

## WEIGHTED QUANTILE REGRESSION THEORY AND ITS APPLICATION

Wei Xiong<sup>\*</sup>, Maozai Tian<sup>2</sup>

<sup>1</sup>*School of Statistics, University of International Business and Economics*

<sup>2</sup>*School of Statistics, Renmin University of China*

### Abstract

As a robust data analysis technique, quantile regression has attracted extensive interest. In this study, the weighted quantile regression (WQR) technique is developed based on sparsity function. We first consider the linear regression model and show that the relative efficiency of WQR compared with least squares (LS) and composite quantile regression (CQR) is greater than 70% regardless of the error distributions. To make the proposed method practically more useful, we consider two nontrivial extensions. The first concerns with a nonparametric model. Local WQR estimate is introduced to explore the nonlinear data structure and shown to be much more efficient compared to other estimates under various non-normal error distributions. The second extension concerns with a multivariate problem where variable selection is needed along with regulation. We couple the WQR with penalization and show that under mild conditions, the penalized WQR enjoys the oracle property. The WQR has an intuitive formulation and can be easily implemented. Simulation is conducted to examine its finite sample performance and compare against alternatives. Analysis of mammal dataset is also conducted. Numerical studies are consistent with the theoretical findings and indicate the usefulness of WQR.

*Key words:* Weighted quantile regression, Local linear regression, Asymptotic relative efficiency, Penalized selection.

---

\* Corresponding author: xiongwei@uibe.edu.cn

## 1 Introduction

In practical regression analysis, it is common that the collected response data display heterogeneity due to either heteroscedastic variance or heavy tails of random errors. As a robust analysis technique, quantile regression (QR; Koenker and Bassett, 1978) is now routinely adopted to accommodate non-normal data. When the relationship between the covariates  $X$  and response  $Y$  evolves across the distribution of  $Y$ , the conditional quantile constitutes a natural tool for depicting the whole distribution. To improve over the standard QR, the composite quantile regression (CQR; Zou and Yuan 2008) has been proposed. It combines strengths across multiple quantile regression models and has been to outperform the standard QR. Despite several nice properties of CQR, its computational cost is high due to the complexity of its loss function.

In this study, we develop the weighted quantile regression (WQR) method to further improve over QR and CQR. Like CQR, WQR combines strengths across multiple quantile regressions efficiently by using data-dependent weights at different quantiles. The weights are obtained based on the sparsity function and have practical meanings. The resulted WQR estimate can be shown to be more robust, reasonable and efficient. Compared to CQR, WQR is computationally more affordable.

In methodological development, we first investigate the simple linear regression model. To fully describe the theoretical efficiency of WQR estimate, we study the relative efficiency (RE) of WQR with respect to both LS and CQR. It is shown that the WQR has satisfactory efficiency even under the worst case scenario. To make WQR practically more useful, we develop two nontrivial extensions of the simple WQR. First, we consider the scenario where the linear assumption is not rich enough to describe the underlying relationship between the response and covariates. Here we consider a nonparametric regression model, develop the local WQR method, and investigate its theoretical properties especially including relative efficiency. Second, we consider the scenario with multiple covariates where variable selection is needed along with estimation and the penalized WQR selection method is adopted. We show that under mild regular conditions, the estimate has the much desired oracle selection and estimation consistency property. Thus, WQR can provide a useful alternative to QR, CQR, and other robust methods.

The rest of the article is organized as follows. Methodological development is presented in Section 2. We first introduce the WQR method under the simple linear regression and study its properties. The nonparametric model and local WQR estimation and the multi-variate model and penalized WQR estimation are then developed. In Section 3, we

conduct simulation study under different models to investigate the finite sample performance. Data analysis example is presented in Section 4. The article concludes with discussions in Section 5. Some technical details and additional numerical study results are presented in Appendix.

## 2 Weighted Quantile Regression Technique and Properties

First consider the linear regression model

$$Y = X\beta + \varepsilon, \quad (1)$$

where  $Y = (Y_1, \dots, Y_n)^T$  is the length- $n$  vector of responses,  $X = (x_1, \dots, x_n)$  is the  $n \times p$  covariate design matrix,  $\beta = (\beta_1, \dots, \beta_p)^T$  is the length- $p$  vector of unknown coefficients, and  $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n)^T$  is vector of random errors.

The standard QR estimate is defined as

$$(\hat{c}_\tau, \hat{\beta}_{QR}) = \operatorname{argmin}_{c, \beta} \sum_{i=1}^n \rho_\tau(Y_i - x_i^T \beta - c), \quad (2)$$

where  $\rho_\tau(u) = \tau u 1(u \geq 0) + (1 - \tau) 1(u < 0)$  is the check function and  $\tau$  is any given quantile.  $c_\tau$  is the  $100\tau\%$  quantile of  $\varepsilon_i$ . When  $\varepsilon_i$ 's are iid, under mild conditions,  $\hat{\beta}_{QR}$  is asymptotically normally distributed,

$$\sqrt{n}(\hat{\beta}_{QR} - \beta) \xrightarrow{D} N\left(0, \frac{\tau(1-\tau)}{f^2(F^{-1}(\tau))} D^{-1}\right). \quad (3)$$

$F(\cdot)$  and  $f(\cdot)$  denote the distribution and density functions of  $\varepsilon$  respectively,  $D$  is a positive definite matrix such that  $\lim_{n \rightarrow \infty} n^{-1} X^T X = D$ , and  $\xrightarrow{D}$  denotes convergence in distribution.

As can be seen from (3), the quantity  $\sqrt{\tau(1-\tau)}/f(F^{-1}(\tau))$  plays a role analogous to the standard deviation in the LS estimation with iid normal errors. Thus to combine strengths across multiple quantile regressions (which shares a similar spirit as CQR) and to estimate  $\beta$  more efficiently, we can use data-dependent weights to make more effective use of the sample information. Consider a set of  $K$  quantiles  $\{\tau_k, k = 1, \dots, K\}$ . The WQR weights are defined as

$$w_k = 1 / \left[ \sqrt{\tau_k(1-\tau_k)} / f(F^{-1}(\tau_k)) \right] = \frac{f(F^{-1}(\tau_k))}{\sqrt{\tau_k(1-\tau_k)}}, \quad k = 1, \dots, K. \quad (4)$$

The quantity  $1/f(F^{-1}(\tau))$  has been referred to as the sparsity function (Tukey, 1965) and quantile-density function (Parzen, 1979), reflecting the density of observations at the quantile of interest. In this paper, we term this quantity *sparsity function* and denote it as  $s(\tau)$ . Then the weight  $w_k$  can be rewritten as  $w_k = [s(k)\sqrt{\tau_k(1-\tau_k)}]$ . Intuitively, by adding the weights  $\{w_k\}_1^K$  into estimation, sample information can be used more effectively, making the weighted estimator more reliable and efficient. Normalized weights are defined as  $w_k^* = w_k/\sum_k w_k$ ,  $k = 1, \dots, K$ .

With a sequence of  $K$  quantiles  $\{\tau_k, k = 1, \dots, K\}$ ,  $\hat{\beta}_{\tau_k}$  is the QR estimate defined in (2) with  $\tau = \tau_k$ . We propose the WQR estimate (5)

$$\hat{\beta}_{WQR} = \sum_{k=1}^K w_k^* \hat{\beta}_{\tau_k}. \quad (5)$$

The asymptotic property of  $\hat{\beta}_{WQR}$  can be summarized as follows.

**Theorem 1** Assume that conditions C1-C3 (Appendix) hold. Then

$$\sqrt{n}(\hat{\beta}_{WQR} - \beta) \xrightarrow{D} N(0, \Sigma_{WQR}),$$

$$\text{where } \Sigma_{WQR} = D^{-1} \Sigma_i \Sigma_j \frac{\tau_i \wedge \tau_j - \tau_i \tau_j}{\sqrt{\tau_i(1-\tau_i)} \sqrt{\tau_j(1-\tau_j)}} / \left( \Sigma_i \frac{f(F^{-1}(\tau_i))}{\sqrt{\tau_i(1-\tau_i)}} \right)^2.$$

## 2.1 Asymptotic relative efficiency

We study the asymptotic relative efficiency (ARE) of WQR with respect to LS estimator (ARE1) and CQR estimator (ARE2) by comparing their MSEs. (1). ARE of WQR with respect to LS

(1) ARE of WQR with respect to LS

As both the WQR and LS estimates are asymptotic unbiased, we only need to compare their asymptotic variances. When  $var(\epsilon) = \sigma^2 < \infty$ , the asymptotic variance of the LS estimate is  $\sigma^2 D^{-1}$ . Thus,

$$\text{ARE}_1(\tau_1, \dots, \tau_K, f) = \frac{\text{MSE}(\hat{\beta}_{LS})}{\text{MSE}(\hat{\beta}_{WQR})} = \frac{\sigma^2 \left( \Sigma_i \frac{f(F^{-1}(\tau_i))}{\sqrt{\tau_i(1-\tau_i)}} \right)^2}{\Sigma_i \Sigma_j \frac{\tau_i \wedge \tau_j - \tau_i \tau_j}{\sqrt{\tau_i(1-\tau_i)} \sqrt{\tau_j(1-\tau_j)}}}. \quad (6)$$

For the convenience of implementation, we take equally spaced quantiles,  $\tau_k = \frac{k}{K+1}$ ,  $k = 1, \dots, K$  and denote the corresponding ARE as  $ARE_1(K, f)$ . As  $K \rightarrow \infty$ , the  $ARE_1(K, f)$  converges to a limit, denoted by  $\theta_1(f)$ . The next theorem establishes the lower bound of the  $ARE_1$ .

**Theorem 2** *The universal lower bound of  $ARE_1$  is that*

$$\lim_{K \rightarrow \infty} \frac{\sum_i \sum_j \frac{\tau_i \wedge \tau_j - \tau_i \tau_j}{\sqrt{\tau_i(1-\tau_i)} \sqrt{\tau_j(1-\tau_j)}}}{\left( \sum_i \frac{f(F^{-1}(\tau_i))}{\sqrt{\tau_i(1-\tau_i)}} \right)^2} \leq \frac{\pi^2 - 8}{16(E[f(\varepsilon)])^2},$$

And

$$\theta_1(f) = \lim_{K \rightarrow \infty} ARE_1(K, f) \geq 16\sigma^2 / (\pi^2 - 8)(E[f(\varepsilon)])^2.$$

To get more insights into  $ARE_1$ , we provide in Table 1  $ARE_1(K, f)$  for some commonly assumed error distributions. Several observations can be made. First, when the error distribution is  $N(0,1)$ , the LS estimator has the best performance as expected. But when  $K$  is large,  $ARE_1$  is very close to 1. Second, for all the non-normal distributions listed in Table 1, WQR has a higher efficiency than LS, particularly when  $K$  is small. For the mixture of two normal distributions that is often assumed for modeling contaminated and heavy-tailed data,  $ARE_1$  is as large as 7.4. For the skewed  $x_2$  distribution,  $ARE_1$  is also larger than 1 regardless of the choice of  $K$ . Third, the value of  $K$  affects  $ARE_1$ . We plot  $ARE_1$  as a function of  $K$  in Figure 1.

Table 1:  $ARE_1$  for commonly assumed error distributions.

Error Distribution	$ARE_1(K, f)$				
	$K = 3$	$K = 5$	$K = 9$	$K = 19$	$K = 99$
$N(0,1)$	0.8589	0.9150	0.9516	0.9731	0.9849
Laplace	1.5538	1.4781	1.4337	1.4090	1.3957
T distribution with df=3	1.8580	1.8631	1.8441	1.8193	1.7974
T distribution with df=4	1.2387	1.2421	1.2294	1.2128	1.1983
$x_3^2$	1.3713	1.5905	1.8025	2.0030	2.2274
$0.95N(0,1) + 0.05N(0,3^2)$	1.0269	1.1637	1.1959	1.2063	1.1986
$0.90N(0,1) + 0.10N(0,3^2)$	1.2169	1.3528	1.3727	1.3656	1.3455
$0.95N(0,1) + 0.05N(0,10^2)$	4.2278	4.7208	4.7945	4.7579	4.6404
$0.90N(0,1) + 0.10N(0,10^2)$	6.8877	7.3963	7.2898	6.9536	6.8037

## (2) ARE of WQR with respect to CQR

For WQR and CQR, the assumption of  $\sigma^2 < \infty$  can be relaxed. For any error distribution, the relative efficiency between WQR and CQR, denoted by  $ARE_2$ , is

$$ARE_2(\tau_1, \dots, \tau_K, f) = \frac{MSE(\hat{\beta}_{CQR})}{MSE(\hat{\beta}_{WQR})} = \frac{\left(\sum_i f(F^{-1}(\tau_i))\right)^2 \left(\sum_i \frac{f(F^{-1}(\tau_i))}{\sqrt{\tau_i(1-\tau_i)}}\right)^2}{\left(\sum_{i,j} \tau_i \wedge \tau_j - \tau_i \tau_j\right) \left(\sum_{i,j} \frac{\tau_i \wedge \tau_j - \tau_i \tau_j}{\sqrt{\tau_i(1-\tau_i)} \sqrt{\tau_j(1-\tau_j)}}\right)}.$$

Still take  $\tau_k = k/(K+1)$ ,  $k = 1, \dots, K$ . Denote the limit of  $ARE_2(K, f)$  as  $\theta_2$  when  $K \rightarrow \infty$ . By the conclusion in Zou and Yuan (2008) that the asymptotic variance of the CQR estimate is  $\Sigma_{CQR} \rightarrow 1 \{12[E(f(\varepsilon))]^2\}$ . Then we have

**Theorem 3** Denote  $F$  as the collection of all density functions.

$$\theta_2 = \inf_{f \in \mathcal{F}} \lim_{K \rightarrow \infty} ARE_2(\tau_1, \dots, \tau_K, f) \geq \frac{4}{3(\pi^2 - 8)} \approx 0.7132$$

That is, regardless of the error distribution, the  $ARE_2$  is bounded below by 0.7132. It indicates that WQR does not incur a serious loss in efficiency even under the worst case scenario. To gain more insights, in Figure 1, we show  $ARE_2$ , for four error distributions. In this figure, we fix  $K = 19$  for CQR, as suggested by Zou and Yuan (2008) and vary the  $K$  value for WQR. Results are consistent with Theorem 3.

## 2.2 Local WQR for the nonlinear regression model

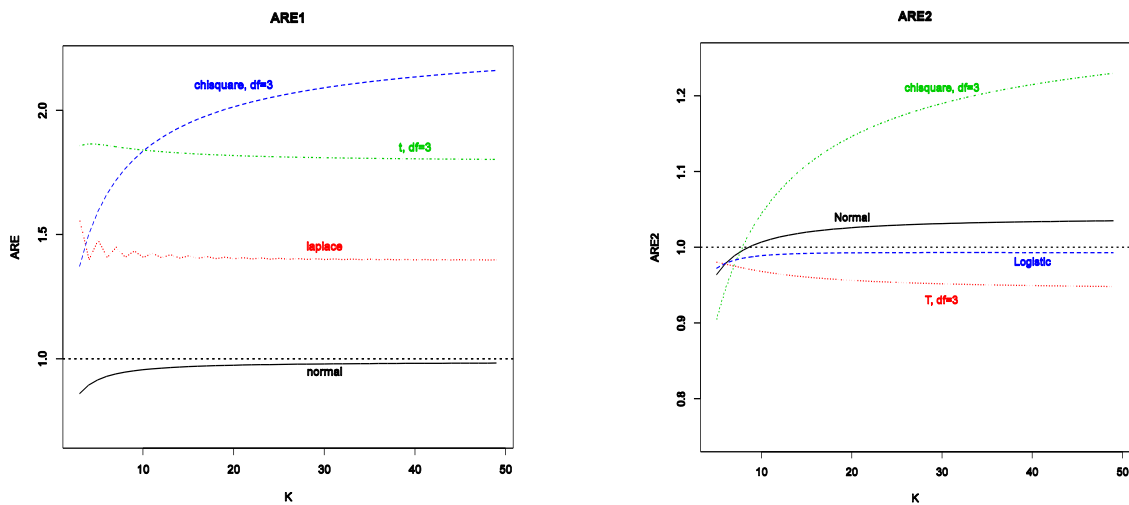


Figure 1:  $ARE_1$  and  $ARE_2$  for different error distributions.

In this section, we further develop the WQR approach for the nonlinear regression model. For simplicity of notation, assume that the covariate is onedimensional. Extension to the additive model for multiple covariates is simple. Suppose that we have a sample  $\{(x_i, y_i), i = 1, \dots, n\}$  satisfying the model

$$y_i = m(x_i) + \varepsilon_i . \quad (8)$$

$m = (x_i) = E(y_i|x_i)$  is a smooth function, error  $\varepsilon_i$  can be heteroscedastic or have infinite variance. Consider estimating  $m(\cdot)$  at a fixed covariate value  $x_0$ . We approximate  $m(x_i)$  locally by a linear function  $m(x_i) \approx m(x_0) + m'(x_0)(x_i - x_0)$ , and then fit a linear model in the neighborhood of  $x_0$ . Let  $K(\cdot)$  be a smooth kernel function. The local linear quantile regression estimate at the  $\tau$ -th quantile can be obtained from

$$(\hat{a}, \hat{b}) = \operatorname{argmin}_{a,b} \sum_{i=1}^n p_\tau(y_i - a - b(x_i - x_0)) K\left(\frac{x_i - x_0}{h}\right), \quad (9)$$

where  $h$  is the smoothing parameter.  $\hat{m}_\tau = \hat{a} + \hat{b}(x_i - x_0)$ . Yu and Jones (1998) show that Bias  $(\hat{m}_\tau(x_0)) = \frac{1}{2} h^2 \mu_2(K) m''(x_0) + o(h^2)$ ,  $\operatorname{Var}(\hat{m}_\tau(x_0)) = \frac{\tau(1-\tau)R(K)}{nhp(x_0)f^2(F^{-1}(\tau))} + o(1/nh)$ .

$\mu_2(K) = \int u^2 K(u) du$ ,  $R(K) = \int K^2(u) du$ , and  $p(\cdot)$  denotes the design density. Following a similar rationale as in the last section, we propose the local-WQR estimate as

$$\hat{m}_{WQR}(x_0) = \sum_{k=1}^K \hat{w}_k^* \hat{m}_{\tau k}(x_0) \quad (10)$$

The asymptotic properties of  $\hat{m}_{WQR}(x_0)$  can be summarized as follows.

**Theorem 4** Assume the same conditions as in Fan et al. (1994). As  $n \rightarrow \infty$ ,  $h = h(n) \rightarrow 0$  and  $nh \rightarrow \infty$ , then

$$\sqrt{nh}(\hat{m}_{WQR}(x_0) - m(x_0)) \xrightarrow{D} N\left(\frac{1}{2} h^2 \mu_2(K) m''(x_0), \Sigma'_{WQR}\right),$$

Where  $\Sigma'_{WQR} = \Sigma_{i,j} \frac{\tau_i \wedge \tau_j - \tau_i \tau_j}{\sqrt{\tau_i(1-\tau_i)} \sqrt{\tau_j(1-\tau_j)}} / \left( \Sigma_i \frac{f(F^{-1}(\tau_i))}{\sqrt{\tau_i(1-\tau_i)}} \right)^2 nhp(x_0)$ .

Remark: For the above result to hold,  $x_0$  cannot be “too close” to the boundaries, and smoothness conditions on  $m(\cdot)$  are needed. The proof can be obtained by combining results

in Yu and Jones (1998) and Theorem 1, is omitted here.

### 2.3 Penalized WQR for the multivariate model

Still consider model (1). When  $p$  is moderate to large, variable selection can be needed along with estimation. Here we use a penalty function for variable selection. Let  $p_\lambda(\cdot)$  be the penalty function with tuning parameter  $\lambda$ . For a sequence of quantiles  $\{\tau_k: k = 1, \dots, K\}$ , the penalized WQR (PWQR) estimate is defined as

$$\hat{\beta}_{\tau_k} = \operatorname{argmin}_\beta \sum_{i=1}^n \rho_{\tau_k}(Y_i - x_i^T \beta) + \sum_{j=1}^p p_{\lambda_k}(|\beta_j|), k = 1, \dots, K,$$

$$\hat{\beta}_{PWQR} = \sum_{k=1}^K w_k^* \hat{\beta}_{\tau_k}, \quad (11)$$

where  $\lambda_k$ 's are the tuning parameters for each quantile regression models. The weights  $w_k^*$ 's have the same definition as in the previous sections.

A large number of penalty functions have been developed, including the Lasso family, SCAD, MCP, and others. Many (if not all) of them are potentially applicable here. As a demonstration, we adopt the adaptive Lasso penalty for its simplicity. In this study, we focus on the relatively simpler case with a fixed  $p$  and  $n \rightarrow \infty$ .

For a sequence of quantiles  $\{\tau_k = k/(K+1), k = 1, \dots, K\}$ , consider the following PWQR procedure. First for each  $\tau_k$ , fit the regular QR using all predictors and denote the estimate as  $\hat{\beta}_{\tau_k}^{(0)}$ . Note that under mild conditions, this estimate is  $\sqrt{n}$  estimation consistent. Compute the adaptive Lasso estimate

$$\hat{\beta}_{\tau_k}^P = \operatorname{argmin}_\beta \sum_{i=1}^n \rho_{\tau_k}(Y_i - x_i^T \beta) + \lambda_k \sum_{j=1}^p \frac{|\beta_j|}{|\hat{\beta}_{\tau_k,j}^{(0)}|^2}.$$

The final PWQR estimate is

$$\hat{\beta}_{PWQR} = \sum_{k=1}^K w_k^* \hat{\beta}_{\tau_k}^P. \quad (12)$$

Let  $\beta_0 = (\beta_{10}^T, \beta_{20}^T)^T$  denote the true value of  $\beta$ , where  $\beta_{10}$  is a  $s$ -vector. Without loss of generality, assume that  $\beta_{20} = 0$  and  $\beta_{10}$  contains all the nonzero components of  $\beta_0$ . Denote  $\hat{\beta}_{1,PWQR}$  and  $\hat{\beta}_{2,PWQR}$  as the components of  $\hat{\beta}_{PWQR}$  corresponding to  $\beta_{10}$  and  $\beta_{20}$ . Properties of the PWQR estimate are summarized as follows.



**Theorem 5** Assume that the regularity conditions in Theorem 3 hold. If  $\lambda_k/\sqrt{n} \rightarrow 0$  and  $\lambda_k \rightarrow \infty, k = 1, \dots, K$ , then  $\hat{\beta}_{PWQR}$  satisfies

- (a) Sparsity, that is,  $\hat{\beta}_{2,PWQR} = 0$ , with probability tending to one;  
 (b) Asymptotic normality, that is,

$$\sqrt{n}(\hat{\beta}_{1,PWQR} - \hat{\beta}_{10}) \xrightarrow{D} (0, \Sigma_{PWQR1}),$$

where

$$\Sigma_{PWQR1} = D_{11}^{-1} \Sigma_i \Sigma_j \frac{\tau_i \wedge \tau_j - \tau_i \tau_j}{\sqrt{\tau_i(1-\tau_i)} \sqrt{\tau_j(1-\tau_j)}} / \left( \Sigma_i \frac{f(F^{-1}(\tau_i))}{\sqrt{\tau_i(1-\tau_i)}} \right)^2,$$

and  $D_{11}$  is the submatrix of  $D$  with the first  $s$  elements in both columns and rows.

That is, the PWQR estimate has the oracle consistency properties. Theorem 5 establishes the asymptotic rate of  $\lambda_k$ 's. Following Wang (2007), in data analysis, we use a BIC-type criterion to select  $\lambda_k$ 's. Practically the optimal  $\lambda$  is selected for each  $\tau_k$  separately. More specifically, the criterion is defined as

$$\text{BIC}(\lambda_k, \tau_k) = \log \left( \frac{1}{n} \sum_{i=1}^n \rho_{\tau_k} \left( Y_i - x_i^T \hat{\beta}_{\tau_k}^P(\lambda_k) \right) \right) + \frac{\log(n)}{n} df_{\lambda_k},$$

where  $df_{\lambda_k}$  is the number of nonzero coefficients in the fitted model.

### 3 Simulation Study

We use Monte Carlo simulation studies to examine the finite sample performance of WQR estimator and compare against LS and CQR estimators. We consider two examples here, Local-WQR and Penalized-WQR. For all examples, the weights are estimated using the plug-in method and 400 data-sets are generated, each consisting of  $n = 200$  observations.

#### (1) Example 1 : Nonparametric regression model

We examine WQR in nonparametric scenario. We consider the same five error distributions as in the previous section. In estimation, the bandwidth is selected by a plug-in bandwidth selector (Ruppert et.al, 1995), and we employ the Gaussian Kernel. The performance of  $\hat{m}(\cdot)$  is assessed by the average squared errors ASE, defined by

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

where  $\{u_i, i = 1, \dots, n_{grid}\}$  are the grid points evenly distributed over the interval at which  $m(\cdot)$  is estimated. We set  $n_{grid} = 200$  and summarize our simulation results using the ratios of average squared errors  $RASE(\hat{m}) = \text{ASE}(\hat{m}_{LS})/\text{ASE}(\hat{m}_{WQR})$  and  $RASE(\hat{m}) = \text{ASE}(\hat{m}_{LS})/\text{ASE}(\hat{m}_{CQR})$ .

Consider the model

$$Y = \sin(2X) + 2\exp(-16X^2) + 0.5\varepsilon ,$$

where  $X$  follows  $U(-2,2)$ . To estimate the mean regression function  $m(x) = \sin(2x) + 2\exp(-16x^2)$ , local linear WQR estimator, local CQR estimator (Kai and Li, 2010) and local linear LS method are employed, and we estimate  $m(\cdot)$  over  $[-1.5, 1.5]$ . For the local linear WQR estimator, we consider  $K = 5, 9, 19$ , and for the local CQR estimator, we set  $K = 9$  following Kai and Li (2010). To see how the estimates behave at a typical point, we present the biases and standard deviations of  $\hat{m}(x_0)$  at  $x_0 = 0.75$  in Table 2. The mean and standard deviation of RASE over 400 simulations are also summarized.

Table 2 presents some information. When the error follows normal distribution, the RASEs of the WQR estimator are slightly less than 1, especially for heteroscedastic errors. For non-normal distributions, the RASEs of the WQR estimator are greater than 1, indicating a gain in efficiency. For estimating the regression function,  $WQR_5$  and  $WQR_9$  seem to have better overall performance than  $WQR_{19}$ . For Cauchy errors, the LS estimator totally breaks down, while the WQR estimator remains reliable, and the RASEs can be as very large. We plot the WQR, CQR, and LS estimates for mean regression function  $m(x)$  under Cauchy errors in Figure 2. We consider  $CQR_9$  and  $WQR_5, WQR_9, WQR_{19}$  for WQR. It is clear that while the LS estimator breaks down, both the CQR and WQR estimators remain well behaved.

## (2) Example 2: Multivariate regression model

Consider the linear model

$$Y = X^T\beta + \sigma\varepsilon ,$$

Table 2: Result of Simulation example 1.

Distribution	RASE and $\hat{m}$ at $x_0 = 0.75$ , Bias (standard deviation)				
	LS	CQR <sub>9</sub>	WQR <sub>5</sub>	WQR <sub>9</sub>	WQR <sub>19</sub>
$N(0,1)$	-0.023 (0.122)	-0.020 (0.125)	-0.021(0.129)	-0.020(0.127)	-0.021 (0.125)
RASE	—	1.012 (0.085)	0.980(0.127)	1.015(0.096)	1.019 (0.064)
$t_3$	-0.037 (0.158)	-0.037 (0.136)	-0.037(0.130)	-0.037(0.134)	-0.036 (0.138)
RASE	—	1.237 (0.499)	1.369(0.531)	1.329(0.747)	1.197 (0.244)
Laplace	-0.025 (0.156)	-0.024 (0.145)	-0.025(0.142)	-0.025(0.142)	-0.024 (0.147)
RASE	—	1.095 (0.1394)	1.165(0.200)	1.121(0.139)	1.107 (0.093)
$x_3^2$	1.454 (0.227)	1.356 (0.226)	1.303(0.232)	1.340(0.231)	1.368 (0.227)
RASE	—	1.135 (0.039)	1.219(0.058)	1.163(0.044)	1.120 (0.031)
Cauchy	-0.654 (18.14)	-0.143 (1.834)	-0.062 (0.248)	-0.304 (5.106)	-0.251 (4.521)
RASE	—	10802 ( $\infty$ )	1462.3 ( $\infty$ )	12200 ( $\infty$ )	4730.2 ( $\infty$ )

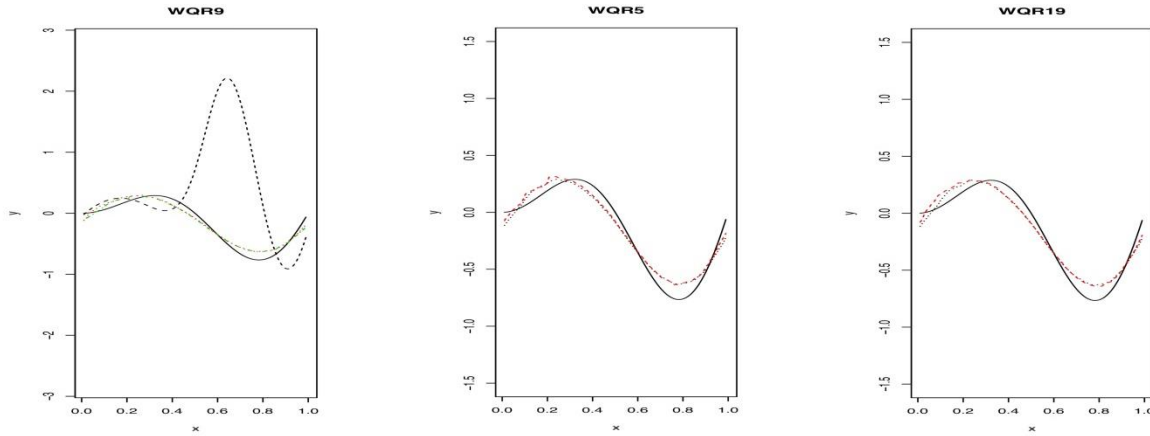


Figure 2: Estimated regression function under Cauchy distribution. First panel: CQR9 (green dashed), WQR9 (red dashed) and LS (dotted) estimates and the true parameter (solid). Last two panels: WQR (red dashed) and CQR (dotted).

where  $\beta = (3, 1.5, 0, 0, 2, 0, 0, 0)$ . Predictors  $X$  follows a multivariate normal distribution with correlation matrix  $\Sigma_{ij} = 0.5^{|i-j|}$ , for  $1 \leq i, j \leq 8$ . Six error distributions are considered:  $N(0, 1)$ ,  $N(0, 3)$ ,  $t$  distribution with  $df = 3$ , Laplace, Cauchy, and mixture of normals  $0.9N(0, 1) + 0.1N(0, 10^2)$ . To obtain the PWQR estimator, tuning parameter  $\lambda_k$  is selected via BIC-like criterion.

*Performance of  $\hat{\beta}_{PWQR}$ ,  $\hat{\beta}_{CQR}$  and  $\hat{\beta}_{LS}$ .* To examine the performance of the proposed  $\hat{\beta}_{PWQR}$ , we consider  $K = 5, 9$  and 19 and take  $\hat{\beta}_{LS}$  and  $\hat{\beta}_{CQR_9}$  for comparison. Means and standard deviations of estimates for nonzero coefficients  $\beta_1$ ,  $\beta_2$  and  $\beta_5$  over 400 simulations are summarized in Table 3.

*Performance of variable selection.* We use generalized mean square error (model error)  $GMSE(\hat{\beta}) = (\hat{\beta} - \beta)^T E(XX^T)(\hat{\beta} - \beta)$  to assess the performance of variable selection procedures. We use notations ‘‘C’’ and ‘‘IC’’ to measure the model complexity. Specifically, column C shows the average number of zero coefficients correctly estimated to be zero, and IC denotes the average number of nonzero coefficients incorrectly estimated to be zero.

Further, we use ‘‘U-fit’’ (under-fit) to denote the trial which excludes nonzero coefficients. ‘‘C-fit’’ (correct-fit) represents the correct model, and ‘‘O-fit’’ (over-fit) is the trial which selects irrelevant predictors besides the three significant predictors. Results over 400 simulations are summarized in Table 4. For each column, we report the mean and standard deviation.

Table 3: Simulation example 2: summary of estimation.

Method	Mean(Standard Deviation)		
	$\hat{\beta}_1 (= 3.0)$	$\hat{\beta}_2 (= 1.5)$	$\hat{\beta}_5 (= 2.0)$
<i>Standard Normal</i>			
LS	3.004 (0.079)	1.485 (0.084)	1.994 (0.076)
CQR <sub>9</sub>	3.004 (0.099)	1.480 (0.136)	1.991 (0.089)
PWQR <sub>5</sub>	3.015 (0.089)	1.465 (0.091)	1.988 (0.076)
PWQR <sub>9</sub>	3.018 (0.089)	1.465 (0.089)	1.985 (0.075)
PWQR <sub>19</sub>	3.018 (0.089)	1.458 (0.089)	1.984 (0.074)
<i>N (0, 3)</i>			
LS	3.006 (0.148)	1.465 (0.155)	1.980 (0.131)
CQR <sub>9</sub>	3.009 (0.170)	1.463 (0.190)	2.001 (0.145)
PWQR <sub>5</sub>	3.017 (0.154)	1.423 (0.175)	1.973 (0.133)
PWQR <sub>9</sub>	3.019 (0.152)	1.414 (0.175)	1.970 (0.169)
PWQR <sub>19</sub>	3.023 (0.152)	1.413 (0.174)	1.970 (0.129)
<i>t – distribution with df = 3</i>			
LS	2.998(0.134)	1.479 (0.146)	1.981 (0.123)
CQR <sub>9</sub>	3.017 (0.127)	1.471 (0.152)	1.999 (0.110)
PWQR <sub>5</sub>	3.012 (0.120)	1.456 (0.128)	1.990 (0.102)
PWQR <sub>9</sub>	3.014 (0.120)	1.455 (0.127)	1.990 (0.102)
PWQR <sub>19</sub>	3.013 (0.119)	1.455 (0.127)	1.990 (0.102)
<i>Laplace</i>			
LS	2.999 (0.120)	1.483 (0.122)	1.987 (0.109)
CQR <sub>9</sub>	3.022 (0.129)	1.465 (0.154)	2.009 (0.101)
PWQR <sub>5</sub>	3.017 (0.108)	1.462 (0.114)	1.980 (0.096)
PWQR <sub>9</sub>	3.017 (0.108)	1.462 (0.114)	1.980 (0.096)
PWQR <sub>19</sub>	3.017 (0.108)	1.462 (0.114)	1.980 (0.096)
<i>0.9N (0, 1) + 0. N (0, 10<sup>2</sup>)</i>			
LS	2.999 (0.270)	1.425 (0.291)	1.934 (0.257)
CQR <sub>9</sub>	3.016 (0.134)	1.475 (0.170)	1.989 (0.113)
PWQR <sub>5</sub>	3.013 (0.122)	1.459 (0.126)	1.990 (0.104)
PWQR <sub>9</sub>	3.013 (0.122)	1.460 (0.126)	1.990 (0.104)
PWQR <sub>19</sub>	3.013 (0.122)	1.460 (0.126)	1.990 (0.104)
<i>Standard Cauchy</i>			
LS	1.743 (1.753)	0.708 (1.409)	1.110 (3.510)
CQR <sub>9</sub>	3.012 (0.185)	1.490 (0.211)	1.988 (0.174)
PWQR <sub>5</sub>	3.021 (0.162)	1.440 (0.188)	1.983 (0.133)
PWQR <sub>9</sub>	3.020 (0.161)	1.441 (0.183)	1.983 (0.133)
PWQR <sub>19</sub>	3.021 (0.162)	1.440 (0.188)	1.983 (0.132)

Table 4: Simulation example 2: performance of variable selection

Method	GMSE		No. of zeros		Proportion of fits		
	Mean (SD)	Median (MAD)	C	IC	U-fit	C-fit	O-fit
<i>Standard Normal</i>							
LS	0.017 (0.015)	0.013 (0.011)	4.908	0.000	0.000	0.918	0.082
CQR <sub>9</sub>	0.028 (0.133)	0.014 (0.012)	4.970	0.000	0.000	0.975	0.025
PWQR5	0.019 (0.016)	0.014 (0.011)	5.000	0.000	0.000	1.000	0.000
PWQR9	0.019 (0.015)	0.015 (0.012)	5.000	0.000	0.000	1.000	0.000
PWQR19	0.019 (0.015)	0.014 (0.012)	5.000	0.000	0.000	1.000	0.000
<i>N(0, 3)</i>							
LS	0.056 (0.046)	0.043 (0.035)	4.903	0.000	0.000	0.910	0.090
CQR <sub>9</sub>	0.072 (0.087)	0.047 (0.041)	4.543	0.000	0.000	0.818	0.182
PWQR5	0.061 (0.053)	0.048 (0.039)	4.993	0.000	0.000	0.993	0.007
PWQR9	0.062 (0.058)	0.047 (0.039)	4.990	0.000	0.000	0.990	0.010
PWQR19	0.080 (0.064)	0.061 (0.057)	4.988	0.000	0.000	0.988	0.012
<i>t - distribution with df = 3</i>							
LS	0.048 (0.047)	0.035 (0.033)	4.893	0.000	0.000	0.903	0.097
CQR <sub>9</sub>	0.060 (0.269)	0.023 (0.022)	4.945	0.000	0.000	0.970	0.003
PWQR5	0.034 (0.027)	0.026 (0.021)	5.000	0.000	0.000	1.000	0.000
PWQR9	0.033 (0.027)	0.026 (0.021)	5.000	0.000	0.000	1.000	0.000
PWQR19	0.033 (0.027)	0.027 (0.021)	5.000	0.000	0.000	1.000	0.000
<i>Laplace</i>							
LS	0.036 (0.031)	0.028 (0.025)	4.883	0.000	0.000	0.895	0.105
CQR <sub>9</sub>	0.051 (0.176)	0.019 (0.016)	4.963	0.000	0.000	0.983	0.017
PWQR5	0.028 (0.026)	0.021 (0.019)	5.000	0.000	0.000	1.000	0.000
PWQR9	0.028 (0.026)	0.021 (0.019)	5.000	0.000	0.000	1.000	0.000
PWQR19	0.028 (0.026)	0.021 (0.019)	5.000	0.000	0.000	1.000	0.000
<i>0.9N(0, 1) + 0.1N(0, 10<sup>2</sup>)</i>							
LS	0.223 (0.206)	0.166 (0.154)	4.873	0.003	0.003	0.877	0.120
CQR <sub>9</sub>	0.034 (0.066)	0.019 (0.018)	4.953	0.000	0.000	0.983	0.017
PWQR5	0.033 (0.029)	0.025 (0.022)	5.000	0.000	0.000	1.000	0.000
PWQR9	0.033 (0.029)	0.026 (0.021)	5.000	0.000	0.000	1.000	0.000
PWQR19	0.033 (0.029)	0.026 (0.021)	5.000	0.000	0.000	1.000	0.000
<i>Cauchy</i>							
LS	25.18 (155.2)	8.255 (10.30)	4.828	1.870	0.840	0.007	0.153
CQR <sub>9</sub>	0.094 (0.177)	0.048 (0.045)	4.383	0.000	0.000	0.777	0.223
PWQR5	0.060 (0.095)	0.040 (0.034)	5.000	0.000	0.000	1.000	0.000
PWQR9	0.059 (0.088)	0.040 (0.034)	5.000	0.000	0.000	1.000	0.000
PWQR19	0.059 (0.089)	0.042 (0.035)	5.000	0.000	0.000	1.000	0.000

## 4 Real Data Analysis

As an illustration, we apply the proposed WQR methods to analyze the mammals data set. First we employ local WQR estimator to a mammals data set, consisting of 107 samples. Of interest is the relationship between the running speed of mammal species and their body mass. The data was collected by Garland in 1983. We applied the local WQR, CQR estimator and local LS estimator to fit the data, we take the response  $Y$  to be the logarithm of speed, and the predictor  $X$  is the logarithm of weight. First we depict the scatter plot and estimate the regression function using full data, depicted in Figure 3. We present the local WQR estimates and CQR estimates with  $K = 9$ , actually the estimates are very similar with different  $K$ s. It is interesting to see from Figure 3 that the overall pattern of these three estimates are the same. The difference between these three estimates becomes slightly larger only when  $x$  is around 2. Interesting enough is that several possible outliers can be detected from the scatter plot when  $x$  is about 2. To analyze the influence of the outliers, we re-estimated the regression function after excluding these outlier observations. Results are depicted in Figure 4, to make comparisons we also present the estimates of the full observations in each panel of Figure 4. We can see that the local WQR estimate remains almost the same, the local CQR estimate changes a little, whereas the LS estimate changes dramatically.

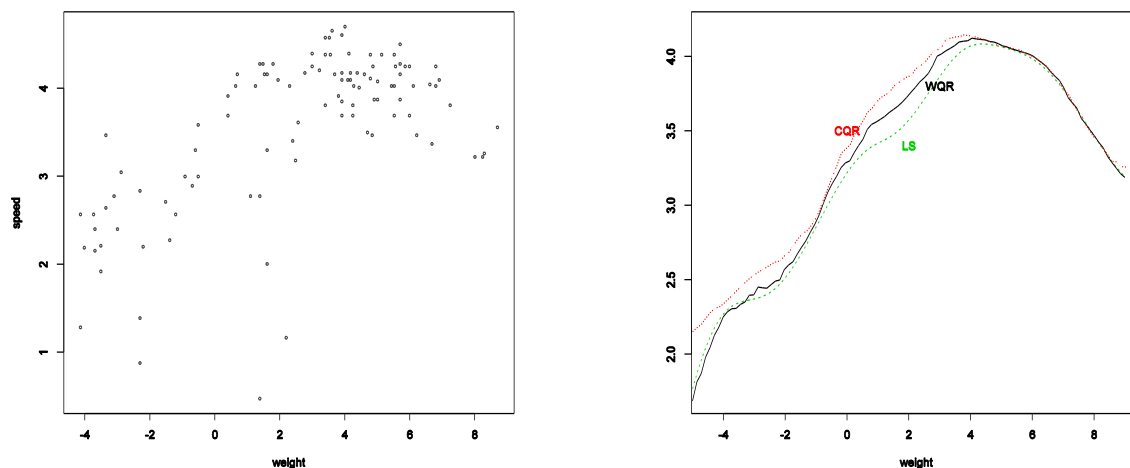


Figure 3: Analysis of the Mammals data using WQR, CQR and LS. The left panel is the scatterplot, the right is the estimated regression functions.

## 5 Discussion

In this paper, we have proposed weighted quantile regression and proved its nice theoretical properties. We have shown that the weighted quantile regression techniques are easy to implement and are very flexible. The weights, derived based on sparsity function, can automatically combine the optimal strength of each quantile into the final model, thus significantly improve the efficiency of WQR estimator. It is shown that the relative efficiency of WQR estimator with respect to both LS and CQR estimator is always larger than 70%. In regularization framework, the penalized WQR can be employed to conduct variable selection. Results of simulation indicate that the penalized WQR estimator have the extraordinary ability to select relevant variables.

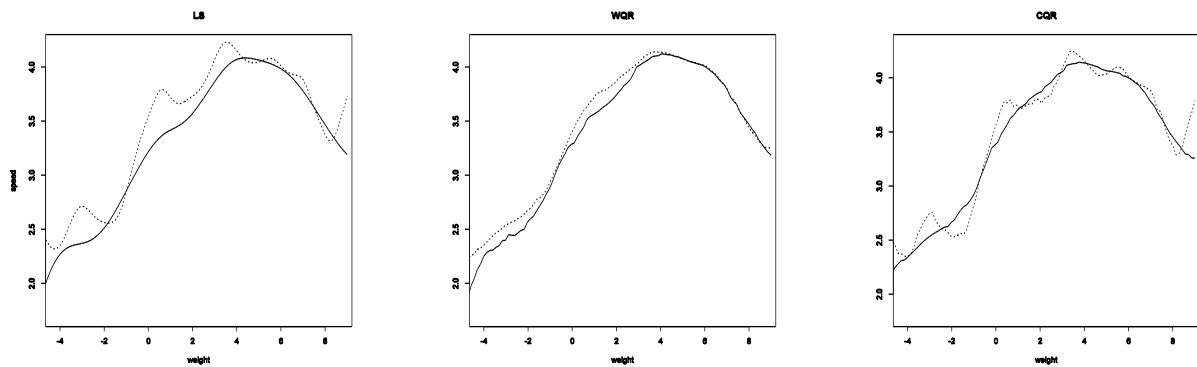


Figure 4: Estimated regression function with original data and data removed several possible outliers.

(--- Original; — Removed)

## References

- [1] Fan, J. (1993) Local linear regression smoothing and their minimax efficiencies. *The Annals of Statistics*, 21, 196-216.
- [2] Fan, J. and Gijbels, I. (1992) Variable bandwidth and local linear regression smoothers. *Ann. Statist.*, 20, 2008-2036.
- [3] Fan, J., and Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association*, 96(456), 1348-1360.
- [4] Kai, B., Li, R. and Zou, H. (2010) Local composite quantile regression smoothing: an efficient and safe alternative to local polynomial regression. *J.R.Statist.Soc.B*, 72, 49-89.
- [5] Koenker, R. (2005) *Quantile Regression*. Cambridge: Cambridge University Press.
- [6] Koenker, R. and Bassett, G. (1982) Robust tests for heteroscedasticity based on regression quantiles. *Econometrica*, 50, 43-61.
- [7] Koenker, R. and Machado, J, A.F. (1999) Goodness of fit and related inference processes for quantile regression. *J.Am.Statist.Ass.*, 94, 1296-1310.
- [8] Meinshausen, N., and Bühlmann, P. (2010). Stability selection. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 72(4), 417- 473.
- [9] Parzen, E. (1979). Nonparametric statistical data modeling. *Journal of the American Statistical Association*, 74(365), 105-121.
- [10] Ruppert, D., Sheather, S. J., and Wand, M. P. (1995). An effective bandwidth selector for local least squares regression. *Journal of the American Statistical Association*, 90(432), 1257-1270.
- [11] Tukey, J. W. (1965). Which part of the sample contains the information?. *Proceedings of the National Academy of Sciences of the United States of America*, 53(1), 127.
- [12] Yu, K. and Jones, M. C. (1998) Local linear quantile regression. *J.Am.Statist.Ass.*,93, 228-237.
- [13] Zhang, C. H. (2010). Nearly unbiased variable selection under minimax concave penalty. *The Annals of Statistics*, 38(2), 894-942.
- [14] Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American statistical association*, 101(476), 1418-1429.
- [15] Zou, H. and Yuan, M. (2008) Composite quantile regression and the oracle model selection theory. *Ann. Statist.*, 36, 1108-1126.