# Composite quantile regression for massive datasets

Rong Jiang[1]; Xueping Hu[2]; Keming Yu[3]; Weimin Qian[4]

*1 Department of Applied Mathematics, College of Science, Donghua University, Shanghai, People's Republic of China*
*2 School of Mathematics and Computational Science, Anqing Normal University, Anqing, People's Republic of China*
*3 Department of Mathematics, Brunel University London, Middlesex, UK*
*4 School of Mathematical Sciences, Tongji University, Shanghai, People's Republic of China*

**Abstract**

Analysis of massive datasets is challenging owing to limitations of computer primary memory. Composite quantile regression (CQR) is a robust and efficient estimation method. In this paper, we extend CQR to massive datasets and propose a divide-and-conquer CQR method. The basic idea is to split the entire dataset into several blocks, applying the CQR method for data in each block, and finally combining these regression results via weighted average. The proposed approach significantly reduces the required amount of primary memory, and the resulting estimate will be as efficient as if the entire data set was analyzed simultaneously. Moreover, to improve the efficiency of CQR, we propose a weighted CQR estimation approach. To achieve sparsity with high-dimensional covariates, we develop a variable selection procedure to select significant parametric components and prove the method possessing the oracle property. Both simulations and data analysis are conducted to illustrate the finite sample performance of the proposed methods.

*Keywords:* Massive dataset; Divide and conquer; Composite quantile regression; Variable selection.

## 1. Introduction

In recent years, statistical analysis of massive data sets has become a subject of increased interest. Datasets grow in size in part because they are increasingly being collected by ubiquitous information sensing mobile devices, remote sensing technologies, and wireless sensor networks, among others. It is not uncommon that databases have hundreds of fields, billions of records and terabytes of information. For instance, Wal-mart handled more than 140 million customer transactions per week in 2014, and Facebook had 1.55 billion monthly active users in the third quarter of 2015. However, the number of observations that can be stored in primary memory is often restricted. The available memory, though large, is finite. Many computing environments also limit the maximum array size allowed and this can be much smaller and even independent of the available memory. Therefore, the surge of massive data presents challenges to both computer scientists and statisticians in terms of data storage, computation, and statistical analysis.

Notwithstanding that new statistical thinking and methods are needed for massive data sets, our focus is on fitting standard statistical models to massive data sets whose size exceeds the capacity of a single computer. There are two major challenges in analyzing massive data sets: 1) the data can be too big to be held in a computer's memory; and 2) the computing task can take too long to wait for the results. These barriers can be approached either with newly developed statistical methodologies or computational methodologies. Some statisticians have made important contributions and are pushing the frontier. Examples are subsampling based approaches (Liang et al. 2013; Kleiner et al. 2014; Ma et al. 2015) and divide and conquer approaches (Lin and Xi 2011; Chen and Xie 2014; Schifano et al. 2016); see Wang et al. (2015) for a review.

In this paper, we consider a divide and conquer approach for massive data sets. "Divide and conquer" (or "divide and recombine", or "split and conquer", or "split and merge"), in particular, has become a popular approach for the analysis of large complex data. The approach is appealing because the data are first divided into subsets and then

[1]*E-mail address*: jrtrying@dhu.edu.cn (R. Jiang), hxprob@163.com (X.P. Hu), Keming.Yu@brunel.ac.uk (K.M. Yu), wmqian2003@aliyun.com (W.M. Qian)

numeric and visualization methods are applied to each of the subsets separately. The divide and conquer approach culminates by aggregating the results from each subset to produce a final solution. Our task is to investigate whether the combined overall result can be as good as the result obtained from analyzing the entire dataset. Recently, Fan et al. (2007) considered a divide and conquer algorithm for the linear model based on least square method. Lin and Xi (2011) developed a computation and storage efficient algorithm for estimating equations in massive data sets using a divide and conquer algorithm. Schifano et al. (2016) extended the work of Lin and Xi (2011) to online updating for stream data. Chen and Xie (2014) considered a divide and conquer approach for generalized linear models where both the sample size and the number of covariates are large. Xu et al. (2017) proposed a novel block average quantile regression method for massive dataset.

Existing estimation procedures for massive data sets were built on either least squares or quantile regression methods. However, the least squares method is sensitive to outliers and does not perform well when the error distribution is heavily skewed. The quantile regression method is an obvious alternative to the least squares. However, the relative efficiency of the quantile regression can be arbitrarily small when compared with the least squares.

In contrast to the above methods, the composite quantile regression (CQR) was first proposed by Zou and Yuan (2008) for estimating the regression coefficients in the classical linear regression model. Zou and Yuan (2008) showed that the relative efficiency of the CQR estimator compared with the least squares estimator is greated than 70% regardless of the error distribution. Furthermore, the CQR estimator could be more efficient and sometimes arbitrarily more efficient than the least squares estimator. Based on CQR, Kai et al. (2010) proposed the local polynomial CQR estimators for estimating the nonparametric regression function and its derivative. Kai et al. (2011) studied semiparametric CQR estimates for the semiparametric varying-coefficient partially linear model. For other references about CQR method see Tang et al. (2012, 2015), Jiang et al. (2012, 2013, 2014, 2015, 2016a, 2016b, 2018), Ning and Tang (2014), Zhang et al. (2016), Tian et al. (2016), Zhao et al. (2017) and so on. These nice theoretical properties of CQR in linear regression motivate us to consider CQR method for massive datasets. Furthermore, we study the construction of confidence intervals and hypothesis tests.

The CQR method is a sum of different quantile regressions with equal weights. Intuitively, equal weights are not optimal in general, and hence Jiang et al. (2012) proposed a weighted CQR (WCQR) estimation. The WCQR is augmented using a data-driven weighting scheme. With the error distribution unspecified, the WCQR estimators share robustness from quantile regression and achieve nearly the same efficiency as the oracle maximum likelihood estimator for a variety of error distributions including the normal, mixed-normal, Student's t, Cauchy distributions, etc. Moreover, by comparing asymptotic relative efficiency theoretically and numerically, the WCQR method all outperforms the CQR method; only when the error density is logistic or close to logistic distribution, standard CQR has good performance compared with weighted CQR (see Zhao and Lian, 2016). Thus, we also consider WCQR method for massive datasets.

In practice, it is common to have a large number of candidate predictor variables available, and they are included in the initial stage of modeling for the consideration of removing potential modeling bias (Fan and Li, 2001). However, it is undesirable to keep irrelevant predictors in the final model since this makes it difficult to interpret the resultant model and may decrease its predictive ability. In the regularization framework, many different types of penalties have been introduced to achieve variable selection. The $L_1$ penalty was used in the LASSO proposed by Tibshirani (1996) for variable selection. Fan and Li (2001) proposed a unified approach via nonconcave penalized least squares regression, which simultaneously performs variable selection and coefficient estimation. By using adaptive weights for penalizing different coefficients in the LASSO penalty, Zou (2006) introduced the adaptive LASSO and demonstrated its oracle properties. Zou and Yuan (2008) studied the LASSO for CQR (CQR-LASSO). The CQR-LASSO is robust and performs nearly like a CQR-oracle model selector. Therefore, we consider CQR-LASSO to study model selection for massive datasets. Model selection with a fixed number of parameters has been widely pursued in the last decades. However, to reduce possible modeling biases, many variables are introduced in practice. Fan and Peng (2004), Lam and Fan (2008) and Fan and Lv (2011) advocated that, in most model selection problems, the number of parameters should be large and grow with the sample size. We allow the number of parameters to depend on the sample size under some certain conditions. We consider data sets as extraordinarily large (massive datasets), if they do not fit into a single computer. We propose a divide and conquer approach to solve the problem and illustrate it using the aforementioned CQR-LASSO method. However, since each selection variable is estimated from a different subset of data, the set of non-zero elements of parameters can differ. To obtain a combined estimator, we use a majority voting method (Meinshausen and Buhlmann, 2010; Shah and Samworth, 2013; and Chen and Xie, 2014) to estimate the true nonzero set.

In this paper, we consider CQR, WCQR and variable selection for massive datasets based on divide and conquer approach. We make the following three major contributions:

(1) Xu et al. (2017) proposed a novel block average quantile regression method for massive dataset. Their method requires a strong assumption that the covariance matrixes of each block are equal, and the number of blocks is fixed. Zhao et al. (2017) proposed a CQR method for massive datasets which is simple average CQR results in each block. However, their estimator requires on a strong assumption that the number of sub-datasets $K$ is of order $O(n^r)$ where $r < 1/3$ and $n$ is the sample size of sub-datasets. We developed two methods: divide-and-conquer CQR (DC-CQR) and divide-and-conquer WCQR (DC-WCQR). DC-CQR (or DC-WCQR) is a form of weighted average of CQR (or WCQR) results in each block, which is asymptotically equivalent to the estimator obtained from analyzing the entire data without additional conditions. Moreover, the order of $K = o\left(n_{\min}^{1/2} (\log\log n_{\min})^{-3/2}\right)$ in Theorem 2.1 is higher than the order of $K$ in Zhao et al. (2017).

(2) We study the construction of confidence intervals and hypothesis tests.

(3) We study an effective and robust variable selection procedure based on the DC-WCQR method to select significant parametric components in the linear model.

The paper is organized as follows. In Section 2, we introduce the composite quantile procedure for massive datasets. In Section 3, the weighted composite quantile regression method is proposed. The variable selection method is developed in Section 4. A numerical implementation is introduced in Section 5. Both simulation examples and the application on real data are given in Section 6 to illustrate the proposed procedures. Final remarks are given in Section 7. All technical proofs are deferred to the Appendix.

## 2. Composite quantile regression

In the section, we propose the divide-and-conquer CQR (DC-CQR) method for massive datasets. The standard CQR method is first reviewed.

### *2.1. Standard CQR method*

In this paper, we consider the following linear model

$$y_i = \mathbf{x}_i^\top \beta_0 + \varepsilon_i, \quad i = 1, \ldots, N, \tag{2.1}$$

where $y_i$ is the univariate response, $\mathbf{x}_i$ is a vector of $p$-dimensional covariates, $\beta_0$ is the unknown parameter, $\varepsilon_i$ is independent and identically distributed unknown random error. Zou and Yuan (2008) proposed CQR method to estimate $\beta_0$ as follows

$$(\hat{b}_1, \ldots, \hat{b}_Q, \hat{\beta}^{CQR}) = \arg \min_{(b_1,\ldots,b_Q,\beta)} \sum_{q=1}^{Q} \sum_{i=1}^{N} \rho_{\tau_q} \left\{ y_i - b_q - \mathbf{x}_i^\top \beta \right\}, \tag{2.2}$$

where $\rho_{\tau_q}(r) = \tau_q r - rI(r < 0), q = 1, \ldots, Q$, are $Q$ check loss functions with $0 < \tau_1 < \ldots < \tau_Q < 1$ and $\hat{b}_q$ is the estimator of $b_{\tau_q}$, $b_{\tau_q}$ is the $\tau_q$ quantile of $\varepsilon_i$, $q = 1, \ldots, Q$.

**Remark 2.1.** In general, given $Q$, one can use the equally spaced quantiles at $\tau_q = q/(Q+1)$ for $q = 1, \ldots, Q$, see Zou and Yuan (2008). Moreover, we can use redefined AIC or BIC criteria (Tian et al, 2016) to select $Q$ as follows

$$AIC(Q) = \frac{2}{N} RS_Q + \frac{2}{N}(2Q + p - 1), \quad Q = 1, \ldots Q_{\max},$$
$$BIC(Q) = \frac{2}{N} RS_Q + \frac{\log(N)}{N}(2Q + p - 1), \quad Q = 1, \ldots Q_{\max},$$

where $Q_{\max}$ is a possible upper bound and $RS_Q = \sum_{q=1}^{Q} \sum_{i=1}^{N} \rho_{\tau_q}\{y_i - \hat{b}_q - \mathbf{x}_i^\top \hat{\beta}^{CQR}\}$ is the residual sum of the estimated CQR model. The resulting optimal value of $Q$ is the smallest redefined AIC or BIC values.

### 2.2. *DC-CQR method for massive datasets*

It is infeasible to solve the optimization problem in (2.2), when the sample size $N$ is too large. Our work builds upon the approach of Lin and Xi (2011) and Chen and Xie (2014) who introduced dividing the dataset into several blocks with each are containable in the computer's memory. We then implement the standard CQR method on data in each subset and combine the results. Concretely, we now show that DC-CQR method can be obtained using the following three key steps.

Step 2.1: Without loss of generality, the entire data set is partitioned into $K$ subsets, so that the $k$th subset contains $n_k$ observations: $(\mathbf{x}_{k,i}, y_{k,i})$, $i = 1, \ldots, n_k$;

Step 2.2: Apply standard composite quantile regression on data within each subset, and obtain the estimators $\hat{\beta}_k^{CQR}$, $k = 1, \ldots, K$, using the methodology in solving equation (2.2);

Step 2.3: The combined estimator DC-CQR method, as a weighted average of $\hat{\beta}_k^{CQR}$, $k = 1, \ldots, K$, is

$$\hat{\beta}^{DC-CQR} = \left( \sum_{k=1}^{K} \mathbf{X}_k^\top \mathbf{X}_k \right)^{-1} \sum_{k=1}^{K} \mathbf{X}_k^\top \mathbf{X}_k \hat{\beta}_k^{CQR},$$

where $\mathbf{X}_k = (\mathbf{x}_{k,1}, \ldots, \mathbf{x}_{k,n_k})^\top$.

**Remark 2.2.** The size of $\left(\mathbf{X}_k^\top \mathbf{X}_k, \hat{\beta}_k^{CQR}\right)$ is $p^2 + p$, so we only need to save $K(p^2 + p)$ numbers, which achieves very efficient compression since both $K$ and $p$ are far less than $N$ in practice.

### 2.3. *Asymptotic normality of the resulting estimator*

To establish the asymptotic properties of the proposed estimators, the following technical conditions are imposed.

**C1.** $\mathbf{C} = \lim_{N\to\infty} \frac{1}{N} \mathbf{X}^\top \mathbf{X}$ is a positive definite matrix, where $\mathbf{X} = (\mathbf{X}_1^\top, \ldots, \mathbf{X}_K^\top)^\top$.

**C2.** The error $\varepsilon$ has cumulative distribution function $F(\cdot)$ and density $f(\cdot)$. The density $f(\cdot)$ is positive and continuous at the $\tau_q$-th quantile $b_{\tau_q}$.

**Theorem 2.1.** Suppose the sample size of the $k$th subset is $n_k = O(N/K)$, $k = 1, \ldots, K$, and that $n_{\max}/n_{\min} = O(1)$. Assume that Conditions **C1** and **C2** are satisfied and $K = o\left(n_{\min}^{1/2} (\log \log n_{\min})^{-3/2}\right)$, then

$$\sqrt{N}(\hat{\beta}^{DC-CQR} - \beta_0) \xrightarrow{L} \mathcal{N}\left(0, \mathbf{C}^{-1} R_1\right), \tag{2.3}$$

where $n_{\max} = \max_{1\le k\le K} n_k$, $n_{\min} = \min_{1\le k\le K} n_k$, $\xrightarrow{L}$ stands for convergence in distribution, and

$$R_1 = \frac{\sum_{q,q'=1}^{Q} \min(\tau_q, \tau_{q'}) \left(1 - \max(\tau_q, \tau_{q'})\right)}{\left[\sum_{q=1}^{Q} f(b_{\tau_q})\right]^2}.$$

**Remark 2.3.** Conditions **C1** and **C2** are basically the same conditions for establishing the asymptotic normality of quantile regression (Koenker, 2005).

**Remark 2.4.** The limiting distribution of $\hat{\beta}^{DC-CQR}$ in (2.3) is that of $\hat{\beta}_N$ in Theorem 2.1 from Zou and Yuan (2008), where the entire data is analyzed. Thus, the DC-CQR estimator is asymptotically equivalent to the corresponding estimator using the full datasets.

**Remark 2.5.** It is shown in Theorem 2.1 that the resulting estimate is robust to the choice of the block size $K$ and subsets size $n_k, k = 1, \ldots, K$. Thus $K$ and $n_k$ are chosen so that the estimation of $\beta_0$ can be easily handled within each block.

### 2.4. *Estimation of the asymptotic variance*

One can see from (2.3) that the asymptotic variance involves the density of the errors $f(\cdot)$ and $b_{\tau_q}$, $q = 1, \ldots, Q$. In practice, the error density $f(\cdot)$ and $b_{\tau_q}$ are generally unknown. We can use kernel estimation

$$\frac{1}{N} \sum_{i=1}^{N} K_h \left(\hat{\varepsilon}_i - \cdot\right)$$

to estimate $f(\cdot)$, where $K_h(\cdot) = K(\cdot/h)/h$, $K(\cdot)$ is a kernel function, and $h$ is a selected bandwidth. The estimator $\hat{b}_q$ of $b_{\tau_q}$ is the sample $\tau_q$-quantile of $\{\hat{\varepsilon}_i, i = 1, \dots, N\}$, where $\hat{\varepsilon}_i = y_i - \mathbf{x}_i^\top \hat{\beta}^{DC-CQR}$. Therefore, we can obtain the estimation of $f(b_{\tau_q})$ by $\hat{f}(\hat{b}_q)$ for $q = 1, \dots, Q$. When the available computer memory is much smaller than $N$, sorting $\{\hat{\varepsilon}_i, i = 1, \dots, N\}$ becomes impossible. To overcome the difficulty, Li et al. (2013) proposed an approach to estimate population parameters from a massive data set. Their method reduces the required amount of primary memory, and the resulting estimate is as efficient as if the entire data set is analyzed simultaneously. However, their method is available under the condition that all subsets are of equal size. Now we extend the method of Li et al. (2013) to different subset size.

Suppose that $s_1, \dots, s_N$ is an independent and identically distributed sample from population $G$. We are interested in estimating parameter $\theta(G)$ of the population. In the same way as in Section 2.2, the entire data set is partitioned into $K$ subsets, so that the $k$th subset contains $n_k$ observations. We use the same estimation method for each block. Denote by $\hat{\theta}_k$ the resulting estimate based on the subsets in the $k$th block. We estimate $\theta(G)$ by weighted averaging of $\hat{\theta}_k$, that is

$$\hat{\theta}_N = \frac{1}{N} \sum_{k=1}^{K} n_k \hat{\theta}_k. \tag{2.4}$$

Denote by $\hat{\theta}$ the estimator based on all samples. Suppose that $\sqrt{N}(\hat{\theta} - \theta) \xrightarrow{L} \mathcal{N}\left(0, \sigma_0^2\right)$ as $N \to \infty$. Thus $\sqrt{n_k}(\hat{\theta}_k - \theta) \xrightarrow{L} \mathcal{N}\left(0, \sigma_0^2\right)$ as $n_k \to \infty$. This implies that

$$\sqrt{N}(\hat{\theta}_N - \theta) = \frac{1}{\sqrt{N}} \sum_{k=1}^{K} n_k \left(\hat{\theta}_k - \theta\right) \xrightarrow{L} \mathcal{N}\left(0, \sigma_0^2\right).$$

This implies that the resulting estimator $\hat{\theta}_N$ is as efficient as $\hat{\theta}$. In other words, the resulting estimate is as efficient as if all data were simultaneously used to compute the estimate.

By the form of (2.4), we can estimate $b_{\tau_q}$, $q = 1, \dots, Q$, as follows

$$\hat{b}_q = \frac{1}{N} \sum_{k=1}^{K} n_k \hat{b}_{q,k}, \tag{2.5}$$

where $\hat{b}_{q,k}$ is the sample $\tau_q$-quantile of $\{\hat{\varepsilon}_{k,i}, i = 1, \dots, n_k\}$ and $\hat{\varepsilon}_{k,i} = y_{k,i} - \mathbf{x}_{k,i}^\top \hat{\beta}^{DC-CQR}$. Moreover, the weighted combined estimator of $f(\cdot)$ is given by

$$\hat{f}_N(\cdot) = \frac{1}{N} \sum_{i=1}^{K} n_k \hat{f}_k(\cdot), \tag{2.6}$$

where $\hat{f}_k(\cdot) = \frac{1}{n_k} \sum_{i=1}^{n_k} K_{h_k}\left(\hat{\varepsilon}_{k,i} - \cdot\right), k = 1, \dots, K$, are kernel density estimation with each subset. Thus, by (2.3),(2.5) and (2.6), $N\left(\mathbf{X}^\top \mathbf{X}\right)^{-1} \hat{R}_1$ is a consistent estimate of the covariance matrix of estimator $\hat{\beta}^{DC-CQR}$, where

$$\hat{R}_1 = \frac{\sum_{q,q'=1}^{Q} \min(\tau_q, \tau_{q'}) \left(1 - \max(\tau_q, \tau_{q'})\right)}{\left[\sum_{q=1}^{Q} \hat{f}(\hat{b}_q)\right]^2}. \tag{2.7}$$

**Remark 2.6.** The bandwidth selection is taken by $h_k = \left(\frac{n_k}{N}\right)^{1/5} h_k^*$ as selected by Li et al. (2013), where $h_k^*$ is selected as $0.9 \times 1.06 \times \sigma_k \times n_k^{-1/5}$ and $\sigma_k$ is the standard deviation of error variable, and we can use $\hat{\sigma}_k = std(\hat{\varepsilon}_{k,1}, \dots, \hat{\varepsilon}_{k,n_k})$ to estimate $\sigma_k$, where $std$ is the sample standard deviation.

### *2.5. Confidence intervals and hypothesis testing*

Based on Theorem 2.1 and (2.7), $N\left(\mathbf{X}^\top \mathbf{X}\right)^{-1} \hat{R}_1$ is a consistent estimate of the covariance matrix of estimator $\hat{\beta}^{DC-CQR}$. We first construct univariate confidence intervals and multivariate confidence regions. Denote $\hat{\sigma}_j^2 =$

$N\left(\mathbf{X}^\top\mathbf{X}\right)^{-1}_{jj}\hat{R}_1$. A confidence interval (CI) at a confidence level $1-\alpha$ for a component $\beta_{0,j}$ of the true parameter is defined as

$$CI_j = \left[\hat{\beta}_j^{DC-CQR} - \frac{\hat{\sigma}_j}{\sqrt{N}}\Phi^{-1}(1-\alpha/2), \hat{\beta}_j^{DC-CQR} + \frac{\hat{\sigma}_j}{\sqrt{N}}\Phi^{-1}(1-\alpha/2)\right],$$

where $\Phi(\cdot)$ is the standard normal cumulative distribution function. Furthermore, for a finite set $L \subset \{1,\ldots,p\}$, we define a confidence region (CR) at a confidence level $1-\alpha$ as

$$CR_L = \left\{\beta_{0,L} \in R^{|L|} : \hat{R}_1^{-1}\left(\hat{\beta}_L^{DC-CQR} - \beta_{0,L}\right)^\top \mathbf{X}_L^\top\mathbf{X}_L\left(\hat{\beta}_L^{DC-CQR} - \beta_{0,L}\right) \le q\chi^2_{|L|}(1-\alpha)\right\},$$

where $q\chi^2_{|L|}(1-\alpha)$ is the $1-\alpha$ quantile of the the chi-square distribution with $|L|$ degrees of freedom and $|L|$ is the cardinality of $L$.

Theorem 2.1 can also be used to test significance of variables. For a component $j \in \{1,\ldots,p\}$, we test

$$H_0^1 : \beta_{0,j} = 0 \;\; versus \;\; H_1^1 : \beta_{0,j} \neq 0,$$

and we have under $H_0^1$,

$$\left|\sqrt{N}\hat{\beta}_j^{DC-CQR}/\hat{\sigma}_j\right| \le \Phi^{-1}(1-\alpha/2),$$

with probability $1-\alpha$. We also test simultaneous significance as follows. For a finite set $L \subset \{1,\ldots,p\}$, we test

$$H_0^2 : \beta_{0,j} = 0 \; for\ all\ components\ j \in L \;\; versus \;\; H_1^2 : \beta_{0,j} \neq 0 \; for\ at\ least\ one\ component\ j \in L,$$

and we have under $H_0^2$,

$$\hat{R}_1^{-1}\left(\hat{\beta}_L^{DC-CQR}\right)^\top \mathbf{X}_L^\top\mathbf{X}_L\hat{\beta}_L^{DC-CQR} \le q\chi^2_{|L|}(1-\alpha),$$

with probability $1-\alpha$.

## 3. Weighted composite quantile regression

Note that the CQR method uses the same weight for different quantile regression models. Jiang et al. (2012, 2016) considered weighted CQR to estimate $\beta_0$ in model (2.1) by minimizing

$$\sum_{q=1}^{Q}\sum_{i=1}^{N} w_q\rho_{\tau_q}\{y_i - b_q - \mathbf{x}_i^\top\beta\},$$

where $\mathbf{w} = (w_1,\ldots,w_Q)^\top$ is a vector of weights and the components in the weight vector $\mathbf{w}$ are allowed to be negative, since $\left\{\sum_{i=1}^N \rho_{\tau_q}(y_i - b_q - \mathbf{x}_i^\top\beta)\right\}_{q=1}^Q$ may not be positively correlated. As in Jiang et al. (2016), under mild assumptions, the asymptotic variance of the estimator is

$$\mathbf{C}^{-1}\left[\sum_{q=1}^{Q} w_q f(b_{\tau_q})\right]^{-2}\sum_{q,q'=1}^{Q} w_q w_{q'}\min(\tau_q,\tau_{q'})\left(1-\max(\tau_q,\tau_{q'})\right).$$

Thus, the optimal weight $\mathbf{w}_{opt}$, which minimizes the asymptotic variance of the estimator, is

$$\mathbf{w}_{opt} = c\Omega^{-1}\mathbf{f},$$

for any constant $c \neq 0$; since $c$ is no effect, we take $c = 1$, where $\mathbf{f} = \left(f(b_{\tau_1}),\ldots,f(b_{\tau_Q})\right)^\top$ and $\Omega$ is a $Q\times Q$ matrix with the $(q,q')$ element $\Omega_{qq'} = \min(\tau_q,\tau_{q'})\left(1-\max(\tau_q,\tau_{q'})\right)$. Furthermore, the usual nonparametric density estimation methods, such as kernel smoothing based on estimated residual $\hat{\varepsilon}_i = y_i - \mathbf{x}_i^\top\hat{\beta}^{CQR}$, can provide a consistent estimation $\hat{\mathbf{f}}$ of $\mathbf{f}$. The $\hat{\mathbf{w}} = \Omega^{-1}\hat{\mathbf{f}}$ is a nonparametric estimator of $\mathbf{w}_{opt}$, where $\hat{\mathbf{f}}$ can be obtained by (2.5) and (2.6) in Section 2.4.

### 3.1. DC-WCQR method for massive datasets

It can be seen that we must obtain the estimation $\hat{b}_q$ of $b_{\tau_q}$, $q = 1, \ldots, Q$ to get the optimal weight $\mathbf{w}_{opt}$. Thus, we use a two step WCQR estimation procedure (Zhao and Lian, 2016) to reduce the computational complexity. In the first step, we use the standard CQR to get the root-n consistent estimators $\hat{b}_q$ of $b_{\tau_q}$, $q = 1, \ldots, Q$. In the second step, these intercepts are plugged into the weighted functional and we only solve for $\beta$,

$$\hat{\beta}^{WCQR} = \arg\min_{\beta} \sum_{q=1}^{Q} \sum_{i=1}^{N} \hat{w}_q \rho_{\tau_q} \left\{ y_i - \hat{b}_q - \mathbf{x}_i^\top \beta \right\}. \tag{3.1}$$

For the massive data, we establish the DC-WCQR method as follows.

Step 3.1: Without loss of generality, the entire data set is partitioned into $K$ subsets, and that the $k$th subset contains $n_k$ observations: $(\mathbf{x}_{k,i}, y_{k,i})$, $i = 1, \ldots, n_k$;

Step 3.2: For each subset, we first use the standard CQR to get the estimators $\hat{b}_q$ and $\hat{w}_q$ of $b_{\tau_q}$ and $w_q$ for $q = 1, \ldots, Q$, respectively. Secondly, apply standard weighted composite quantile regression on data with $\hat{b}_q$ and $\hat{w}_q$, $q = 1, \ldots, Q$, and obtain the estimators $\hat{\beta}_k^{WCQR}$, $k = 1, \ldots, K$ using the methodology in solving equation (3.1);

Step 3.3: The combined estimator DC-WCQR method, as a weighted average of $\hat{\beta}_k^{WCQR}$, $k = 1, \ldots, K$, is

$$\hat{\beta}^{DC-WCQR} = \left( \sum_{k=1}^{K} \mathbf{X}_k^\top \mathbf{X}_k \right)^{-1} \sum_{k=1}^{K} \mathbf{X}_k^\top \mathbf{X}_k \hat{\beta}_k^{WCQR}.$$

### 3.2. Asymptotic normality of the resulting estimator

To reveal the merits of the proposed DC-WCQR method, we now establish the asymptotic normality of $\hat{\beta}^{DC-WCQR}$.

**Theorem 3.1.** Under the same conditions as in Theorem 2.1, we have

$$\sqrt{N}(\hat{\beta}^{DC-WCQR} - \beta_0) \xrightarrow{L} \mathcal{N}\left(0, \mathbf{C}^{-1}\left(\mathbf{f}^\top \Omega^{-2} \mathbf{f}\right)^{-1}\right). \tag{3.2}$$

**Remark 3.1.** The limiting distribution of $\hat{\beta}^{DC-WCQR}$ in (3.2) is that of $\hat{\beta}_N$ in Theorem 2 from Jiang et al. (2016), where the entire data is analyzed. Thus, the DC-WCQR estimator is asymptotically equivalent to the corresponding estimator using the full data sets. Moreover, as for the results in Jiang et al. (2016), the WCQR method achieves nearly the same efficiency as the maximum likelihood estimator.

### 3.3. Confidence intervals and hypothesis testing

By (3.2), $N(\mathbf{X}^\top \mathbf{X})^{-1}\left(\hat{\mathbf{f}}^\top \Omega^{-2} \hat{\mathbf{f}}\right)^{-1}$ is a consistent estimate of the covariance matrix of estimator $\hat{\beta}^{DC-WCQR}$. Denote $\tilde{\sigma}_j^2 = N(\mathbf{X}^\top \mathbf{X})_{jj}^{-1}\left(\hat{\mathbf{f}}^\top \Omega^{-2} \hat{\mathbf{f}}\right)^{-1}$. Thus, based on Theorem 3.1, we can construct confidence intervals and hypothesis testing as follows.

$$CI_j = \left[ \hat{\beta}_j^{DC-WCQR} - \frac{\tilde{\sigma}_j}{\sqrt{N}} \Phi^{-1}(1-\alpha/2), \hat{\beta}_j^{DC-WCQR} + \frac{\tilde{\sigma}_j}{\sqrt{N}} \Phi^{-1}(1-\alpha/2) \right],$$

$$CR_L = \left\{ \beta_{0,L} \in R^{|L|} : \left(\hat{\beta}_L^{DC-WCQR} - \beta_{0,L}\right)^\top \mathbf{X}_L^\top \mathbf{X}_L \left(\hat{\mathbf{f}}^\top \Omega^{-2} \hat{\mathbf{f}}\right) \left(\hat{\beta}_L^{DC-WCQR} - \beta_{0,L}\right) \le q\chi_{|L|}^2(1-\alpha) \right\}.$$

To test

$$H_0^1 : \beta_{0,j} = 0 \;\; versus \;\; H_1^1 : \beta_{0,j} \neq 0,$$

and we have under $H_0^1$,

$$\left| \sqrt{N} \hat{\beta}_j^{DC-WCQR} / \tilde{\sigma}_j \right| \le \Phi^{-1}(1-\alpha/2),$$

with probability $1-\alpha$. To test

$$H_0^2 : \beta_{0,j} = 0 \; for\ all\ components\ j \in L \;\; versus \;\; H_1^2 : \beta_{0,j} \neq 0 \; for\ at\ least\ one\ component\ j \in L,$$

and we have under $H_0^2$,

$$\left(\hat{\beta}_L^{DC-WCQR}\right)^\top \mathbf{X}_L^\top \mathbf{X}_L \left(\hat{\mathbf{f}}^\top \Omega^{-2} \hat{\mathbf{f}}\right) \hat{\beta}_L^{DC-WCQR} \le q\chi_{|L|}^2(1-\alpha),$$

with probability $1-\alpha$.

## 4. Variable selection

The adaptive LASSO (see Zou, 2006) can be viewed as a generalization of the LASSO penalty. Basically the idea is to penalize the coefficients of different covariates at a different level by using adaptive weights. Suppose the dataset of size $N$ is divided into $K$ subsets, and that the $k$th subset has $n_k$ observations $(\mathbf{x}_{k,i}, y_{k,i})$, $i = 1, \ldots, n_k$. The adaptive LASSO penalized weighted composite quantile regression estimator (PWCQR) for the $k$th subset, $k = 1, \ldots, K$ , denoted by $\hat{\beta}_k^{PWCQR}$, is the minimizer of the following function

$$\sum_{q=1}^{Q}\sum_{i=1}^{n_k} \hat{w}_q \rho_{\tau_q}\left\{y_{k,i} - b_q - \mathbf{x}_{k,i}^\top \beta\right\} + n_k \lambda_k \sum_{j=1}^{p} \frac{\left|\beta_j\right|}{\left|\hat{\beta}_j^{DC-WCQR}\right|^2}, \tag{4.1}$$

where $\lambda_k$ is a nonnegative regularization parameter and $\hat{w}_q$ is defined in Section 3. Under the setup, the penalized estimator $\hat{\beta}_k^{PWCQR}$ has the oracle property, see Theorem 6 in Jiang et al. (2012). Denote by $\hat{A}_k = \left\{j : \hat{\beta}_{k,j}^{PWCQR} \neq 0\right\}$ the set of non-zero elements of $\hat{\beta}_k^{PWCQR}$. Since each $\hat{\beta}_k^{PWCQR}$ is estimated from a different subset of data, the $\hat{A}_k$ can differ, $k = 1, \ldots, K$. To obtain a combined estimator of $\beta_0$ from $\hat{\beta}_k^{PWCQR}$, $k = 1, \ldots, K$, we use a majority voting method to estimate the true nonzero set $A^* = \{j : \beta_{0j} \neq 0\}$. We take

$$\hat{A}^* = \left\{ j : \sum_{k=1}^{K} I\left(\hat{\beta}_{k,j}^{PWCQR} \neq 0\right) > d \right\},$$

as the set of selected variables of the combined estimator, where $d \in [0, K)$ is a prespecified threshold and $I$ is the indicator function. The theoretical development suggests that the choice of a fixed threshold, does not affect the asymptotic results, see Meinshausen and B*ü*hlmann (2010), Shah and Samworth (2013) and Chen and Xie (2014). In Section 6, we use $d = K/2$ as selected by Chen and Xie (2014).

Take $B = diag(v_1, \ldots, v_p)$ to be the $p \times p$ voting matrix with $v_j = 1$, if $\sum_{k=1}^{K} I\left(\hat{\beta}_{k,j}^{PWCQR} \neq 0\right) > d$ and 0 otherwise, and let $A = B_{\hat{A}^*}$ be the $p \times |\hat{A}^*|$ selection matrix, where $B_{\hat{A}^*}$ stands for an $p \times |\hat{A}^*|$ sub-matrix of $B$ formed by columns, whose indices are in $\hat{A}^*$. Our combined estimator as a weighted average of $\hat{\beta}_k^{PWCQR}$, $k = 1, \ldots, K$, is

$$\hat{\beta}^{DC-PWCQR} = A\left(\sum_{k=1}^{K} A^\top \{\mathbf{X}_k^\top \mathbf{X}_k\} A\right)^{-1} \sum_{k=1}^{K} A^\top \{\mathbf{X}_k^\top \mathbf{X}_k\} A A^\top \hat{\beta}_k^{PWCQR}.$$

We show that the adaptive LASSO penalized WCQR estimator enjoys the oracle properties of the WCQR-oracle. For any indices set $S$, denote by $\hat{\beta}_S$ a $|S| \times 1$ vector formed by the elements of $\hat{\beta}$, whose indices are in $S$. The result allows $p$ to depend on the sample size $N$. To stress dependence on the sample size, we rewrite $p$ as $p_N$.

**Theorem 4.1** (Consistency). Under the same conditions as in Theorem 3.1. If $p_N^3 K/N \to 0$, $\lambda_{\max}\sqrt{N/K} \to 0$ as $N \to \infty$, then

$$\left\|\hat{\beta}^{DC-PWCQR} - \beta_0\right\| = O_p\left(\sqrt{p_N K/N}\right),$$

and $\hat{\beta}_{\bar{A}^*}^{DC-PWCQR} = \mathbf{0}$, where $\lambda_{\max} = \max\{\lambda_1, \ldots \lambda_K\}$ and $\bar{A}^*$ is the complement of the true zero set $\{j : \beta_{0,j} = 0\}$.

**Theorem 4.2** (Asymptotic normality). Under the same conditions as in Theorem 4.1. If $\lambda_{\min}\left(NK^{-1}p_N^{-1}\right)^{(\gamma+1)/2} \to \infty$, for some $\gamma > 0$ and $\min_{1 \leq j \leq A^*} |\beta_{0,j}|/(\lambda_{\max}\sqrt{N/K}) \to \infty$ as $N \to \infty$, then

$$\sqrt{N}\left(\hat{\beta}_{A^*}^{DC-PWCQR} - \beta_{0,A^*}\right) \xrightarrow{L} \mathcal{N}\left(0, \mathbf{C}_{A^*}^{-1}\left(\mathbf{f}^\top \Omega^{-2} \mathbf{f}\right)^{-1}\right),$$

where $\lambda_{\min} = \min\{\lambda_1, \ldots \lambda_K\}$.

**Remark 4.1.** The limiting distribution of $\hat{\beta}_{A^*}^{DC-PWCQR}$ in Theorem 4.2 is that of $\hat{\beta}_N$ in Theorem 6 from Jiang et al. (2012), where the entire data is analyzed. Thus, the DC-PWCQR estimator is asymptotically equivalent to the corresponding estimator using the full datasets.

**Remark 4.2.** For the penalized WCQR estimators, one has to select tuning parameters $\lambda_k, k = 1, \ldots, K$. Many selection criteria such as cross validation (CV), generalized cross validation (GCV), BIC and AIC selection can be used. Wang et al. (2007a) pointed out that the GCV approach tends to produce overfitted models even as the sample size goes to infinity. For this reason, Wang et al. (2007b) developed a BIC-type selection criterion, which motivates us to consider the following BIC criterion

$$BIC(\lambda_k) = \log\left(\sum_{q=1}^{Q}\sum_{i=1}^{n_k}\hat{w}_q\rho_{\tau_q}\left\{y_{k,i} - \hat{b}_q^{PWCQR} - \mathbf{x}_{k,i}^{\top}\hat{\beta}^{PWCQR}\right\}\right) + df_{\lambda_k}\log(n_k)/n_k,$$

where $\hat{b}_q^{PWCQR}$ is the minimizer of (4.1), and $df_{\lambda_k}$ is the number of nonzero coefficients in $\hat{\beta}^{PWCQR}$ as a simple estimate for the degrees of freedom, see Zou et al. (2007). We can select $\hat{\lambda}_k = \arg\min_{\lambda_k} BIC(\lambda_k)$, for $k = 1, \ldots, K$.

## 5. Numerical implementation

DC-CQR, DC-WCQR and DC-PWCQR estimations need to solve convex programming problems of (2.2), (3.1) and (4.1). Similar to idea in Wu and Liu (2009), we can modify the optimization problem (2.2) and (3.1) to the following constrained linear programming problem:

$$\begin{aligned}
&\min\sum_{q=1}^{Q}\sum_{i=1}^{N}\hat{w}_q\left(\tau_q\xi_{iq} + (1-\tau_q)\zeta_{iq}\right),\\
&subject\ to\ \xi_{iq} - \zeta_{iq} = y_i - b_q - \mathbf{x}_i^{\top}\beta,\\
&\xi_{iq} \geq 0,\ \zeta_{iq} \geq 0,\\
&q = 1, \ldots, Q,\ i = 1, \ldots, N,
\end{aligned} \tag{5.1}$$

when $\hat{\mathbf{w}}$ is the estimation of $\mathbf{w}_{opt}$, (5.1) is equal to (3.1), and (5.1) is equal to (2.2) under $\hat{\mathbf{w}} \equiv \mathbf{1}$.

Given fixed tuning parameter $\lambda_1, \ldots, \lambda_K$ and $\hat{\beta}^{DC-WCQR}$, the minimization problem (4.1) can also be casted into a constrained linear programming problem as follows:

$$\begin{aligned}
&\min\sum_{q=1}^{Q}\sum_{i=1}^{N}\hat{w}_q\left(\tau_q\xi_{iq} + (1-\tau_q)\zeta_{iq}\right) + \sum_{j=1}^{p}\frac{n_k\lambda_k}{|\hat{\beta}_j^{DC-WCQR}|^2}\varphi_j,\\
&subject\ to\ \xi_{iq} - \zeta_{iq} = y_i - b_q - \mathbf{x}_i^{\top}\beta,\\
&\xi_{iq} \geq 0,\ \zeta_{iq} \geq 0,\ q = 1, \ldots, Q,\ i = 1, \ldots, N,\\
&\varphi_j \geq \beta_j,\ \varphi_j \geq -\beta_j,\ j = 1, \ldots, p_N.
\end{aligned} \tag{5.2}$$

The linear programming problem (5.1) and (5.2) can be easily achieved by using optimization software, for example, optimization toolbox in Matlab.

## 6. Numerical studies

In this section, we first use Monte Carlo simulation studies to assess the finite sample performance of the proposed procedures and then demonstrate the application of the proposed methods with a real data analysis. Tian et al. (2016) proposed redefined AIC and BIC to select number of composite quantiles $Q$. However, the performances of CQR method with different $Q$ are very similar in their the simulation part. Moreover, Zou and Yuan (2008) recommended $Q = 19$ for linear model. Therefore, we choose $Q = 19$ as a compromise between estimation and computation efficiency of the CQR method and let the equally spaced quantile levels be $\tau_q = q/20, q = 1, \ldots, 19$. All programs are written in Matlab and our computer has a 2.4GHz Pentium processor and 4G memory.

### 6.1. *Example for estimation procedure*

In this section, we study the performances of DC-CQR and DC-WCQR method. Furthermore, we include four competitors in our comparison:

(1) the least square method (LS): $\hat{\beta}^{LS} = \left(\sum_{k=1}^{K} \mathbf{X}_k^\top \mathbf{X}_k\right)^{-1} \sum_{k=1}^{K} \mathbf{X}_k^\top \mathbf{Y}_k$, see Draper and Smith (1998);

(2) the least absolute deviation method (LAD): $\hat{\beta}^{LAD} = \frac{1}{K}\sum_{k=1}^{K} \hat{\beta}_k^{LAD}$, where $\hat{\beta}_k^{LAD} = \arg\min_\beta \sum_{i=1}^{n_k} \left|y_i - \mathbf{x}_i^\top \beta\right|$, $k = 1, \ldots, K$, see Xu et al (2017);

(3) it is noted that the CQR estimator degenerates to the median regression estimator when $Q = 1$ (DC-CQR$_1$);

(4) the aggregated CQR estimator with $Q = 19$ (ACQR$_{19}$): $\hat{\beta}^{ACQR_{19}} = \frac{1}{K}\sum_{k=1}^{K} \hat{\beta}_k^{CQR_{19}}$, see Zhao et al (2017).

We conduct a simulation study with $N = 100,000$ and the data are generated from model (2.1), where $\beta_0 = (\beta_{0,1}, \beta_{0,2}, \beta_{0,3}, \beta_{0,4}, \beta_{0,5})^\top = (-2, -1, 0, 1, 2)^\top$. In our simulation, we consider three error distributions for $\varepsilon$: standard normal distribution (N(0,1)), a t distribution with 3 degrees of freedom (t(3)) and a Chi-square distribution with 5 degrees of freedom ($\chi^2(5)$), and two cases of $\mathbf{X}$.

Case 1: $\mathbf{X}$ follows a multivariate normal distribution: $N(\mathbf{0}, \Sigma)$ with correlation matrix $\Sigma_{ij} = 0.5^{i-j}$ for $1 \le i, j \le 5$.

Case 2: $\mathbf{X}_i, i = 1, \ldots, 5$, follows a mixture distribution: $N(0,1)$, $N(0,2)$, $N(0,3)$, $N(0,4)$, $N(0,5)$, $U(0,1)$, $U(0,2)$, $U(0,3)$, $U(0,4)$, $U(0,5)$ and each distribution contains $N/10$ observations.

All of the simulations are run for 100 replicates. Three block numbers are considered: $K = 1, 10, 100$, and each subset contains equal observations. The bias, absolute bias and standard deviations of parameters are summarized in Table 1 and Table 2. From Tables 1 and 2, one can see that DC-CQR$_{19}$ and DC-WCQR$_{19}$ are close to the true value, because the bias and absolute bias are all very small. Furthermore, Tables 3-5 depict the root-mean squared errors (RMSE) and mean absolute deviation (MAD) of the estimate $\hat{\beta}$ to assess the accuracy of proposed methods comparing with others methods,

$$RMSE = \sqrt{\frac{1}{5}\sum_{j=1}^{5}(\hat{\beta}_j - \beta_{0j})^2}, \quad MAD = \frac{1}{5}\sum_{j=1}^{5}\left|\hat{\beta}_j - \beta_{0j}\right|.$$

From Tables 3-5, the following conclusions can be drawn:

(i) The LS is the most efficient and DC-WCQR$_{19}$ is very close to LS under standard normal error distribution. For other error distributions, DC-WCQR$_{19}$ is consistently superior to the other five methods.

(ii) The DC-CQR and DC-WCQR methods perform well when comparing the resulting estimator with entire data estimator.

(iii) One can see from Table 3-5 under Case 1 that the results of LAD are close to DC-CQR$_1$ and the results of ACQR$_{19}$ are close to DC-CQR$_{19}$. The reason is that under Case 1, $\lim_{n_1\to\infty} \mathbf{X}_1^\top \mathbf{X}_1/n_1 = \ldots = \lim_{n_K\to\infty} \mathbf{X}_K^\top \mathbf{X}_K/n_K$, thus

$$\hat{\beta}^{DC-CQR} = \left(\sum_{k=1}^{K} \mathbf{X}_k^\top \mathbf{X}_k\right)^{-1} \sum_{k=1}^{K} \mathbf{X}_k^\top \mathbf{X}_k \hat{\beta}_k^{CQR} \longrightarrow \frac{1}{K}\sum_{k=1}^{K} \hat{\beta}_k^{CQR}.$$

When $\mathbf{X}$ follows Case 2, DC-CQR$_1$ and DC-CQR$_{19}$ are better than LAD and ACQR$_{19}$, respectively.

### 6.2. *Example for confidence intervals and hypothesis testing*

In this example, we first study the coverage probability (C.P.) of the interval estimate and the length of the confidence interval (Length) for CQR and WCQR methods:

$$C.P. = P\left(\hat{\beta}_i - \frac{\hat{\sigma}_j}{\sqrt{N}}\Phi^{-1}(1-\alpha/2) \le \beta_{0,i} \le \hat{\beta}_i + \frac{\hat{\sigma}_j}{\sqrt{N}}\Phi^{-1}(1-\alpha/2)\right),$$
$$Length = \frac{\hat{\sigma}_j}{\sqrt{N}}\Phi^{-1}(1-\alpha/2),$$

where $\hat{\beta}_i$ and $\hat{\sigma}_j$ are defined in Section 2 and 3. Theoretically, C.P. is approximately equal to $1-\alpha$. For a given confidence level $1-\alpha$, Length depends on the standard error of the estimate, thus a shorter interval is preferred. All the settings are the same as in Example 6.1. Since the results are similar for all $\beta_{0,i}$'s, only the results on $\beta_{0,3} = 0$ is reported here. Table 6 lists the C.P. and Length ($\alpha = 0.05$) with 100 simulation runs for various error distributions. It can be seen that the C.P.s of CQR and WCQR are all around the nominal level (0.95), and the Lengths demonstrate WCQR is more efficient than CQR under all three noise conditions.

Next, we consider the test probability (T.P.) and P-value under $H_0 : \beta_{0,3} = 0$ for CQR and WCQR methods:

$$T.P. = P\left(\left|\sqrt{N}\hat{\beta}_j^{DC-CQR}/\hat{\sigma}_j\right| \le \Phi^{-1}(1-\alpha/2)\right) = C.P.,$$
$$P - value = 2P\left(\bar{X} > \left|\sqrt{N}\hat{\beta}_j^{DC-CQR}/\hat{\sigma}_j\right|\right),$$

where $\bar{X}$ is a random variable which follows a standard normal distribution. Theoretically, T.P. is approximately equal to $1-\alpha$. For a significance level $\alpha$, because of $\beta_{0,3} = 0$, under $H_0 : \beta_{0,3} = 0$, the P-value should be larger than $\alpha$. Table 4 also lists the T.P. ($\alpha = 0.05$) and P-value with 100 simulation runs for various error distributions. It can be seen that the T.P.s of CQR and WCQR are all around the nominal level (0.95), and P-values are all larger than $\alpha$, thus we should accept the original hypothesis $H_0$.

### 6.3. Example for variable selection

In this example, we consider the model (2.1) with $\beta_0 = (-2, -1, 0, 1, 2, 0, 0, 0, 0, 0)^\top$, and the covariate vector $\mathbf{X}$ follows a multivariate normal distribution: $N(\mathbf{0}, \Sigma)$ with correlation matrix $\Sigma_{ij} = 0.5^{i-j}$ for $1 \le i, j \le 10$. Other settings are the same as those in Example 6.1.

To assess the performance of variable selection procedures for the parametric component, we consider the generalized mean square error (GMSE), as defined in Kai et al. (2011).

$$GMSE(\hat{\beta}) = (\hat{\beta} - \beta_0)^\top E(\mathbf{X}^\top \mathbf{X})(\hat{\beta} - \beta_0).$$

For each procedure, we calculate the relative GMSE (RGMSE), which is defined to be the ratio of GMSE of a selected final model to that of the un-penalized estimate under the full model.

$$RGMSE = GMSE(\hat{\beta}^{DC-WCQR})/GMSE(\hat{\beta}^{DC-PWCQR}).$$

In addition, we calculated model selection sensitivity (the number of truly selected variables divided by the true model size) and model selection specificity (the number of truly removed variables divided by the number of noise variables). The simulation results are shown in Table 7. According to $RGMSE$ in Table 7, the DC-PWCQR estimators performed better than DC-WCQR estimators. In all cases, the DC-PWCQR estimators had good model selection results with high model selection sensitivity and specificity that were similar to those of the penalized estimators analyzing the full dataset.

### 6.4. Real data example: airline on-time data

The airline on-time performance data from the 2009 ASA Data Expo (http://stat-computing.org/dataexpo/2009/the-data.html) is used as a case study. The data is publicly available and has been used for demonstration with massive data by Wang et al. (2015) and Schifano et al. (2016). It consists of flight arrival and departure details for all commercial flights within the USA, from October 1987 to April 2008. About 12 million flights were recorded with 29 variables. Due to the over long computing time, only the data of 2007 for $N = 7,275,288$ observations with complete data is considered.

We considered arrival delay (ArrDelay) as a continuous variable by modeling $\log(ArrDelay - \min(ArrDelay) + 1)$, denoted as $Y$, as a linear function of departure hour (rang 0 to 24), distance (in 1000 miles), night flight (1 if departure between 8 p.m. and 5 a.m., 0 otherwise), and weekend flight (1 if departure occurred during the weekend, 0 otherwise). This model was also studied by Schifano et al. (2016).

We estimate the above regression model using LS, LAD, DC-CQR$_1$, ACQR$_{19}$, DC-CQR$_{19}$, DC-WCQR$_{19}$ and DC-PWCQR$_{19}$ methods. For the purpose of comparison, we evaluate the performance of these estimators based on their out-of-sample prediction. In particular, we estimate the above regression model based on 5,108,211 data. We use a subset size of $n_k = 510,821$ for $k = 1, \ldots, 9$, and $n_{10} = 510,822$. Then, we use the estimated coefficients to construct forecast of the other 2,167,077 data. We compare both the mean squared prediction error (MSE) and the mean absolute deviation (MAD) of the predictions,

$$MSE = \frac{1}{n}\sum_i (Y_i - \hat{Y}_i)^2, MAD = \frac{1}{n}\sum_i |Y_i - \hat{Y}_i|,$$

where $\hat{Y}_i$ is the fitted value of $Y_i$, $i = 1, \ldots, n$, and here $n = 2,167,077$. The MSE and MAD of seven methods are given in Table 8. DC-PWCQR$_{19}$ method shows that the coefficient of departure hour should be zero. The results of MSE are very close to different methods. The results of MAD show that the performs of DC-PWCQR$_{19}$ method are the best except LAD.

## 7. Conclusion

In the linear model, we developed a divide-and-conquer composite quantile regression method for massive data sets. The proposed approach significantly reduces the required amount of primary memory, and the resulting estimates were as efficient as if the entire data set was analyzed simultaneously. Moreover, to improve the efficiency of CQR, we proposed a weighted CQR estimation approach. To achieve sparsity with high-dimensional covariates, we provided a variable selection procedure to select significant parametric components.

The methods in this article were designed for small to moderate covariate dimensionality $p$, but large $N$. The use of de-biasing technique for ultra-high dimensional data ($p$ is larger than $N$) is an interesting consideration, see Van de Geer et al.(2014) and Zhao et al. (2015).


## Acknowledgments

This research was supported by the Fundamental Research Funds for the Central Universities (Series number: 2232016D3-17). Additional partial support is also provided by the Shanghai Sailing Program (Series number: 17YF1400800); the National Natural Science Foundation of China (Series number: 11626057); and the China Scholarship Council (Series number: 201706635019).


## Appendix

**Proof of Theorem 2.1.** Let $\sqrt{n_k}(\hat{\beta}_k^{CQR} - \beta_0) = \mathbf{u}_{n_k}$ and $\sqrt{n_k}(\hat{b}_{q,k}^{CQR} - b_{\tau_q}) = v_{n_k,q}$, where $\hat{b}_{q,k}^{CQR}$ is the minimizer of (2.2) for the $k$th subset. Then $(v_{n_k,1}, \ldots, v_{n_k,Q}, \mathbf{u}_{n_k})$, $k = 1, \ldots, K$, are the minimizer of the following criterion:

$$L_{n_k} = \sum_{q=1}^{Q}\sum_{i=1}^{n_k}\left\{\rho_{\tau_q}\left(\varepsilon_i - b_{\tau_q} - [v_q + x_{k,i}^\top \mathbf{u}]/\sqrt{n_k}\right) - \rho_{\tau_q}\left(\varepsilon_i - b_{\tau_q}\right)\right\}, k = 1, \ldots, K,$$

To apply the identity (Knight, 1998)

$$\rho_\tau(x - y) - \rho_\tau(x) = y\{I(x < 0) - \tau\} + \int_0^y \{I(x \le z) - I(x \le 0)\}dz.$$

Thus, we rewrite $L_{n_k}$ as follows:

$$\begin{aligned} L_{n_k} &= \sum_{q=1}^{Q}\sum_{i=1}^{n_k}\frac{v_q + x_{k,i}^\top \mathbf{u}}{\sqrt{n_k}}[I(\varepsilon_i < b_{\tau_q}) - \tau_q] \\ &+ \sum_{q=1}^{Q}\sum_{i=1}^{n_k}\int_0^{[v_q + x_{k,i}^\top \mathbf{u}]/\sqrt{n_k}}\left[I(\varepsilon_i \le b_{\tau_q} + t) - I(\varepsilon_i \le b_{\tau_q})\right]dt \\ &\equiv \sum_{q=1}^{Q}\sum_{i=1}^{n_k}\frac{v_q + x_{k,i}^\top \mathbf{u}}{\sqrt{n_k}}[I(\varepsilon_i < b_{\tau_q}) - \tau_q] + \sum_{q=1}^{Q} B_{n_k}^{(q)}, \end{aligned}$$

where $B_{n_k}^{(q)} = \sum_{i=1}^{n_k} \int_0^{[v_q + x_{k,i}^\top \mathbf{u}]/\sqrt{n_k}} [I(\varepsilon_i \le b_{\tau_q} + t) - I(\varepsilon_i \le b_{\tau_q})]dt$. Then, we have

$$\begin{aligned}
E[B_{n_k}^{(q)}] &= \sum_{i=1}^{n_k} \int_0^{[v_q + x_{k,i}^\top \mathbf{u}]/\sqrt{n_k}} [F(b_{\tau_q} + t) - F(b_{\tau_q})]dt \\
&= \frac{1}{n_k} \sum_{i=1}^{n_k} \int_0^{[v_q + x_{k,i}^\top \mathbf{u}]} \sqrt{n_k}[F(b_{\tau_q} + t/\sqrt{n_k}) - F(b_{\tau_q})]dt \\
&\to \frac{1}{2} f(b_{\tau_q})(v_q, \mathbf{u}^\top) \begin{bmatrix} 1 & 0 \\ 0 & \mathbf{C}_k \end{bmatrix} (v_q, \mathbf{u}^\top)^\top. \\
Var[B_{n_k}^{(q)}] &= \sum_{i=1}^{n_k} E\left( \int_0^{[v_q + x_{k,i}^\top \mathbf{u}]/\sqrt{n_k}} ([I(\varepsilon_i \le b_{\tau_q} + t) - I(\varepsilon_i \le b_{\tau_q})] - [F(b_{\tau_q} + t) - F(b_{\tau_q})])dt \right)^2 \\
&\le \sum_{i=1}^{n_k} E\left[ \left| \int_0^{[v_q + x_{k,i}^\top \mathbf{u}]/\sqrt{n_k}} ([I(\varepsilon_i \le b_{\tau_q} + t) - I(\varepsilon_i \le b_{\tau_q})] - [F(b_{\tau_q} + t) - F(b_{\tau_q})])dt \right| \right] \\
&\quad \times 2 \left| \frac{v_q + x_{k,i}^\top \mathbf{u}}{\sqrt{n_k}} \right| \le 4E[B_{n_k}^{(q)}] \frac{\max_{1 \le i \le n_k} |v_q + x_{k,i}^\top \mathbf{u}|}{\sqrt{n_k}} \to 0.
\end{aligned}$$

Hence, $B_{n_k}^{(q)} \xrightarrow{P} \frac{1}{2} f(b_{\tau_q})(v_q, \mathbf{u}^\top) \begin{bmatrix} 1 & 0 \\ 0 & \mathbf{C}_k \end{bmatrix} (v_q, \mathbf{u}^\top)^\top$. Thus it follows that

$$L_{n_k} \xrightarrow{P} \sum_{q=1}^{Q} \sum_{i=1}^{n_k} \frac{v_q + x_{k,i}^\top \mathbf{u}}{\sqrt{n_k}} [I(\varepsilon_i < b_{\tau_q}) - \tau_q] + \frac{1}{2} \sum_{q=1}^{Q} f(b_{\tau_q})(v_q, \mathbf{u}^\top) \begin{bmatrix} 1 & 0 \\ 0 & \mathbf{C} \end{bmatrix} (v_q, \mathbf{u}^\top)^\top.$$

Since $L_{n_k}$ is a convex function, then following Knight (1998) and Koenker (2005), we have

$$\mathbf{C}_k \cdot \sqrt{n_k} \left( \hat{\beta}_k^{CQR} - \beta_0 \right) = \left[ \sum_{q=1}^{Q} f(b_{\tau_q}) \right]^{-1} \cdot \frac{1}{\sqrt{n_k}} \sum_{i=1}^{n_k} x_{k,i}^\top \sum_{q=1}^{Q} [I(\varepsilon_i < b_{\tau_k}) - \tau_k] + R_{n_k},$$

where $\mathbf{C}_k = \lim_{n_k \to \infty} \frac{1}{n_k} \mathbf{X}_k^\top \mathbf{X}_k$, and following Bahadur (1966), we can obtain $O\left(n_k^{-1/4} (\log\log n_k)^{3/4}\right)$. Since $\mathbf{X}_i$ is independent for $i = 1, \ldots, N$, $\mathbf{X}^\top \mathbf{X} = \sum_{k=1}^{K} \mathbf{X}_k^\top \mathbf{X}_k$ and $K = o\left(n_{\min}^{1/2} (\log\log n_{\min})^{-3/2}\right)$, we can obtain

$$\frac{1}{\sqrt{N}} \sum_{k=1}^{K} \mathbf{X}_k^\top \mathbf{X}_k \left( \hat{\beta}_k^{CQR} - \beta_0 \right) \xrightarrow{L} \mathcal{N}(0, \mathbf{C} R_1).$$

Therefore, using the DC-CQR estimator, we have

$$\begin{aligned}
\sqrt{N} \left( \hat{\beta}_k^{DC-CQR} - \beta_0 \right) &= \sqrt{N} \left( \sum_{k=1}^{K} \mathbf{X}_k^\top \mathbf{X}_k \right)^{-1} \sum_{k=1}^{K} \mathbf{X}_k^\top \mathbf{X}_k \left( \hat{\beta}_k^{CQR} - \beta_0 \right) \\
&= \left( \frac{\sum_{k=1}^{K} \mathbf{X}_k^\top \mathbf{X}_k}{N} \right)^{-1} \cdot \frac{1}{\sqrt{N}} \sum_{k=1}^{K} \mathbf{X}_k^\top \mathbf{X}_k \left( \hat{\beta}_k^{CQR} - \beta_0 \right) \\
&= \left( \frac{\mathbf{X}^\top \mathbf{X}}{N} \right)^{-1} \cdot \frac{1}{\sqrt{N}} \sum_{k=1}^{K} \mathbf{X}_k^\top \mathbf{X}_k \left( \hat{\beta}_k^{CQR} - \beta_0 \right) \xrightarrow{L} \mathcal{N}\left(0, \mathbf{C}^{-1} R_1\right).
\end{aligned}$$

This completes the proof.

**Proof of Theorem 3.1.** Let $\sqrt{n_k}(\hat{\beta}_k^{WCQR} - \beta_0) = \mathbf{u}_{n_k}^*$ and $\sqrt{n_k}(\hat{b}_{q,k}^{WCQR} - b_{\tau_q}) = v_{n_k,q}^*$, where $\hat{b}_{q,k}^{CQR}$ is the minimizer of (3.1) for the $k$th subset. Then $(v_{n_k,1}^*, \ldots, v_{n_k,Q}^*, \mathbf{u}_{n_k}^*)$, $k = 1, \ldots, K$, are the minimizer of the following criterion:

$$L_{n_k}^* = \sum_{q=1}^{Q} \sum_{i=1}^{n_k} \hat{w}_q \left[ \rho_{\tau_q} \left( \varepsilon_i - b_{\tau_q} - [v_q^* + x_{k,i}^\top \mathbf{u}^*]/\sqrt{n_k} \right) - \rho_{\tau_q} \left( \varepsilon_i - b_{\tau_q} \right) \right], k = 1, \ldots, K.$$

We rewrite $L_{n_k}$ as follows:

$$\begin{aligned}
L_{n_k}^* &= \sum_{q=1}^{Q}\sum_{i=1}^{n_k}\frac{v_q^* + x_{k,i}^\top \mathbf{u}^*}{\sqrt{n_k}}\hat{w}_q[I(\varepsilon_i < b_{\tau_q}) - \tau_q] \\
&\quad + \sum_{q=1}^{Q}\hat{w}_q\sum_{i=1}^{n_k}\int_0^{[v_q + x_{k,i}^\top \mathbf{u}^*]/\sqrt{n_k}}\left[I(\varepsilon_i \le b_{\tau_q} + t) - I(\varepsilon_i \le b_{\tau_q})\right]dt \\
&\equiv \sum_{q=1}^{Q}\sum_{i=1}^{n_k}\frac{v_q^* + x_{k,i}^\top \mathbf{u}^*}{\sqrt{n_k}}\hat{w}_q[I(\varepsilon_i < b_{\tau_q}) - \tau_q] + \sum_{q=1}^{Q}\hat{w}_q B_{n_k}^{*(q)},
\end{aligned}$$

where$B_{n_k}^{*(q)} = \sum_{i=1}^{n_k}\int_0^{[v_q^* + x_{k,i}^\top \mathbf{u}^*]/\sqrt{n_k}}[I(\varepsilon_i \le b_{\tau_q} + t) - I(\varepsilon_i \le b_{\tau_q})]dt$. Then, we have

$$\begin{aligned}
E[B_{n_k}^{*(q)}] &= \sum_{i=1}^{n_k}\int_0^{[v_q^* + x_{k,i}^\top \mathbf{u}^*]/\sqrt{n_k}}[F(b_{\tau_q} + t) - F(b_{\tau_q})]dt \\
&= \frac{1}{n_k}\sum_{i=1}^{n_k}\int_0^{[v_q + x_{k,i}^\top \mathbf{u}^*]}\sqrt{n_k}[F(b_{\tau_q} + t/\sqrt{n_k}) - F(b_{\tau_q})]dt \\
&\to \frac{1}{2}f(b_{\tau_q})(v_q^*, \mathbf{u}^{*\top})\begin{bmatrix} 1 & 0 \\ 0 & \mathbf{C}_k \end{bmatrix}(v_q^*, \mathbf{u}^{*\top})^\top. \\
Var[B_{n_k}^{*(q)}] &= \sum_{i=1}^{n_k}E\left(\int_0^{[v_q^* + x_{k,i}^\top \mathbf{u}^*]/\sqrt{n_k}}([I(\varepsilon_i \le b_{\tau_q} + t) - I(\varepsilon_i \le b_{\tau_q})] - [F(b_{\tau_q} + t) - F(b_{\tau_q})])dt\right)^2 \\
&\le \sum_{i=1}^{n_k}E\left[\left|\int_0^{[v_q^* + x_{k,i}^\top \mathbf{u}^*]/\sqrt{n_k}}([I(\varepsilon_i \le b_{\tau_q} + t) - I(\varepsilon_i \le b_{\tau_q})] - [F(b_{\tau_q} + t) - F(b_{\tau_q})])dt\right|\right] \\
&\quad \times 2\left|\frac{v_q^* + x_{k,i}^\top \mathbf{u}^*}{\sqrt{n_k}}\right| \le 4E[B_{n_k}^{*(q)}]\frac{\max_{1\le i\le n_k}|v_q^* + x_{k,i}^\top \mathbf{u}^*|}{\sqrt{n_k}} \to 0.
\end{aligned}$$

Hence, $B_{n_k}^{*(q)} \xrightarrow{P} \frac{1}{2}f(b_{\tau_q})(v_q^*, \mathbf{u}^{*\top})\begin{bmatrix} 1 & 0 \\ 0 & \mathbf{C}_k \end{bmatrix}(v_q^*, \mathbf{u}^{*\top})^\top$. Thus it follows that

$$L_{n_k}^* \xrightarrow{P} \sum_{q=1}^{Q}\sum_{i=1}^{n_k}\frac{v_q^* + x_{k,i}^\top \mathbf{u}^*}{\sqrt{n_k}}\hat{w}_q[I(\varepsilon_i < b_{\tau_q}) - \tau_q] + \frac{1}{2}\sum_{q=1}^{Q}\hat{w}_q f(b_{\tau_q})(v_q^*, \mathbf{u}^{*\top})\begin{bmatrix} 1 & 0 \\ 0 & \mathbf{C}_k \end{bmatrix}(v_q^*, \mathbf{u}^{*\top})^\top.$$

Since $L_{n_k}^*$ is a convex function, then following Knight (1998), Koenker (2005) and Bahadur (1966), we have

$$\begin{aligned}
\mathbf{C}_k \cdot \sqrt{n_k}\left(\hat{\beta}_k^{WCQR} - \beta_0\right) = \left[\sum_{q=1}^{Q}\hat{w}_q f(b_{\tau_q})\right]^{-1} \cdot \frac{1}{\sqrt{n_k}}\sum_{i=1}^{n_k}x_{k,i}^\top\sum_{q=1}^{Q}\hat{w}_q\left[I(\varepsilon_i < b_{\tau_k}) - \tau_k\right] \\
+ O\left(n_k^{-1/4}(\log\log n_k)^{3/4}\right).
\end{aligned}$$

Therefore, using the DC-WCQR estimator, we have

$$\begin{aligned}
\sqrt{N}\left(\hat{\beta}_k^{DC-WCQR} - \beta_0\right) &= \sqrt{N}\left(\sum_{k=1}^{K}\mathbf{X}_k^\top\mathbf{X}_k\right)^{-1}\sum_{k=1}^{K}\mathbf{X}_k^\top\mathbf{X}_k\left(\hat{\beta}_k^{WCQR} - \beta_0\right) \\
&= \left(\frac{\sum_{k=1}^{K}\mathbf{X}_k^\top\mathbf{X}_k}{N}\right)^{-1} \cdot \frac{1}{\sqrt{N}}\sum_{k=1}^{K}\mathbf{X}_k^\top\mathbf{X}_k\left(\hat{\beta}_k^{WCQR} - \beta_0\right) \\
&= \left(\frac{\mathbf{X}^\top\mathbf{X}}{N}\right)^{-1} \cdot \frac{1}{\sqrt{N}}\sum_{k=1}^{K}\mathbf{X}_k^\top\mathbf{X}_k\left(\hat{\beta}_k^{WCQR} - \beta_0\right) \xrightarrow{L} \mathcal{N}\left(0, \mathbf{C}^{-1}\left(\mathbf{f}^\top\Omega^{-2}\mathbf{f}\right)^{-1}\right).
\end{aligned}$$

This completes the proof.

**Proof of Theorem 4.1.** From Theorem 5 of Jiang et al. (2012), for each subsets $n_k, k = 1, \ldots, K$, we have

$$\left\|\hat{\beta}_k^{PWCQR} - \beta_0\right\| = O_p\left(\sqrt{p_N/n_k}\right),$$

and by the condition $n_k = O(N/K)$, we can obtain

$$\left\|\hat{\beta}_k^{PWCQR} - \beta_0\right\| = O_p\left(\sqrt{p_N K/N}\right).$$

When $N$ is large enough, we have $\hat{A}_k = A^*$ for all subsets and $\hat{A}^* = A^*$. In this case, $\hat{\beta}_{\bar{A}^*}^{DC-PWCQR} = \mathbf{0}$ then

$$\begin{aligned}\left\|\hat{\beta}^{DC-PWCQR} - \beta_0\right\| &= \left\|A\left(\sum_{k=1}^{K} A^\top\{\mathbf{X}_k^\top \mathbf{X}_k\}A\right)^{-1} \sum_{k=1}^{K} A^\top\{\mathbf{X}_k^\top \mathbf{X}_k\}AA^\top\left(\hat{\beta}_k^{PWCQR} - \beta_0\right)\right\| \\ &\leq O_p\left(\sqrt{p_N K/N}\right).\end{aligned}$$

**Proof of Theorem 5.1.** By the form of $\hat{\beta}^{DC-PWCQR}$, $\hat{A}_k = A^*$ for all subsets and $\hat{A}^* = A^*$, we have

$$\hat{\beta}_{A^*}^{DC-PWCQR} = \left(\sum_{k=1}^{K} \mathbf{X}_{k,A^*}^\top \mathbf{X}_{k,A^*}\right)^{-1} \sum_{k=1}^{K} \mathbf{X}_{k,A^*}^\top \mathbf{X}_{k,A^*} \hat{\beta}_{k,A^*}^{PWCQR}.$$

By Theorem 6 of Jiang et al. (2012) and by using the techniques of the proof of Theorem 3.1, we can proof the Theorem 5.1.

Table 1: The average bias (standard deviation) of coefficients. ($10^{-2}$)

| $\varepsilon$ | **X** | K | Methods | $\beta_{0,1}$ | $\beta_{0,2}$ | $\beta_{0,3}$ | $\beta_{0,4}$ | $\beta_{0,5}$ |
|---|---|---|---|---|---|---|---|---|
| N(0,1) | Case 1 | K=1 | DC-CQR$_{19}$ | 0.0507 (0.3691) | -0.0937 (0.4126) | 0.1690 (0.3483) | 0.1204 (0.3028) | -0.1507 (0.3596) |
| | | | DC-WCQR$_{19}$ | 0.0633 (0.3843) | -0.1213 (0.3947) | 0.1607 (0.3281) | 0.0996 (0.3272) | -0.1553 (0.3745) |
| | | K=10 | DC-CQR$_{19}$ | 0.1075 (0.2146) | 0.0732 (0.3447) | 0.0029 (0.4960) | 0.0535 (0.2220) | -0.0519 (0.2657) |
| | | | DC-WCQR$_{19}$ | 0.0701 (0.2008) | 0.0836 (0.3682) | -0.0035 (0.5913) | 0.0771 (0.2223) | -0.0694 (0.2864) |
| | | K=100 | DC-CQR$_{19}$ | -0.0562 (0.3720) | 0.1467 (0.4864) | -0.0521 (0.4633) | 0.1392 (0.3490) | -0.1066 (0.3348) |
| | | | DC-WCQR$_{19}$ | -0.0411 (0.3791) | 0.1181 (0.4843) | -0.0247 (0.4227) | 0.0801 (0.3697) | -0.0869 (0.3408) |
| | Case 2 | K=1 | DC-CQR$_{19}$ | 0.0906 (0.3476) | -0.0824 (0.2440) | 0.0237 (0.1498) | -0.0435 (0.2207) | -0.0142 (0.1490) |
| | | | DC-WCQR$_{19}$ | 0.1083 (0.3557) | -0.1005 (0.2514) | -0.0022 (0.1487) | -0.0103 (0.1977) | -0.0313 (0.1402) |
| | | K=10 | DC-CQR$_{19}$ | 0.0530 (0.2493) | 0.1792 (0.3162) | 0.0344 (0.2823) | 0.1771 (0.1068) | -0.0424 (0.3776) |
| | | | DC-WCQR$_{19}$ | -0.0062 (0.2131) | 0.1441 (0.2636) | 0.0031 (0.2138) | 0.1285 (0.0890) | -0.0743 (0.3337) |
| | | K=100 | DC-CQR$_{19}$ | 0.1196 (0.3021) | 0.0181 (0.2366) | 0.1062 (0.2537) | 0.1124 (0.2457) | -0.0525 (0.1355) |
| | | | DC-WCQR$_{19}$ | 0.0575 (0.2666) | -0.0418 (0.1870) | 0.0432 (0.2363) | 0.0479 (0.1947) | -0.0861 (0.0808) |
| t(3) | Case 1 | K=1 | DC-CQR$_{19}$ | 0.0999 (0.4197) | -0.1275 (0.5426) | 0.1550 (0.2697) | -0.1239 (0.3644) | 0.0651 (0.5703) |
| | | | DC-WCQR$_{19}$ | 0.1163 (0.4369) | -0.1309 (0.5035) | 0.1354 (0.2975) | -0.1149 (0.3594) | 0.0859 (0.5143) |
| | | K=10 | DC-CQR$_{19}$ | 0.0222 (0.4155) | 0.0394 (0.3826) | 0.3391 (0.3802) | -0.2128 (0.5193) | -0.0866 (0.6786) |
| | | | DC-WCQR$_{19}$ | -0.0140 (0.3734) | 0.0882 (0.3337) | 0.2245 (0.3912) | -0.1314 (0.5580) | -0.1017 (0.6750) |
| | | K=100 | DC-CQR$_{19}$ | -0.1446 (0.6234) | -0.1579 (0.6916) | 0.1658 (0.5568) | 0.0631 (0.4875) | -0.0532 (0.3585) |
| | | | DC-WCQR$_{19}$ | -0.1499 (0.5561) | -0.1234 (0.6059) | 0.1317 (0.4984) | 0.0631 (0.5281) | -0.0366 (0.3973) |
| | Case 2 | K=1 | DC-CQR$_{19}$ | 0.0062 (0.2837) | -0.0309 (0.2472) | 0.0056 (0.2012) | -0.0827 (0.2014) | 0.0086 (0.2921) |
| | | | DC-WCQR$_{19}$ | 0.0383 (0.2718) | -0.0267 (0.2407) | -0.0031 (0.2039) | -0.1104 (0.1554) | 0.0018 (0.2779) |
| | | K=10 | DC-CQR$_{19}$ | -0.2371 (0.3156) | 0.0278 (0.4198) | -0.1019 (0.3707) | -0.1795 (0.2839) | -0.0407 (0.3637) |
| | | | DC-WCQR$_{19}$ | -0.1461 (0.2116) | 0.0871 (0.3279) | -0.0384 (0.2961) | -0.0537 (0.3166) | 0.0402 (0.2796) |
| | | K=100 | DC-CQR$_{19}$ | -0.5647 (0.2239) | -0.4681 (0.3424) | -0.5673 (0.3311) | -0.4232 (0.2615) | -0.3951 (0.3831) |
| | | | DC-WCQR$_{19}$ | -0.2410 (0.2231) | -0.1231 (0.3202) | -0.2695 (0.3065) | -0.0916 (0.2221) | -0.0259 (0.2791) |
| $\chi^2(5)$ | Case 1 | K=1 | DC-CQR$_{19}$ | 0.0023 (0.8998) | 0.2578 (1.1780) | -0.1102 (1.2520) | 0.1129 (0.9846) | -0.4624 (1.1949) |
| | | | DC-WCQR$_{19}$ | -0.1708 (0.7041) | 0.1627 (0.8415) | -0.0267 (0.5486) | 0.2247 (0.7281) | -0.5244 (1.1771) |
| | | K=10 | DC-CQR$_{19}$ | 0.1413 (0.7054) | -0.3281 (0.7803) | 0.0588 (1.0212) | 0.2581 (1.0325) | -0.1482 (1.3243) |
| | | | DC-WCQR$_{19}$ | 0.4099 (0.6574) | -0.2484 (0.8631) | -0.0114 (0.5191) | 0.2100 (0.7918) | 0.1625 (0.6933) |
| | | K=100 | DC-CQR$_{19}$ | -0.1119 (1.0410) | 0.0889 (0.9194) | -0.3875 (1.1497) | -0.2194 (0.8035) | -0.0275 (0.9936) |
| | | | DC-WCQR$_{19}$ | -0.1237 (0.6555) | 0.0357 (1.0025) | -0.3008 (1.0426) | 0.1429 (0.7106) | -0.0428 (0.9181) |
| | Case 2 | K=1 | DC-CQR$_{19}$ | -0.2228 (0.2986) | 0.0307 (0.4965) | 0.2576 (0.5331) | -0.0459 (0.4782) | -0.2346 (0.4632) |
| | | | DC-WCQR$_{19}$ | -0.1429 (0.4885) | 0.0759 (0.3751) | 0.1706 (0.4112) | -0.2486 (0.4227) | -0.0389 (0.3446) |
| | | K=10 | DC-CQR$_{19}$ | 0.0143 (0.8062) | 0.1418 (1.0016) | 0.1078 (0.8279) | 0.0807 (0.9426) | 0.1216 (1.1651) |
| | | | DC-WCQR$_{19}$ | -0.0653 (0.3995) | 0.0926 (0.5791) | 0.0233 (0.3316) | -0.0208 (0.5100) | 0.0855 (0.5751) |
| | | K=100 | DC-CQR$_{19}$ | 0.6875 (0.7310) | 0.1316 (0.5766) | -0.0255 (0.4717) | 0.4189 (0.5127) | 0.2956 (0.7259) |
| | | | DC-WCQR$_{19}$ | 0.3191 (0.4351) | -0.0092 (0.2806) | -0.1741 (0.3507) | 0.2275 (0.4356) | 0.1336 (0.5231) |

Table 2: The average absolute bias (standard deviation) of coefficients. ($10^{-2}$)

| $\varepsilon$ | **X** | K | Methods | $\beta_{0,1}$ | $\beta_{0,2}$ | $\beta_{0,3}$ | $\beta_{0,4}$ | $\beta_{0,5}$ |
|---|---|---|---|---|---|---|---|---|
| N(0,1) | Case 1 | K=1 | DC-CQR$_{19}$ | 0.2632 (0.2492) | 0.2849 (0.2996) | 0.2928 (0.2403) | 0.2021 (0.2498) | 0.2869 (0.2512) |
| | | | DC-WCQR$_{19}$ | 0.2827 (0.2516) | 0.3076 (0.2589) | 0.2517 (0.2567) | 0.2150 (0.2583) | 0.3257 (0.2217) |
| | | K=10 | DC-CQR$_{19}$ | 0.2019 (0.1167) | 0.2765 (0.1995) | 0.3947 (0.2701) | 0.1630 (0.1515) | 0.1727 (0.2012) |
| | | | DC-WCQR$_{19}$ | 0.1614 (0.1297) | 0.3124 (0.1870) | 0.4729 (0.3181) | 0.1964 (0.1148) | 0.1837 (0.2234) |
| | | K=100 | DC-CQR$_{19}$ | 0.2796 (0.2437) | 0.3774 (0.3305) | 0.3466 (0.3016) | 0.2647 (0.2616) | 0.2811 (0.2021) |
| | | | DC-WCQR$_{19}$ | 0.2872 (0.2422) | 0.3857 (0.3043) | 0.3027 (0.2879) | 0.2816 (0.2449) | 0.2978 (0.1755) |
| | Case 2 | K=1 | DC-CQR$_{19}$ | 0.2611 (0.2328) | 0.1820 (0.1739) | 0.1078 (0.1008) | 0.1707 (0.1358) | 0.1251 (0.0710) |
| | | | DC-WCQR$_{19}$ | 0.2706 (0.2413) | 0.1918 (0.1832) | 0.1221 (0.0746) | 0.1463 (0.1241) | 0.1066 (0.0902) |
| | | K=10 | DC-CQR$_{19}$ | 0.1848 (0.1653) | 0.2376 (0.2701) | 0.2181 (0.1678) | 0.1771 (0.1068) | 0.3059 (0.2014) |
| | | | DC-WCQR$_{19}$ | 0.1758 (0.1054) | 0.2192 (0.1979) | 0.1750 (0.1080) | 0.1358 (0.0760) | 0.2809 (0.1727) |
| | | K=100 | DC-CQR$_{19}$ | 0.2467 (0.1988) | 0.2018 (0.1052) | 0.2220 (0.1487) | 0.2191 (0.1451) | 0.1046 (0.0962) |
| | | | DC-WCQR$_{19}$ | 0.2340 (0.1179) | 0.1615 (0.0891) | 0.1957 (0.1240) | 0.1657 (0.0998) | 0.0872 (0.0796) |
| t(3) | Case 1 | K=1 | DC-CQR$_{19}$ | 0.3059 (0.2886) | 0.4344 (0.3207) | 0.2348 (0.1954) | 0.2742 (0.2574) | 0.4578 (0.3116) |
| | | | DC-WCQR$_{19}$ | 0.3255 (0.2969) | 0.3986 (0.3098) | 0.2439 (0.2069) | 0.2874 (0.2282) | 0.3954 (0.3147) |
| | | K=10 | DC-CQR$_{19}$ | 0.2952 (0.2763) | 0.2992 (0.2205) | 0.3950 (0.3147) | 0.4456 (0.3153) | 0.5019 (0.4347) |
| | | | DC-WCQR$_{19}$ | 0.2812 (0.2276) | 0.2786 (0.1837) | 0.3557 (0.2617) | 0.4583 (0.3117) | 0.4764 (0.4636) |
| | | K=100 | DC-CQR$_{19}$ | 0.5439 (0.2884) | 0.6013 (0.3230) | 0.4583 (0.3275) | 0.3172 (0.3609) | 0.3060 (0.1663) |
| | | | DC-WCQR$_{19}$ | 0.4824 (0.2750) | 0.5099 (0.3085) | 0.4134 (0.2790) | 0.3339 (0.3993) | 0.3181 (0.2165) |
| | Case 2 | K=1 | DC-CQR$_{19}$ | 0.2190 (0.1651) | 0.1797 (0.1621) | 0.1702 (0.0912) | 0.1760 (0.1172) | 0.1923 (0.2105) |
| | | | DC-WCQR$_{19}$ | 0.2037 (0.1715) | 0.1668 (0.1668) | 0.1723 (0.0927) | 0.1468 (0.1172) | 0.2005 (0.1805) |
| | | K=10 | DC-CQR$_{19}$ | 0.2892 (0.2629) | 0.3698 (0.1586) | 0.3075 (0.2094) | 0.2898 (0.1518) | 0.3127 (0.1596) |
| | | | DC-WCQR$_{19}$ | 0.1813 (0.1788) | 0.2726 (0.1828) | 0.2331 (0.1702) | 0.2772 (0.1343) | 0.2207 (0.1608) |
| | | K=100 | DC-CQR$_{19}$ | 0.5647 (0.2239) | 0.5088 (0.2704) | 0.5673 (0.3311) | 0.4232 (0.2615) | 0.4610 (0.2899) |
| | | | DC-WCQR$_{19}$ | 0.2954 (0.1318) | 0.2744 (0.1889) | 0.3039 (0.2684) | 0.1827 (0.1469) | 0.2371 (0.1273) |
| $\chi^2(5)$ | Case 1 | K=1 | DC-CQR$_{19}$ | 0.6802 (0.5437) | 1.0100 (0.5728) | 1.0058 (0.6761) | 0.8602 (0.4020) | 1.0246 (0.7063) |
| | | | DC-WCQR$_{19}$ | 0.5698 (0.4092) | 0.6527 (0.5139) | 0.4266 (0.3156) | 0.6403 (0.3615) | 1.0812 (0.6263) |
| | | K=10 | DC-CQR$_{19}$ | 0.5397 (0.4428) | 0.5846 (0.5905) | 0.7915 (0.5921) | 0.7929 (0.6645) | 1.0197 (0.7892) |
| | | | DC-WCQR$_{19}$ | 0.6113 (0.4514) | 0.7426 (0.4481) | 0.4238 (0.2645) | 0.6406 (0.4690) | 0.6014 (0.3289) |
| | | K=100 | DC-CQR$_{19}$ | 0.8096 (0.6078) | 0.7034 (0.5517) | 0.9288 (0.7281) | 0.6111 (0.5332) | 0.7862 (0.5489) |
| | | | DC-WCQR$_{19}$ | 0.4832 (0.4327) | 0.8089 (0.5287) | 0.8366 (0.6402) | 0.5875 (0.3797) | 0.7725 (0.4264) |
| | Case 2 | K=1 | DC-CQR$_{19}$ | 0.3213 (0.1722) | 0.3973 (0.2688) | 0.4343 (0.3851) | 0.3433 (0.3163) | 0.4156 (0.2894) |
| | | | DC-WCQR$_{19}$ | 0.3975 (0.2929) | 0.3011 (0.2154) | 0.3399 (0.2703) | 0.3261 (0.3594) | 0.2649 (0.2060) |
| | | K=10 | DC-CQR$_{19}$ | 0.5879 (0.5158) | 0.7566 (0.6241) | 0.6658 (0.4536) | 0.6908 (0.6046) | 0.9551 (0.6001) |
| | | | DC-WCQR$_{19}$ | 0.3082 (0.2425) | 0.4756 (0.3058) | 0.2443 (0.2104) | 0.3890 (0.3041) | 0.5114 (0.2201) |
| | | K=100 | DC-CQR$_{19}$ | 0.8307 (0.5411) | 0.4409 (0.3685) | 0.3803 (0.2500) | 0.5064 (0.4159) | 0.5409 (0.5467) |
| | | | DC-WCQR$_{19}$ | 0.4263 (0.3170) | 0.2092 (0.1737) | 0.3545 (0.1303) | 0.3174 (0.3678) | 0.4305 (0.2958) |

Table 3: The mean of RMSE and MAD (standard deviation) when $\varepsilon \sim N(0, 1)$. $(10^{-2})$

| | **X** | Methods | K=1 | K=10 | K=100 |
|---|---|---|---|---|---|
| RMSE | Case 1 | LS | 0.3488 (0.1633) | 0.3488 (0.1633) | 0.3488 (0.1633) |
| | | LAD | 0.4872 (0.2037) | 0.4863 (0.2274) | 0.4898 (0.2224) |
| | | DC-CQR$_1$ | 0.4884 (0.2042) | 0.4852 (0.2287) | 0.4757 (0.1977) |
| | | ACQR$_{19}$ | 0.3686 (0.1595) | 0.3699 (0.1578) | 0.3723 (0.1545) |
| | | DC-CQR$_{19}$ | 0.3686 (0.1595) | 0.3697 (0.1599) | 0.3686 (0.1603) |
| | | DC-WCQR$_{19}$ | 0.3592 (0.1563) | 0.3594 (0.1562) | 0.3597 (0.1558) |
| MAD | Case 1 | LS | 0.3675 (0.1535) | 0.2753 (0.1336) | 0.2347 (0.1191) |
| | | LAD | 0.4954 (0.2521) | 0.3470 (0.1790) | 0.3201 (0.0931) |
| | | DC-CQR$_1$ | 0.4909 (0.2558) | 0.3418 (0.1778) | 0.3184 (0.1171) |
| | | ACQR$_{19}$ | 0.3835 (0.1814) | 0.2632 (0.1256) | 0.2667 (0.1035) |
| | | DC-CQR$_{19}$ | 0.3835 (0.1814) | 0.2616 (0.1276) | 0.2577 (0.1107) |
| | | DC-WCQR$_{19}$ | 0.3819 (0.1552) | 0.2767 (0.1334) | 0.2417 (0.0980) |
| RMSE | Case 2 | LS | 0.2004 (0.0697) | 0.2004 (0.0697) | 0.2004 (0.0697) |
| | | LAD | 0.2531 (0.0839) | 0.4793 (0.1831) | 0.4883 (0.1936) |
| | | DC-CQR$_1$ | 0.2584 (0.0814) | 0.3282 (0.0984) | 0.3203 (0.1062) |
| | | ACQR$_{19}$ | 0.2087 (0.0719) | 0.4402 (0.1238) | 0.4383 (0.1340) |
| | | DC-CQR$_{19}$ | 0.2087 (0.0716) | 0.2662 (0.0892) | 0.2695 (0.0777) |
| | | DC-WCQR$_{19}$ | 0.2055 (0.0708) | 0.2145 (0.0726) | 0.2149 (0.0737) |
| MAD | Case 2 | LS | 0.1689 (0.0566) | 0.1616 (0.0637) | 0.1936 (0.0623) |
| | | LAD | 0.1944 (0.0466) | 0.3872 (0.1477) | 0.3726 (0.1240) |
| | | DC-CQR$_1$ | 0.1955 (0.0499) | 0.2514 (0.1170) | 0.2993 (0.1228) |
| | | ACQR$_{19}$ | 0.1758 (0.0345) | 0.3912 (0.1233) | 0.3934 (0.0886) |
| | | DC-CQR$_{19}$ | 0.1758 (0.0345) | 0.2228 (0.0933) | 0.2634 (0.0984) |
| | | DC-WCQR$_{19}$ | 0.1739 (0.0406) | 0.1722 (0.0689) | 0.1993 (0.0791) |

Table 4: The mean of RMSE and MAD (standard deviation) when $\varepsilon \sim t(3)$. $(10^{-2})$

| | **X** | Methods | K=1 | K=10 | K=100 |
|---|---|---|---|---|---|
| RMSE | Case 1 | LS | 0.6781 (0.1716) | 0.6781 (0.1716) | 0.6781 (0.1716) |
| | | LAD | 0.5259 (0.0721) | 0.5152 (0.0796) | 0.5327 (0.1182) |
| | | DC-CQR$_1$ | 0.5240 (0.0694) | 0.5195 (0.1015) | 0.5262 (0.0849) |
| | | ACQR$_{19}$ | 0.4847 (0.1225) | 0.4860 (0.1228) | 0.4867 (0.1292) |
| | | DC-CQR$_{19}$ | 0.4847 (0.1225) | 0.4863 (0.1229) | 0.4864 (0.1180) |
| | | DC-WCQR$_{19}$ | 0.4783 (0.1330) | 0.4778 (0.1336) | 0.4790 (0.1338) |
| MAD | Case 1 | LS | 0.5197 (0.1802) | 0.5288 (0.1870) | 0.4790 (0.1005) |
| | | LAD | 0.3752 (0.1419) | 0.4032 (0.0740) | 0.4209 (0.1325) |
| | | DC-CQR$_1$ | 0.3756 (0.1363) | 0.3839 (0.0695) | 0.4399 (0.1064) |
| | | ACQR$_{19}$ | 0.3563 (0.1290) | 0.3258 (0.0747) | 0.3586 (0.1103) |
| | | DC-CQR$_{19}$ | 0.3563 (0.1290) | 0.3274 (0.0749) | 0.3644 (0.1081) |
| | | DC-WCQR$_{19}$ | 0.3414 (0.1379) | 0.3071 (0.0845) | 0.3578 (0.1189) |
| RMSE | Case 2 | LS | 0.3539 (0.1216) | 0.3539 (0.1216) | 0.3539 (0.1216) |
| | | LAD | 0.2752 (0.1032) | 0.5008 (0.1962) | 0.5031 (0.1921) |
| | | DC-CQR$_1$ | 0.2801 (0.1023) | 0.3583 (0.1256) | 0.3629 (0.1273) |
| | | ACQR$_{19}$ | 0.2504 (0.0847) | 0.5371 (0.1863) | 0.5702 (0.1884) |
| | | DC-CQR$_{19}$ | 0.2504 (0.0847) | 0.3317 (0.1047) | 0.3577 (0.1272) |
| | | DC-WCQR$_{19}$ | 0.2425 (0.0851) | 0.2540 (0.0782) | 0.2932 (0.0814) |
| MAD | Case 2 | LS | 0.2531 (0.0926) | 0.2501 (0.0809) | 0.3239 (0.0851) |
| | | LAD | 0.2234 (0.0756) | 0.4678 (0.1745) | 0.5380 (0.2383) |
| | | DC-CQR$_1$ | 0.2289 (0.0732) | 0.3338 (0.1091) | 0.3181 (0.1097) |
| | | ACQR$_{19}$ | 0.1753 (0.0299) | 0.4498 (0.1197) | 0.4037 (0.1684) |
| | | DC-CQR$_{19}$ | 0.1753 (0.0299) | 0.2630 (0.1258) | 0.5084 (0.1126) |
| | | DC-WCQR$_{19}$ | 0.1897 (0.0282) | 0.2197 (0.0971) | 0.2777 (0.0714) |

Table 5: The mean of RMSE and MAD (standard deviation) when $\varepsilon \sim \chi^2(5)$. $(10^{-2})$

| | **X** | Methods | K=1 | K=10 | K=100 |
|---|---|---|---|---|---|
| RMSE | Case 1 | LS | 2.1899 (0.7643) | 2.1899 (0.7643) | 2.1899 (0.7643) |
| | | LAD | 19.484 (3.8009) | 9.9637 (3.4820) | 5.6553 (2.0022) |
| | | DC-CQR$_1$ | 1.4196 (0.5407) | 1.4252 (0.5552) | 1.4330 (0.6088) |
| | | ACQR$_{19}$ | 1.1078 (0.3937) | 1.1162 (0.4024) | 1.1220 (0.3967) |
| | | DC-CQR$_{19}$ | 1.1069 (0.3937) | 1.1146 (0.4025) | 1.1179 (0.4018) |
| | | DC-WCQR$_{19}$ | 0.7746 (0.3006) | 0.7767 (0.3018) | 0.7804 (0.3040) |
| MAD | Case 1 | LS | 1.6769 (0.4736) | 1.9430 (0.8388) | 1.3456 (0.5321) |
| | | LAD | 15.146 (3.0817) | 9.1055 (4.1343) | 3.8980 (1.4488) |
| | | DC-CQR$_1$ | 1.1164 (0.4795) | 1.0964 (0.3975) | 1.0163 (0.2893) |
| | | ACQR$_{19}$ | 0.7817 (0.2909) | 0.8256 (0.3023) | 0.8017 (0.2481) |
| | | DC-CQR$_{19}$ | 0.7817 (0.2909) | 0.8262 (0.3036) | 0.8145 (0.2617) |
| | | DC-WCQR$_{19}$ | 0.6598 (0.1680) | 0.5879 (0.1531) | 0.5469 (0.1203) |
| RMSE | Case 2 | LS | 42.473 (0.1658) | 42.473 (0.1658) | 42.473 (0.1658) |
| | | LAD | 43.473 (0.2020) | 37.892 (2.0429) | 37.871 (1.0911) |
| | | DC-CQR$_1$ | 0.7075 (0.2118) | 0.8533 (0.2826) | 0.8777 (0.2748) |
| | | ACQR$_{19}$ | 0.5217 (0.1730) | 1.1870 (0.3536) | 1.1885 (0.3558) |
| | | DC-CQR$_{19}$ | 0.5217 (0.1730) | 0.6743 (0.2269) | 0.7693 (0.2664) |
| | | DC-WCQR$_{19}$ | 0.4331 (0.1301) | 0.4520 (0.1210) | 0.4673 (0.1224) |
| MAD | Case 2 | LS | 42.487 (0.1235) | 42.575 (0.1637) | 42.632 (0.2770) |
| | | LAD | 43.413 (0.1808) | 38.451 (2.4846) | 38.521 (0.5111) |
| | | DC-CQR$_1$ | 0.5991 (0.2198) | 0.8613 (0.3102) | 0.9259 (0.2663) |
| | | ACQR$_{19}$ | 0.4440 (0.1417) | 0.9487 (0.2861) | 1.1850 (0.4549) |
| | | DC-CQR$_{19}$ | 0.4440 (0.1417) | 0.6930 (0.2561) | 0.6254 (0.1859) |
| | | DC-WCQR$_{19}$ | 0.3388 (0.1097) | 0.3336 (0.0751) | 0.2572 (0.0834) |

Table 6: The means of confidence intervals and hypothesis testing for $\beta_3$ for Case 1.

| $\epsilon$ | K | Methods | $\hat{\beta}_3(\times 10^{-2})$ | Length$(\times 10^{-2})$ | C.P.(=T.P.) | P-value |
|---|---|---|---|---|---|---|
| N(0,1) | 1 | CQR | -0.0399 | 0.6571 | 0.94 | 0.5642 |
| | | WCQR | -0.0313 | 0.6421 | 0.94 | 0.5770 |
| | 10 | CQR | -0.0419 | 0.6571 | 0.96 | 0.5683 |
| | | WCQR | -0.0314 | 0.6421 | 0.96 | 0.5765 |
| | 100 | CQR | -0.0439 | 0.6571 | 0.96 | 0.5549 |
| | | WCQR | -0.0316 | 0.6420 | 0.96 | 0.5680 |
| t(3) | 1 | CQR | -0.1087 | 0.8039 | 0.92 | 0.4062 |
| | | WCQR | -0.1025 | 0.7847 | 0.94 | 0.4152 |
| | 10 | CQR | -0.1091 | 0.8085 | 0.91 | 0.4057 |
| | | WCQR | -0.1058 | 0.7895 | 0.94 | 0.4155 |
| | 100 | CQR | -0.1113 | 0.8095 | 0.91 | 0.4059 |
| | | WCQR | -0.0973 | 0.7845 | 0.94 | 0.4148 |
| $\chi^2(5)$ | 1 | CQR | 0.0218 | 1.7583 | 0.91 | 0.4822 |
| | | WCQR | -0.0010 | 1.3172 | 0.90 | 0.5107 |
| | 10 | CQR | 0.0278 | 1.7583 | 0.90 | 0.4860 |
| | | WCQR | 0.0029 | 1.3171 | 0.89 | 0.5100 |
| | 100 | CQR | 0.0154 | 1.7583 | 0.90 | 0.4871 |
| | | WCQR | -0.0062 | 1.3160 | 0.89 | 0.5153 |

Table 7: The simulation results for variable selection.

| | N(0,1) | | | | t(3) | | | |
|---|---|---|---|---|---|---|---|---|
| | RGMSE | GMSE(DC-WCQR) | sensitivity | specificity | RGMSE | GMSE(DC-WCQR) | sensitivity | specificity |
| K=1 | 1.9179 | 0.0010 | 0.9700 | 0.9667 | 1.4317 | 0.0013 | 0.9300 | 0.9667 |
| K=10 | 1.8773 | 0.0012 | 0.9400 | 0.9667 | 1.4701 | 0.0017 | 0.9200 | 0.9667 |
| K=100 | 1.9048 | 0.0011 | 0.9500 | 0.9667 | 1.4679 | 0.0014 | 0.9100 | 0.9500 |

Table 8: The estimates for the airline on-time data.

| Mthod | departure hour | distance | night | weekend | MSE | MAD |
|---|---|---|---|---|---|---|
| LS | 0.0012 | -0.0091 | -0.0037 | -0.0144 | 0.0732 | 0.1906 |
| LAD | -0.0009 | -0.0093 | -0.0057 | -0.0117 | 0.0744 | 0.1857 |
| DC-$CQR_1$ | 0.0018 | 0.0021 | -0.0009 | -0.0048 | 0.0737 | 0.1961 |
| $ACQR_{19}$ | 0.0007 | -0.0057 | -0.0030 | -0.0071 | 0.0733 | 0.1904 |
| DC-$CQR_{19}$ | 0.0007 | -0.0055 | -0.0030 | -0.0069 | 0.0733 | 0.1904 |
| DC-$WCQR_{19}$ | 0.0007 | -0.0065 | -0.0018 | -0.0044 | 0.0734 | 0.1904 |
| DC-$PWCQR_{19}$ | 0.0000 | -0.0055 | -0.0030 | -0.0069 | 0.0736 | 0.1884 |

## References


Bahadur, R. R. (1966). A note on quantiles in large samples. Annals of Mathematical Statistics, 37, 577-580.

Chen, X. and Xie, M. (2014). A split-and-conquer approach for analysis of extraordinarily large data. Statistica Sinica, 24, 1655-1684.

Draper, N. R. and Smith, H. (1998). Applied Regression Analysis, 3rd ed., Wiley, New York.

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

Fan, J. and Peng, H. (2004). On non-concave penalized likelihood with diverging number of parameters. Annals of Statistics, 32, 928-961.

Fan, J. and Lv, J. (2011). Non-concave penalized likelihood with NP-dimensionality. IEEE transaction on Information Theory, 57, 5467-5484.

Fan, T. H., Lin, D. and Cheng, K. F. (2007). Regression analysis for massive datasets. Data & Knowledge Engineering, 61, 554-562.

Van de Geer, S., Buhlmann P, Ritov, Y, and Dezeure R. (2014). On asymptotically optimal confidence regions and tests for high-dimensional models. The Annals of Statistics, 42, 1166-1202.

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

Jiang, X. J., Li, J. Z., Xia, T. and Yan, W. F. (2016). Robust and efficient estimation with weighted composite quantile regression. Physica A, 457, 413-423.

Jiang, R., Zhou, Z. G., Qian, W. M. and Shao, W. Q. (2012). Single-index composite quantile regression. Journal of the Korean Statistical Society, 3, 323-332.

Jiang, R., Zhou, Z. G., Qian, W. M. and Chen, Y. (2013). Two step composite quantile regression for single-index models. Computational Statistics and Data Analysis, 64, 180-191.

Jiang, R., Qian, W. M. and Li, J. R. (2014). Testing in linear composite quantile regression models. Computational Statistics, 29, 1381-1402.

Jiang, R., Qian, W. M. and Zhou, Z. G. (2015). Composite quantile regression for linear errors-in-variables models. Hacettepe Journal of Mathematics and Statistics, 44, 707-713.

Jiang, R., Qian, W. M. and Zhou, Z. G. (2016a). Weighted composite quantile regression for single-index models. Journal of Multivariate Analysis, 148, 34-48.

Jiang, R., Qian, W. M. and Zhou, Z. G. (2016b). Single-index composite quantile regression with heteroscedasticity and general error distributions. Statistical Papers, 57, 185-203.

Jiang, R., Qian, W. M. and Zhou, Z. G. (2018). Weighted composite quantile regression for partially linear varying coefficient models. Communications in Statistics-Theory and Methods, 47, 3987-4005.

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, Series B, 72, 49-69.

Kai, B., Li, R. and Zou, H. (2011). New efficient estimation and variable selection methods for semiparametric varying-coefficient partially linear models. Annals of Statistics, 39, 305-332.

Kleiner, A., Talwalkar, A., Sarkar, P. and Jordan, M. I. (2014). A scalable bootstrap for massive data. Journal of the Royal Statistical Society, Series B, 76, 795-816.

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

Lam, C. and Fan, J. (2008). Profile-Kernel likelihood inference with diverging number of parameters. Annals of Statistics, 36, 2232-2260.

Li, R., Lin, D. and Li, B. (2013). Statistics inference in massive data sets. Applied Stochastic Models in Business and Industry, 29, 399-409.

Liang, F., Cheng, Y., Song, Q., Park, J. and Yang, P. (2013). A resampling-based stochastic approximation method for analysis of large geostatistical data. Journal of the American Statistical Association, 108, 325-339.

Lin, N. and Xi, R. (2011). Aggregated estimating equation estimation. Statistics and Its Interface, 4, 73-83.

Ma, P., Mahoney, M. W. and Yu, B. (2015). A statistical perspective on algorithmic leveraging. The Journal of Machine Learning Research, 6, 861-911.

Meishausen, N. and Buhlmann, P. (2010). Stability selection. Journal of the Royal Statistical Society, Series B, 72, 417-473.

Ning, Z. and Tang, L. (2014). Estimation and test procedures for composite quantile regression with covariates missing at random. Statistics and Probability Letters, 95, 15-25.

Schifano, E. D., Wu, J., Wang, C., Yan, J. and Chen, M. H. (2016). Online updating of statistical inference in the big data setting. Technometrics, 58, 393-403.

Shah, R. and Samworth, R. J. (2013). Variable selection with error control: Another look at stability selection. Journal of the Royal Statistical Society, Series B, 75, 55-80.

Tang, L., Zhou, Z. and Wu, C. (2012). Weighted composite quantile estimation and variable selection method for censored regression model. Statistics and Probability Letters, 3, 653-663.

Tang, L. and Zhou, Z. (2015). Weighted local linear CQR for varying-coefficient models with missing covariates. Test, 24, 583-604.

Tian, Y., Zhu, Q. and Tian, M. (2016). Estimation of linear composite quantile regression using EM algorithm. Statistics and Probability Letters, 117, 183-191.

Tibshirani, R. J. (1996). Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society, Series B, 58, 267-288.

Wang, C., Chen, M. H., Schifano, E. D., Wu, J. and Yan, J. (2015). Statistical methods and computing for big data. Statistics and its Interface, arXiv:1502.07989.

Wang, H., Li, G. and Jiang, G. (2007a). Robust regression shrinkage and consistent variable selection via the LAD-LASSO. Journal of Business and Economic Statistics, 20, 347-355.

Wang, H., Li, R. and Tsai, C. L. (2007b). Tuning parameter selectors for the smoothly clipped absolute deviation method. Biometrika, 94, 553-568.

Wu, Y. and Liu, Y. (2009). Variable selection in quantile regression. Statistica Sinica, 19, 801-817.

Xu, Q., Cai, C., Jiang, C., Sun, F. and Huang, X. (2017). Block average quantile regression for massive dataset. Statistical Papers, DOI:10.1007/s00362-017-0932-6 .

Zhao, T., Kolar, M. and Liu, H. (2017). A general framework for robust testing and confidence regions in high-dimensional quantile regression. Tech. rep.

Zhao, K. and Lian, H. (2016). A note on the efficiency of composite quantile regression. Journal of Statistical Computation and Simulation, 86, 1334-1341.

Zhao, W., Lian, H. and Song, X. (2017). Composite quantile regression for correlated data. Computational Statistics and Data Analysis, 109, 15-33.

Zou. H. (2006). The adaptive LASSO and its oracle properties. Journal of The American Statistical Association, 101, 1418-1429.

Zou, H., Hastie, T. and Tibshirani, R. (2007). On the degrees of freedom of the Lasso. Annals of Statistics, 35, 2173-2192.

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