$  are locally constant.

We assume the following linear quantile regression model at the upper tails

$$Q_Y(\tau|\mathbf{x}) = \alpha_0(\tau) + \mathbf{x}^T \boldsymbol{\beta}_0, \quad \tau \in \mathcal{T}. \quad (2.3)$$

Different from (2.1), here the quantile slope is assumed to be constant at the upper tails across  $\tau \in \mathcal{T}$ , with  $\alpha_0(\tau)$  still an increasing function of  $\tau$ . Based on the extreme value theory, we develop two optimally combined estimators for tail quantile regression, which are obtained by aggregating information across  $K$  quantiles  $\tau_1 < \dots < \tau_K \in \mathcal{T}$ .

function ‘anova.rq’ in R package *quantreg* (Koenker (2005), Page 76).

## 2.2. Optimally weighted quantile average estimator

For (2.3) at  $K$  upper quantiles, the unknown parameters consist of  $K$  distinct intercepts  $\alpha_0(\tau_k)$  and one common slope vector  $\boldsymbol{\beta}_0$ . Denote the vector of true

parameters  $\boldsymbol{\theta}_0 = (\alpha_{0,1}, \dots, \alpha_{0,K}, \boldsymbol{\beta}_0^T)^T$ , where  $\alpha_{0,k} = \alpha_0(\tau_k)$ ,  $k = 1, \dots, K$ . Let  $\varpi_k$  be the weight assigned to  $\tau_k$ ,  $k = 1, \dots, K$ . For identifiability, we assume that  $\mathbf{1}_K^T \boldsymbol{\varpi} = 1$ , where  $\boldsymbol{\varpi} = (\varpi_1, \dots, \varpi_K)^T$ , and  $\mathbf{1}_K$  denotes a  $K$ -dimensional vector of ones. We take the weighted quantile average estimator of  $\boldsymbol{\theta}$  as

$$\widehat{\boldsymbol{\beta}}_{\text{WQAE}} = \sum_{k=1}^K \varpi_k \widehat{\boldsymbol{\beta}}(\tau_k), \quad (2.4)$$

where  $\widehat{\boldsymbol{\beta}}(\tau_k)$  is the local quantile slope estimator obtained by minimizing the objective function in (2.2) at the  $\tau_k$ th quantile. The WQAE is the weighted average over conventional estimators at multiple high quantiles, and thus can be viewed as a special case of the L-estimator with discrete weights. Portnoy and Koenker (1989) and Koenker (1984) studied the L-estimator for the slope in the location-shift linear model, which implies that  $\boldsymbol{\beta}(\tau)$  is constant across the entire quantile region  $\tau \in (0, 1)$ .

The asymptotic properties of WQAE at central quantiles have been studied by Koenker (1984) and Zhao and Xiao (2014). In our Theorem 1, we establish the asymptotic distribution of  $\widehat{\boldsymbol{\beta}}_{\text{WQAE}}$  at the tails. For any sequences  $a(z)$  and  $b(z)$ , we use the notation  $a(z) \sim b(z)$  to mean that  $a(z)/b(z) \rightarrow 1$  as a specified limit is taken over  $z$ . We make the following assumptions.

- A1.** The distribution function  $F_Y(y|\mathbf{x}_i)$  is absolutely continuous with continuous density  $f_Y(y|\mathbf{x}_i)$  that is uniformly bounded away from zero and infinity and has a bounded first derivative around  $Q_Y(\tau_k|\mathbf{x}_i)$  for any  $k = 1, \dots, K$ .
- A2.** The distribution of  $\mathbf{X}$  has a compact support  $\mathcal{X}$ . The expectation  $E(\mathbf{X}) = \mathbf{0}$ , and  $E(\mathbf{X}\mathbf{X}^T) = \mathbf{D}$  exists and is positive definite.
- A3.** Let  $U = Y - \mathbf{X}^T\boldsymbol{\beta}_0$ . There exists some distribution function  $F_0(\cdot)$  in the maximum domain of attraction with extreme value index  $\xi$ , such that  $1 - F_U(z|\mathbf{x}) \sim 1 - F_0(z)$  as  $z \rightarrow s_u$  uniformly in  $\mathbf{x} \in \mathcal{X}$ , where  $s_u$  is the upper end-point of  $U$ .
- A4.**  $\partial F_0^{-1}(1 - \tau)/\partial \tau$  is regularly varying at 0 with exponent  $-\xi - 1$ .
- A5.** For  $k = 1, \dots, K$ ,  $(1 - \tau_k)/(1 - \tau) = l_k + o(1)$  for some  $\tau \rightarrow 1$  and  $n(1 - \tau) \rightarrow \infty$ , where the  $l_k > 0$  are constants.

Here **A3** requires that  $Y$ , after linear transformation, is tail equivalent to a distribution  $F_0$  in the maximum domain of attraction. The domain of attraction assumption covers such common distributions as the Gaussian, Beta, and  $t$  distribution; see de Haan and Ferreira (2006) for more details about domains of attraction. Condition **A4** is a von Mises-type condition that specifies how the density function decays at the right tail. It is basic for a distribution to belong to the maximum domain of attraction. Condition **A5** restricts attention to the

intermediate order of extreme quantiles with  $\tau_k \rightarrow 1$  and  $n(1 - \tau_k) \rightarrow \infty$  as  $n \rightarrow \infty$ . In finite sample analysis, we can let  $\tau = \tau_1$  and  $l_k = (1 - \tau_k)/(1 - \tau_1)$ . We define

$$a_n = \frac{\sqrt{(1 - \tau)n}}{F_0^{-1}(\tau) - F_0^{-1}(\tilde{\tau}_m)}, \text{ where } \tilde{\tau}_m = 1 - m(1 - \tau) \text{ for some } m > 1.$$

**Theorem 1.** *Suppose that (2.3) and A1–A5 hold. With  $\varpi = (\varpi_1, \dots, \varpi_K)^T$ , as  $n \rightarrow \infty$ ,*

$$a_n(\hat{\beta}_{WQAE} - \beta) \xrightarrow{d} N\left(0, \sigma_{WQAE}^2(\varpi) \begin{pmatrix} m^{-\xi} - 1 \\ -\xi \end{pmatrix}^{-2} \mathbf{D}^{-1}\right),$$

where  $\sigma_{WQAE}^2(\varpi) = \varpi^T \Phi^{-1}(\xi) \mathbf{\Gamma} \Phi^{-1}(\xi) \varpi$ ,  $\mathbf{\Gamma}$  is a  $K \times K$  matrix with the  $(k, k')$ th element  $\min(l_k, l_{k'})$ ,  $\Phi(\xi) = \text{diag}\{l_1^{\xi+1}, \dots, l_K^{\xi+1}\}$ , and  $\mathbf{D} = E(\mathbf{X}\mathbf{X}^T)$ .

Theorem 1 suggests that the asymptotic covariance of  $\hat{\beta}_{WQAE}$  depends on the weights  $\varpi$  only through a scalar function  $\sigma_{WQAE}^2(\varpi)$  of  $\varpi$ . Therefore, the optimal weight that maximize the efficiency of  $\hat{\beta}_{WQAE}$  is

$$\varpi_{\text{opt}} = \underset{\varpi}{\text{argmin}} \varpi^T \Phi^{-1}(\xi) \mathbf{\Gamma} \Phi^{-1}(\xi) \varpi, \text{ subject to } \mathbf{1}_K^T \varpi = 1. \quad (2.5)$$

The minimization in (2.5) is a standard constrained optimization problem, with

$$\varpi_{\text{opt}} = \Phi(\xi) \mathbf{\Gamma}^{-1} \phi(\xi) \{\phi^T(\xi) \mathbf{\Gamma}^{-1} \phi(\xi)\}^{-1}. \quad (2.6)$$

Therefore, the minimal value of  $\sigma_{WQAE}^2$  is  $\{\phi^T(\xi) \mathbf{\Gamma} \phi(\xi)\}^{-1}$ . We refer to the optimal WQAE of  $\beta_0$  based on  $\varpi_{\text{opt}}$  as  $\hat{\beta}_{\text{OWQAE}}$ .

### 2.3. Weighted composite quantile estimator

An alternative estimator of the common slope  $\beta$  is the weighted composite regression of quantiles (WCRQ) defined as

$$\hat{\theta}_{WCRQ} = \underset{\alpha_1, \dots, \alpha_K, \beta}{\text{argmin}} \sum_{k=1}^K \omega_k \sum_{i=1}^n \rho_{\tau_k}(y_i - \alpha_k - \mathbf{x}_i^T \beta), \quad (2.7)$$

where  $\omega_k$  is the weight assigned to  $\tau_k$ . The estimator  $\hat{\theta}_{WCRQ}$  is referred to as the weighted composite regression of quantiles (WCRQ) estimator.

Here  $\hat{\theta}_{WCRQ}$  depends on the prespecified weights  $\omega_k$ . Focusing on central quantiles, Zou and Yuan (2008), Kai, Li and Zou (2010), Jiang, Wang and Bondell (2013) considered a combined quantile objective function that is similar to that in (2.7), but assigns equal weights to different quantiles. Since neighboring quantiles

are correlated, especially at the tails, assigning equal weights is not an efficient way to combine information across quantiles in general.

When the components in  $\boldsymbol{\omega}_{\text{opt}}$  are non-negative, the combined objective function in (2.7) is convex. The optimization in (2.7) can be recast into a linear programming problem and solved by using existing software such as the function “make.lp” in the R package *lpSolveAPI* or function “rq.fit.fnb” in the R package *quantreg*. When some weights are negative, the objective function in (2.7) might not be convex, leading to difficulty in minimization. To avoid potential computational difficulty, we first consider WCRQ with nonnegative weights.

**Theorem 2.** *If (2.3) and A1-A5 hold, and  $\omega_k \geq 0$  for  $k = 1, \dots, K$ , then as  $n \rightarrow \infty$ ,*

$$a_n(\widehat{\boldsymbol{\beta}}_{\text{WCRQ}} - \boldsymbol{\beta}_0) \xrightarrow{d} N\left(0, \sigma_{\text{WCRQ}}^2(\boldsymbol{\omega}) \begin{pmatrix} m^{-\xi} - 1 \\ -\xi \end{pmatrix}^{-2} \mathbf{D}^{-1}\right),$$

$$\sigma_{\text{WCRQ}}^2(\boldsymbol{\omega}) = \{\boldsymbol{\omega}^T \boldsymbol{\phi}(\xi)\}^{-2} \boldsymbol{\omega}^T \boldsymbol{\Gamma} \boldsymbol{\omega}, \text{ where } \boldsymbol{\phi}(\xi) = (l_1^{\xi+1}, \dots, l_K^{\xi+1})^T.$$

**Sub-optimal Estimator.** Similar to WQAE, the optimal weights for WCRQ can be obtained by minimizing the asymptotic variance of  $\widehat{\boldsymbol{\beta}}_{\text{WCRQ}}$ , which depends on the weights only through a scalar  $\sigma_{\text{WCRQ}}^2(\boldsymbol{\omega})$ . Let

$$\boldsymbol{\omega}_{\text{sub}} = \underset{\omega_1 \geq 0, \dots, \omega_K \geq 0}{\operatorname{argmin}} \{\boldsymbol{\omega}^T \boldsymbol{\phi}(\xi)\}^{-2} \boldsymbol{\omega}^T \boldsymbol{\Gamma} \boldsymbol{\omega}, \text{ subject to } \mathbf{1}_K^T \boldsymbol{\omega} = \mathbf{1}. \quad (2.8)$$

We refer to  $\boldsymbol{\omega}_{\text{sub}}$  as the sub-optimal weight since it minimizes the asymptotic variance of  $\widehat{\boldsymbol{\beta}}_{\text{WCRQ}}$  under the nonnegative restriction. The optimization in (2.8) is a standard quadratic programming problem, and can be solved by existing quadratic programming software such as the function “solve.QP” in the R package *quadprog*. We denote the minimizer of (2.7) based on  $\boldsymbol{\omega}_{\text{sub}}$  as  $\widehat{\boldsymbol{\theta}}_{\text{WCRQ+}}$ .

Compared with OWQAE, the additional restriction in  $\widehat{\boldsymbol{\theta}}_{\text{WCRQ+}}$  can cause some loss of efficiency for the estimation of  $\boldsymbol{\beta}$  as indicated in Proposition 1. Without the nonnegative constraint, we find

$$\boldsymbol{\omega}_{\text{opt}} = \frac{\boldsymbol{\Gamma}^{-1} \boldsymbol{\phi}(\xi)}{\{\mathbf{1}_K^T \boldsymbol{\Gamma}^{-1} \boldsymbol{\phi}(\xi)\}}. \quad (2.9)$$

Following the argument in the proof of Theorem 2, we can show that the asymptotic normality in Theorem 2 holds for the WCRQ estimator based on the weights  $\boldsymbol{\omega}_{\text{opt}}$ , and with the same asymptotic variance. We refer to the WCRQ estimator based on  $\boldsymbol{\omega}_{\text{opt}}$  as the OWCRQ (optimally weighted composite regression of quantile).

**Proposition 1.** *For any weight  $\boldsymbol{\omega} \in \mathbb{R}^K$ ,  $\sigma_{\text{WCRQ}}^2(\boldsymbol{\omega}) \geq \{\boldsymbol{\phi}^T(\xi) \boldsymbol{\Gamma} \boldsymbol{\phi}(\xi)\}^{-1}$ , and the equality holds if and only if  $\boldsymbol{\omega} = \boldsymbol{\omega}_{\text{opt}}$ .*

**Remark 1.** In Theorems 1 and 2, the normalizing sequence  $a_n$  and the asymptotic variances involve  $m$  while the optimal weights  $\varpi_{\text{opt}}$  and  $\omega_{\text{opt}}$  do not. Thus users need not choose  $m$  for computing  $\hat{\beta}_{\text{OWQAE}}$  and  $\hat{\beta}_{\text{OWCRQ}}$ .

**Remark 2.** For tail quantile regression, the occurrence of negative optimal weights in  $\omega_{\text{opt}}$  depends on the heaviness of the tail of the response distribution. We can show that when the response distribution is heavy-tailed with  $\xi > 0$ , the least extreme quantile receives positive weight and all the others negative weights. For light-tailed distributions with  $\xi = 0$ , as  $\tau \rightarrow 1$ , all the weights are put into the least extreme quantile level  $\tau_1$  while other higher quantiles receive zero weight. For short-tailed distributions with  $\xi < 0$ , all the weights at high quantiles are positive. Some justifications are provided in the supplementary material. Table 2 gives the optimal weights for normal,  $t_2$ ,  $t_1$ , Beta(2, 5), and Beta(2, 2.5) distributions.

**One-step Estimator.** When  $\omega_{\text{opt}}$  contains negative weights, the weighted objective function in (2.7) may be non-convex, making it hard to minimize. To utilize the optimal weights  $\omega_{\text{opt}}$  and avoid the non-convex optimization, we consider an alternative one-step estimator. Let  $\mathbf{z}_{i,k} = (\mathbf{e}_k^T, \mathbf{x}_i^T)^T$ , where  $\mathbf{e}_k$  is a  $K$ -dimensional vector with the  $k$ th entry 1 and the others 0. Let

$$A(\boldsymbol{\theta}) = \sum_{k=1}^K \sum_{i=1}^n \omega_k^{(o)} \mathbf{z}_{i,k} \{I(y_i - \mathbf{z}_{i,k}^T \boldsymbol{\theta} < 0) - \tau_k\}, B(\boldsymbol{\theta}) = \sum_{k=1}^K \sum_{i=1}^n \omega_k^{(o)} \mathbf{z}_{i,k} \mathbf{z}_{i,k}^T f_Y(\mathbf{z}_{i,k}^T \boldsymbol{\theta} | \mathbf{x}_i),$$

where  $\omega_k^{(o)}$  is the  $k$ th element of  $\omega_{\text{opt}}$ . The one-step estimator of  $\boldsymbol{\theta}_0$  is

$$\hat{\boldsymbol{\theta}}_{\text{OS}} = \tilde{\boldsymbol{\theta}} - \{B(\tilde{\boldsymbol{\theta}})\}^{-1} A(\tilde{\boldsymbol{\theta}}), \quad (2.10)$$

where  $\tilde{\boldsymbol{\theta}}$  is any  $a_n$ -consistent estimator of  $\boldsymbol{\theta}_0$ .

This one-step approach was first discussed by Bickel (1975) for estimating the location parameter in a linear model. Recently, Bradic, Fan, and Wang (2011) studied the one-step method for variable selection based on penalized composite quasi-likelihood. Following their arguments, we can show that for tail quantiles, the one-step slope estimator  $\hat{\beta}_{\text{OS}}$  achieves the same asymptotic efficiency as the OWQAE and OWCRQ estimators as long as the initial slope estimator  $\tilde{\beta}$  is consistent of the same rate  $a_n$ .

**Theorem 3.** If **A1-A5** hold and  $\|\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\| = O_p(1/a_n)$ , then

$$a_n(\hat{\beta}_{\text{OS}} - \beta_0) \xrightarrow{d} N\left(\mathbf{0}, \frac{1}{\boldsymbol{\phi}^T(\xi)\boldsymbol{\Gamma}\boldsymbol{\phi}(\xi)} \left(\frac{m^{-\xi} - 1}{-\xi}\right)^{-2} \mathbf{D}^{-1}\right).$$

To achieve this asymptotic efficiency, the initial estimator  $\tilde{\boldsymbol{\theta}}$  can be any  $a_n$ -consistent estimator, such as the conventional quantile regression estimator or  $\hat{\boldsymbol{\theta}}_{WCRQ}$  with any given nonnegative weights. In our implementation, we use  $\hat{\boldsymbol{\theta}}_{WCRQ+}$  as the initial estimator.

In (2.10),  $B(\boldsymbol{\theta})$  involves the unknown quantity  $f_Y(\mathbf{z}_{i,k}^T \boldsymbol{\theta} | \mathbf{x}_i) = f_Y\{Q_Y(\tau_k | \mathbf{x}_i) | \mathbf{x}_i\}$ , which has to be estimated in practice. In general, we can estimate  $f_Y\{Q_Y(\tau | \mathbf{x}) | \mathbf{x}\}$  by using the difference quotient method proposed by Hendricks and Koenker (1992), or estimate  $\sum_{i=1}^n \mathbf{z}_{i,k} \mathbf{z}_{i,k}^T f_Y(\mathbf{z}_{i,k}^T \boldsymbol{\theta} | \mathbf{x}_i)$  by using the kernel method proposed by Powell (1991). Our numerical investigation suggests that these two methods work well for large sample sizes but they sometimes lead to unstable results for small samples. Model (2.3) and **A3** imply that as  $\tau \rightarrow 1$ ,  $f_Y\{Q_Y(\tau | \mathbf{x}) | \mathbf{x}\} = f_U\{\alpha(\tau) | \mathbf{x}\} \sim f_0\{\alpha(\tau)\}$ , which is common across  $\mathbf{x}$ . Therefore, we suggest using the nonparametric kernel density estimation based on the estimated residuals  $\hat{\epsilon}_i = y_i - \mathbf{x}_i^T \hat{\boldsymbol{\beta}}_{WCRQ+}$  as in Zhao and Xiao (2014).

The calculation of WQAE requires only minimizing the convex quantile objective function in (2.2) at each quantile level  $\tau_k$  separately. Therefore, negative weights in  $\boldsymbol{\varpi}_{\text{opt}}$  do not cause any computational difficulty. In contrast, the WCRQ method requires solving the combined objective function, and a one-step iteration is needed for calculating the optimal estimator when some of the optimal weights are negative. Despite the possible computational complication, the WCRQ method has advantages. First, it can be used to accommodate general interquantile commonality in a more direct way, for instance, locally linear quantile slopes with  $\boldsymbol{\beta}(\tau_k) = \boldsymbol{\beta}(\tau_1) + (\tau_k - \tau_1)\boldsymbol{\gamma}$ , where  $\boldsymbol{\gamma}$  is an unknown parameter. Further, penalization can be incorporated in the weighted composite quantile loss function for variable selection and inter-quantile shrinkage; see for instance Bradic, Fan, and Wang (2011), Jiang, Wang and Bondell (2013), and Jiang, Bondell and Wang (2014).

## 2.4. Estimation of extreme value index

The optimal and sub-optimal weights involve the unknown extreme value index  $\xi$ , a measurement of the heaviness of the tail distribution. We estimate  $\xi$  by the maximum likelihood estimator. Suppose  $z_1, \dots, z_n$  is a random sample of  $Z$  that has distribution function  $F$ . Pickands (1975) showed that if  $F$  is in the domain of attraction with the extreme value index  $\xi$  then, as  $u \rightarrow z_F$ , the endpoint of  $Z$ ,  $P(Z \leq u + z | Z > u)$  converges to the distribution function of the generalized Pareto distribution (GPD) with parameters  $\boldsymbol{\Psi} = (\xi, \sigma, u)^T$ . The density function of the GPD can be written as

$$g(z | \boldsymbol{\Psi}) = \begin{cases} \frac{1}{\sigma} \left\{ 1 + \frac{\xi(z-u)}{\sigma} \right\}^{-(1+\xi)/\xi}, & \text{if } \xi \neq 0, \\ \frac{1}{\sigma} \exp\left\{-\frac{(z-u)}{\sigma}\right\}, & \text{if } \xi = 0, \end{cases}$$



Table 1. The asymptotic relative efficiency of QAE, CRQ, WCRQ+ and OWCRQ with respect to OWQAE (OWCRQ) for five types of distributions.

Tail type	QAE	CRQ	WCRQ+
Light ( $\xi = 0$ )	0.64	0.81	1.00
Heavy ( $\xi = 0.5$ )	0.23	0.52	0.90
Heavy ( $\xi = 1$ )	0.06	0.33	0.77
Short ( $\xi = -0.2$ )	0.84	0.92	1.00
Short ( $\xi = -0.4$ )	0.95	0.95	1.00

where  $z > u$  for  $\xi \geq 0$  and  $0 \leq z - u \leq -\sigma/\xi$  for  $\xi < 0$ . that is,  $\{z_i - u : z_i \geq u, i = 1, \dots, n\}$ . it was shown that the maximum likelihood estimator of  $\xi$  is consistent and asymptotically normal; see Smith (1987) and Theorem 3.4.2 in de Haan and Ferreira (2006) for details. In our setup, we take  $u_i = \widehat{Q}_Y(\tau_0|\mathbf{x}_i), i = 1, \dots, n$ , where  $\tau_0 \rightarrow 1$  as  $n \rightarrow \infty$ , and  $\widehat{Q}_Y(\tau_0|\mathbf{x})$  is any consistent estimator of  $Q_Y(\tau_0|\mathbf{x})$ , for instance, the conventional quantile regression estimator. We then estimate  $\xi$  by maximizing the GPD likelihood based on the exceedances  $\{y_i - u_i : y_i \geq u_i, i = 1, \dots, n\}$ . If  $F_0(\cdot)$  satisfies the second-order condition on page 44 of de Haan and Ferreira (2006) and  $\xi > -1/2$ , it follows by Smith (1987) and Theorem 3.4.2 in de Haan and Ferreira (2006) that the maximum likelihood estimator of  $\xi$  is consistent and asymptotically normal. Throughout our numerical studies, we choose  $\tau_0 = 0.95$  and obtain the maximum likelihood estimator of  $\xi$  by using the function “gpd.fit” in the R package *ismev*.

## 2.5. Comparison of asymptotic efficiency

Theorems 1–3 suggest that OWQAE, OWCRQ, and the one-step estimator achieve the same asymptotic efficiency, while WCRQ+ is asymptotically less efficient when some of the optimal weights are negative.

We assess the efficiency gain of using optimal weights by comparing the asymptotic efficiency of OWCRQ/OWQAE, WCRQ+ with CRQ and QAE, the composite and quantile average estimators based on equal weights. We consider five cases:  $\xi = 0$  corresponding to such light-tailed as the exponential and normal,  $\xi = 0.5$  and 1 corresponding to such heavy-tailed distributions as  $t_2$  and  $t_1$ ,  $\xi = -0.2$  and  $-0.4$  corresponding to short-tailed distributions such as Beta(2, 5) and Beta(2, 2.5). We consider quantiles  $\tau_k = 0.95, 0.96, 0.97, 0.98$  and  $0.99$ . Table 1 summarizes the asymptotic relative efficiency of QAE, CRQ and WCRQ+ with respect to OWQAE (or equivalently OWCRQ), and Table 2 presents the optimal weights for the WCRQ and WQAE methods.

Table 2 suggests that, for the optimal WCRQ and WQAE methods, less weights (and sometimes even negative weights) are put on coefficients corresponding to the more extreme quantiles. This pattern is in line with Remark 2.

Table 2. Optimal weights for the WCRQ and WQAE methods at  $\tau_k = 0.95, 0.96, 0.97, 0.98$  and  $0.99$ .

Tail type	$\omega_{\text{opt}}$ for WCRQ					$\varpi_{\text{opt}}$ for WQAE				
	0.95	0.96	0.97	0.98	0.99	0.95	0.96	0.97	0.98	0.99
Light ( $\xi = 0$ )	1	0	0	0	0	1	0	0	0	0
Heavy ( $\xi = 0.5$ )	2.96	-0.34	-0.39	-0.48	-0.74	1.26	-0.10	-0.08	-0.05	-0.03
Heavy ( $\xi = 1$ )	7.41	-1.60	-1.60	-1.60	-1.60	1.35	-0.19	-0.11	-0.05	-0.12
Short ( $\xi = -0.2$ )	0.61	0.03	0.04	0.07	0.25	0.80	0.03	0.04	0.04	0.09
Short ( $\xi = -0.4$ )	0.35	0.03	0.05	0.10	0.47	0.53	0.05	0.06	0.09	0.28

When the quantile level gets more extreme, the local estimator becomes more unstable, and assigning less weights can reduce the variance of the weighted estimator.

Table 1 shows that the optimal estimators have higher efficiency across different types of distributions when compared to the estimators QAE and CRQ with equal weights. The efficiency gains of OWCRQ and OWQAE also depend on the heaviness of the tail, and are more pronounced for heavy-tailed distributions. One explanation is that, for heavy-tailed error distributions, the data sparsity leads to larger variances to the local quantile regression estimator at more extreme tails. The optimally weighted estimators improve efficiency by assigning more weights to the less extreme quantiles. For distributions with light and short tails, estimators based on equal weights lose some efficiency but the efficiency loss is not as substantial as for heavy-tailed distributions, and OWCRQ and WCRQ+ are the same since all the optimal weights are non-negative. For heavy-tailed distributions, optimal weights outperform sub-optimal weights.

### 3. Simulation Study

To demonstrate the finite sample performance of the proposed methods, we conducted simulations. We considered: univariate predictors with constant quantile slope, univariate predictors with constant quantile slope only at upper quantiles, and multivariate predictors with constant slopes across quantiles.

**Example 1.** The data were generated from

$$y_i = x_i\beta + \epsilon_i, \quad i = 1, \dots, n, \quad (3.1)$$

where  $x_i \sim N(0, 1)$ ,  $\beta = 1$ , and the  $\epsilon_i$  independent and identically distributed. Here the  $\tau$ th conditional quantile of  $Y$  is  $Q_Y(\tau|x) = F_\epsilon^{-1}(\tau) + x$ , where  $F_\epsilon$  is the cumulative distribution function of  $\epsilon_i$ .

**Example 2.** The slope is constant only at the upper quantiles with  $\tau > 0.9$ . The quantile function is  $Q_Y(\tau|x) = \alpha(\tau) + \beta(\tau)x$ , where  $\alpha(\tau) = F_\epsilon^{-1}(\tau)$  and

$$\beta(\tau) = \begin{cases} \beta - F_\epsilon^{-1}(0.90) + F_\epsilon^{-1}(\tau) & \text{if } 0 < \tau < 0.90, \\ \beta & \text{if } 0.90 \leq \tau < 1, \end{cases} \quad (3.2)$$

with  $\beta = 1$ . To generate the data, we first generated  $x_i \sim U(0, 1)$  and quantile levels  $u_i \sim U(0, 1)$ , and then let  $y_i = \alpha(u_i) + \beta(u_i)x_i$ ,  $i = 1, \dots, n$ . Therefore,  $\beta(\tau)$  varies for  $\tau < 0.9$ , but it is constant for  $\tau \geq 0.9$ .

**Example 3.** We generated data from

$$y_i = x_{i,1}\beta_1 + x_{i,2}\beta_2 + \epsilon_i, \quad i = 1, \dots, n, \quad (3.3)$$

where  $x_{i,1} \sim N(0, 1)$ ,  $x_{i,2} \sim N(0, 1)$ ,  $\beta_1 = 1$  and  $\beta_2 = 2$ , and  $\epsilon_i$  are i.i.d. with distribution  $F_\epsilon$ . The  $\tau$ th conditional quantile of  $Y$  given  $x_1$  and  $x_2$  is  $Q_Y(\tau|x_1, x_2) = F_\epsilon^{-1}(\tau) + x_1 + 2x_2$ , and the slopes are constant across  $\tau \in (0, 1)$ .

For these examples, we took  $\epsilon$  to be standard normal,  $t_2$ , and Beta(2, 5). We considered sample sizes:  $n=500$  and 1,000. We chose the equally spaced extreme upper quantiles as  $\tau_k = 1 - (6 - k)n^{-3/4}$ , where  $k = 1, \dots, 5$ . The simulation was repeated 500 times for each scenario.

Five estimators of  $\beta$  are included for comparison: the quantile average estimator with equal weights (QAE), the composite estimator with equal weights (CRQ), the optimally weighted quantile average estimator (OWQAE), the one-step composite estimator (referred to as OWCRQ as the two have the same asymptotic efficiency), and the weighted composite estimator based on the sub-optimal nonnegative weights (WCRQ+). For the OWQAE, OWCRQ and WCRQ+ methods, the extreme value index  $\xi$  was estimated by the maximum likelihood estimator introduced in Section 2.3.

Tables 3 and 4 summarize the mean squared error (MSE) of different estimators of  $\beta$  in Examples 1 and 2. For Example 3, we report the mean integrated squared error

$$\text{MISE} = \frac{1}{500} \sum_{j=1}^{500} \{(\hat{\beta}_{j1} - \beta_1)^2 + (\hat{\beta}_{j2} - \beta_2)^2\},$$

where  $\hat{\beta}_{j1}$  and  $\hat{\beta}_{j2}$  are the estimators of  $\beta_1$  and  $\beta_2$  from the  $j$ th simulation. Since  $t_2$  is a heavy-tailed distribution with infinite variance while the normal has variance 1 and the Beta has variance 0.026, MSE and MISE under the  $t_2$  distribution have much larger magnitudes.

Tables 3–5 OWCRQ and OWQAE have similar performance, which agrees with the asymptotic theory. For the heavy-tailed  $t_2$  distribution, OWQAE and OWCRQ are more efficient than WCRQ+. For the  $t_2$  distribution, the three methods involving weight estimation perform significantly better than the two methods QAE and CRQ based on equal weights. For normal errors, optimally

Table 3. The  $10^3 \times \text{MSE}$  of different estimators of  $\beta$  in Example 1. The values in the parentheses are the standard errors of  $10^3 \times \text{MSE}$ .

Distribution	$n$	QAE	<b>OWQAE</b>	CRQ	<b>WCRQ+</b>	<b>OWCRQ</b>
Normal	500	11.54 (0.76)	10.29 (0.72)	10.52 (0.74)	9.96 (0.66)	9.96 (0.66)
	1,000	8.82 (0.61)	7.16 (0.51)	7.78 (0.54)	7.19 (0.52)	7.17 (0.52)
$t_2$	500	349.89 (20.08)	101.01 (6.43)	173.39 (11.01)	104.41 (6.71)	100.58 (6.40)
	1,000	628.92 (39.72)	148.91 (8.78)	302.17 (19.09)	167.87 (10.22)	159.20 (9.69)
Beta(2, 5)	500	0.39 (0.03)	0.39 (0.03)	0.37 (0.03)	0.39 (0.02)	0.40 (0.03)
	1,000	0.26 (0.01)	0.25 (0.01)	0.24 (0.01)	0.24 (0.01)	0.24 (0.01)

Table 4. The  $10^3 \times \text{MSE}$  of different estimators of  $\beta$  in Example 2. The values in the parentheses are the standard errors of  $10^3 \times \text{MSE}$ .

Distribution	$n$	QAE	<b>OWQAE</b>	CRQ	<b>WCRQ+</b>	<b>OWCRQ</b>
Normal	500	129.68 (8.48)	124.94 (8.22)	114.67 (7.54)	119.63 (7.96)	119.66 (7.90)
	1,000	99.75 (5.94)	88.31 (5.60)	89.78 (5.28)	85.21 (5.29)	85.09 (5.25)
$t_2$	500	5493.83 (406.81)	1459.47 (95.32)	2452.45 (161.92)	1498.51 (99.53)	1467.96 (97.95)
	1,000	8044.79 (518.23)	2116.36 (144.46)	3807.19 (223.98)	2241.42 (145.35)	2182.90 (143.02)
Beta(2, 5)	500	4.25 (0.28)	4.54 (0.29)	3.95 (0.26)	4.54 (0.30)	4.57 (0.30)
	1,000	2.97 (0.18)	2.88 (0.18)	2.86 (0.17)	2.79 (0.17)	2.79 (0.16)

weighted methods still have some efficiency gain over the equally weighted estimators but the gain is less obvious than it is for the  $t_2$  distribution. For the short-tailed Beta(2, 5) distribution, the optimally weighted methods have no obvious efficiency gain, and this agrees with our asymptotic efficiency comparison in Section 2.5. For the  $t_2$  distribution, MSE with  $n = 1,000$  is surprisingly higher than that with  $n = 500$ . That is due to the fact that larger  $n$  is associated with more extreme quantiles  $\tau_k$ , and the gain from increased sample size is less than the loss due to the increased disturbance at the more extreme tail for this heavy-tailed distribution.

Table 5. The  $10^3 \times \text{MISE}$  of different estimators in Example 3. The values in the parentheses are the standard errors of  $10^3 \times \text{MISE}$ .

Distribution	$n$	QAE	OWQAE	CRQ	WCRQ+	OWCRQ
Normal	500	22.68 (1.10)	21.15 (1.03)	20.66 (0.98)	20.66 (1.00)	20.66 (0.99)
	1,000	17.40 (0.75)	13.63 (0.63)	15.16 (0.67)	13.46 (0.63)	13.36 (0.63)
$t_2$	500	780.16 (38.45)	231.48 (11.24)	428.49 (22.78)	249.95 (12.00)	240.52 (11.58)
	1,000	1007.67 (42.22)	290.43 (12.88)	537.08 (24.15)	312.2 (13.55)	298.63 (12.97)
Beta(2, 5)	500	0.76 (0.03)	0.77 (0.04)	0.70 (0.03)	0.74 (0.03)	0.74 (0.03)
	1,000	0.52 (0.02)	0.45 (0.02)	0.48 (0.02)	0.45 (0.02)	0.45 (0.02)

#### 4. Application to Chicago Precipitation Data

An important topic in climate studies is quantifying extremal phenomena such as heavy precipitation or high temperature, for which quantile regression serves as a promising tool. We applied the proposed methods to the statistical downscaling of daily precipitations at the Aurora station of Chicago. The response  $Y$  is the observed daily precipitation (inches) at the station from 1957 to 2002, and the covariate  $X$  is the simulated daily precipitation from the ERA-40 reanalysis model introduced in Uppala et al. (2005).

Since we are interested in estimating the extremely heavy precipitation conditioning on  $X$ , we only include the wet days data. In this data set 30% of the days are wet with  $y_i > 0$ . Since in climate studies it is commonly assumed that the percentage of wet days in the future is the same as in the past and prediction is based on the simulated daily precipitation, we define the wet days data as the pairs of  $(y_i, x_i)$  with  $x_i$  exceeding its 70th sample percentile. This yields a data set of 4,816 observations.

As a preliminary analysis, we applied conventional quantile regression at high quantiles  $\tau > 0.99$ . The slope estimates appear to be stable from  $\tau = 0.990$  to  $\tau = 0.995$ ; see Figure 1. We apply a variation of the Wald-type test (Koenker, 2005, page 76) for testing the null hypothesis  $H_0 : \beta(0.990) = \beta(0.991) = \dots = \beta(0.995)$  and obtain a  $p$ -value=0.794. However, the test for  $H_0 : \beta(0.990) = \dots = \beta(0.996) = \beta(0.997) = \beta(0.998)$  yields a  $p$ -value of  $4^{-10}$ , suggesting a violation of the common slope assumption for  $\tau \geq 0.996$ . Therefore, we chose six quantile levels 0.990, 0.991,  $\dots$ , 0.995 to apply the proposed combined estimation methods.

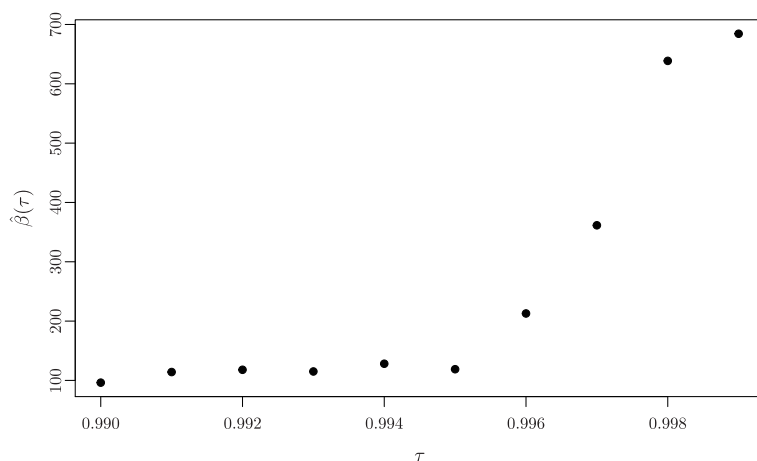


Figure 1. Local quantile regression estimates of the slope at upper quantiles for the precipitation data in Aurora station.

Table 6. The estimated common slope  $\hat{\beta}$  for the proposed methods and the bootstrap standard error (s.e.), and the quantile-specific weights given by different methods based on the estimated extreme value index  $\hat{\xi} = 0.29$ .

Estimate	QAE	CRQ	OWQAE	WCRQ+	OWCRQ
$\hat{\beta}$	115.07	132.63	91.21	96.25	96.78
bootstrap s.e.	(103.08)	(117.21)	(30.72)	(40.93)	(40.20)
$\tau_1 = 0.990$	1/6	1/6	1.22	1	1.55
$\tau_2 = 0.991$	1/6	1/6	-0.03	0	-0.05
$\tau_3 = 0.992$	1/6	1/6	-0.03	0	-0.05
$\tau_4 = 0.993$	1/6	1/6	-0.03	0	-0.06
$\tau_5 = 0.994$	1/6	1/6	-0.03	0	-0.06
$\tau_6 = 0.995$	1/6	1/6	-0.10	0	-0.32

Table 6 summarizes the estimated slopes from different methods. The values in the parentheses are the bootstrap standard errors based on 1,000 bootstrap samples obtained by sampling the paired observations  $(y_i, x_i)$  with replacement. Here  $\hat{\beta}_{QAE}$  and  $\hat{\beta}_{CRQ}$  are larger than the other estimates, which is partially due to the fact that these two methods enforce equal weights at the six quantiles including the higher quantile  $\tau = 0.994$ , at which the quantile slope estimation  $\hat{\beta}(0.994)$  is the largest among the six slope estimations. In addition, the estimators  $\hat{\beta}_{QAE}$  and  $\hat{\beta}_{CRQ}$  have larger variances, which lead to insignificance in the slope. The OWQAE, OWCRQ and WCRQ+ estimators have smaller variances, and all three showed that at the high quantiles, the simulated daily precipitation has a significant positive effect on the observed precipitation and thus can serve as a good predictor for high precipitation.

Table 7. The prediction error of different methods at quantile levels 0.990, 0.992 and 0.995. The values in the parentheses are the standard errors of the prediction errors based on 500 times of cross-validation.

$\tau$	RQ	QAE	CRQ	OWQAE	WCRQ+	OWCRQ
0.990	65.27 (0.34)	66.75 (0.30)	68.74 (0.29)	65.27 (0.33)	65.27 (0.34)	65.25 (0.33)
0.992	57.10 (0.32)	57.92 (0.30)	60.29 (0.29)	56.77 (0.33)	56.79 (0.33)	56.76 (0.33)
0.995	45.01 (0.25)	43.76 (0.26)	46.47 (0.3)	42.31 (0.32)	42.38 (0.32)	42.37 (0.32)

We conducted a cross-validation study to evaluate the prediction accuracy of each method. We randomly choose half of the data set as the training data (total 2,408 observations) and the other half as the validation data. For each method, we used training data to obtain the estimated coefficients  $\hat{\alpha}(\tau)$  and  $\hat{\beta}(\tau)$  and predicted the  $\tau$ th conditional quantile of  $Y$  as  $\hat{Q}_Y(\tau|x_i) = \hat{\alpha}(\tau) + x_i\hat{\beta}(\tau)$ . The prediction error is

$$PE = \sum_{i=1}^{2408} \rho_{\tau}\{y_i - \hat{Q}_Y(\tau|x_i)\}, \tau \in \{0.990, \dots, 0.995\},$$

where  $\{(y_i, x_i), i = 1, \dots, 2408\}$  are in the validation set. The cross validation was repeated 500 times and the mean of PE is reported at  $\tau = 0.990, 0.992$  and  $0.995$  in Table 7. Results show that the optimally weighted methods perform generally better than the equally-weighted methods QAE and CRQ and the local quantile regression method (RQ). Moreover, the gain in the prediction accuracy of the optimally weighted methods increases as the quantile level gets more extreme.

For the precipitation data, it is interesting but challenging to estimate the quantiles in the more extreme tail. Since the slope in this data set does not appear to be constant for  $\tau \in [0.996, 1)$ , the proposed methods cannot be directly used for this. We could consider joint-quantile estimation by incorporating some other type of commonality such as local linearity in the slope; further research is needed in this direction.

### Supplementary Material

Proofs for Theorems 1–3, Proposition 1 and the statements in Remark 2 are provided in the online supplementary material.

### Acknowledgement

The authors would like to thank two referees, an associate editor and the Co-Editor for helpful comments. The research was partially supported by the National Science Foundation CAREER Award DMS-1525692.

## References

- Abrevaya, J. (2001). The effects of demographics and maternal behavior on the distribution of birth outcomes. *Empirical Econom.* **26**, 247-257.
- Bickel, P. J. (1975). One-step Huber estimates in the linear model. *J. Amer. Statist. Assoc.* **70**, 428-434.
- Bradic, J., Fan, J. and Wang, W. (2011). Penalized composite quasi-likelihood for ultrahigh dimensional variable selection. *J. Roy. Statist. Soc. Ser. B* **73**, 325-349.
- Chernozhukov, V. and Du, S. (2008). Extremal quantiles and value-at-risk. In *The New Palgrave Dictionary of Economics* (Edited by S. N. Durlauf and L. E. Blume). Palgrave Macmillan, Basingstoke.
- de Haan, L. and Ferreira, A. (2006). *Extreme Value Theory: An Introduction*. Springer, New York.
- Friederichs, P. (2010). Statistical downscaling of extreme precipitation events using extreme value theory. *Extremes* **13**, 109-132.
- Guo, J., Tian, M. and Zhu, K. (2012). New efficient and robust estimation in varying-coefficient models with heteroscedasticity. *Statist. Sinica* **22**, 1075-1101.
- Hendricks, W. and Koenker, R. (1992). Hierarchical spline models for conditional quantiles and the demand for electricity. *J. Amer. Statist. Assoc.* **87**, 58-68.
- Hogg, R. (1980). Personal communication with Koenker, R.
- Jiang, L., Bondell, H. D. and Wang, H. (2014). Interquantile shrinkage and variable selection in quantile regression. *Comput. Statist. Data Anal.* **69**, 208-219.
- Jiang, L., Wang, H. and Bondell, H. D. (2013). Interquantile shrinkage in regression models. *J. Comput. Graph. Statist.* **22**, 970-986.
- Jiang, R., Qian, W. and Zhou, Z. (2012). Variable selection and coefficient estimation via composite quantile regression with randomly censored data. *Statistics and Probability Letters* **82**, 308-317.
- Jiang, R., Zhou, Z.-G., Qian, W.-M. and Chen, Y. (2013). Two step composite quantile regression for single-index models. *Comput. Statist. Data Anal.* **64**, 180-191.
- Jiang, R., Zhou, Z.-G., Qian, W.-M. and Shao, W.-Q. (2012). Single-index composite quantile regression. *Journal of the Korean Statistical Society* **41**, 323-332.
- Jiang, X., Jiang, J. and Song, X. (2012). Oracle model selection for nonlinear models based on weighted composite quantile regression. *Statist. Sinica* **22**, 1479-1506.
- Kai, B., Li, R. and Zou, H. (2010). Local composite quantile regression smoothing: an efficient and safe alternative to local polynomial regression. *Journal of the Royal Statistical Society B* **72**, 49-69.
- Koenker, R. (1984). A note on l-estimates for linear models. *Statist. Probab. Lett.* **2**, 323-325.
- Koenker, R. (2005). *Quantile Regression*. Cambridge University Press, New York.
- Koenker, R. and Bassett, G. (1978). Regression quantiles. *Econometrica* **46**, 33-50.
- Pandey, G. and Nguyen, V.-T.-V. (1999). A comparative study of regression based methods in regional flood frequency analysis. *J. Hydrology* **225**, 92-101.
- Pickands, J. (1975). Statistical inference using extreme order statistics. *Ann. Statist.* **3**, 119-131.
- Portnoy, S. and Koenker, R. (1989). Adaptive l-estimation for linear models. *Ann. Statist.* **17**, 362-381.



- Powell, J. L. (1991). Estimation of monotonic regression models under quantile restrictions. In *Nonparametric and Semiparametric Methods in Econometrics*, 357-384, Cambridge University Press, New York.
- Smith, R. L. (1987). Estimating tails of probability distributions. *Ann. Statist.* **15**, 1174-1207.
- Tang, L., Zhou, Z. and Wu, C. (2012). Weighted composite quantile estimation and variable selection method for censored regression model. *Statist. Probab. Lett.* **82**, 653-663.
- Uppala, S. M., Källberg, P., Simmons, A., Andrae, U., Bechtold, V., Fiorino, M., Gibson, J., Haseler, J., Hernandez, A., Kelly, G., et al.(2005). The era-40 re-analysis. *Quarterly J. Roy. Meteorological Soc.* **131**, 2961-3012.
- Wang, H., Li, D. and He, X. (2012). Estimation of high conditional quantiles for heavy-tailed distributions. *J. Amer. Statist. Assoc.* **107**, 1453-1464.
- Zhao, Z. and Xiao, Z. (2014). Efficient regressions via optimally combining quantile information. *Econom. Theory* **30**, 1-43.
- Zou, H. and Yuan, M. (2008). Composite quantile regression and the oracle model selection theory. *Ann. Statist.* **36**, 1108-1126.

PPD Inc, 3900 Paramount Parkway, Morrisville, NC 27560-7200, USA.

E-mail: kehui.wang@ppdi.com

Department of Statistics, George Washington University, Raleigh, NC 27695-8203, USA.

E-mail: judywang@gwu.edu

(Received February 2014; accepted February 2015)