Weighted Type of Quantile Regression and its Application

Xuejun Jiang, Tian Xia, and Dejun Xie

Abstract—In this paper we introduce a weighted composite
quantile regression (CQR) estimation approach and study its
application in nonlinear models such as exponential models
and ARCH type of models. The weighted CQR is augmented
by using a data-driven weighting scheme. With the error dis-
tribution unspecified, the proposed estimators share robustness
from quantile regression and achieve nearly the same efficiency
as the oracle maximum likelihood estimator (MLE) for a variety
of error distributions including the normal, mixed-normal,
Student’s t, Cauchy distributions and etc.. We also suggest an
algorithm for fast implementation of the proposed methodology.
Simulations are conducted to compare the performance of
different estimators, and the proposed approach is used to
analyse the daily S&P 500 Composite index, which endorse
our theoretical results.

Index Terms—weighted CQR, Oracle MLE, Extended inte-
rior algorithm, Double threshold ARCH models.

I. INTRODUCTION

Quantile regression (QR), introduced by Koenker and
Bassett (1978), receives increasing attention in econometrics
and statistics for its advantages over mean regression at
least in two aspects: (i) the stochastic relationship between
random variables can be portrayed much better and with
much more accuracy using quantile regression than simple
mean regression (see for example Chaudhuri, Doksum and
Samarov, 1997); (ii) quantile regression provides more robust
and consequently more efficient estimates than the mean re-
gression when the error is non-normal (Koenker and Bassett,
1978; Koenker and Zhao, 1996). This motivates us to work
along the line of QR for developing more efficient estimation
methods.

The CQR in Zou and Yuan (2008) is robust compared
to traditional QR. The CQR they used is a sum of differ-
ent quantile regression (QR) [Koenker and Bassett (1978)]
at predetermined quantiles, which uses equal weights for
different QR (see Section 2 for details). Intuitively, equal
weights are not optimal in general, and hence a more efficient
CQR should exist. Therefore, in this article we introduce a
“weighted CQR” estimation method and let the data decide
the weights to improve efficiency while keeping robustness
from the QR. The weighted CQR method is applicable to
various models, but in this article we focus only on the
nonlinear model

\[ y_i = f(x_i', \beta) + \varepsilon_i, \quad i = 1, \ldots, n, \]  

(1.1)

where \( \varepsilon_i \)'s are independent random errors with unknown
distribution function \( G(\cdot) \) and density \( g(\cdot) \), and the function

\[ f(\cdot, \beta) \]  

is known up to a \( p \)-dimensional vector of parameters
\( \beta \).

We will address the issue of choosing weights when using
the weighted CQR. Since the weights in the weighted CQR
are allowed to be negative, the proposed weighted CQR
is different from the common QR and the CQR (see also
Section 2). When the weights are all equal and the model is
linear with a fixed number of parameters, our method reduces
to that of Zou and Yuan (2008). Since the proposed weighted
CQR involves a vector of weights, we develop a data-
driven weighting strategy which maximizes the efficiency
of the weighted CQR estimators. The resulting estimation
is adaptive in the sense that it performs asymptotically the
same as if the theoretically optimal weights were used. The
adaptive estimation is robust against outliers and heavy-
tailed error distributions, like the Cauchy distribution, and
efficient as nearly as the oracle MLE for a variety of error
distributions (see Table I). This is a great advantage of the
proposed estimation method, since the adaptive weighted
CQR estimators do not require the form of error distribution
and achieves nearly the Cramér-Rao lower bound.

The weighted CQR estimators admit no close form and
involve minimizing convoluted nonlinear functions, so it is
challenging to derive asymptotic properties and to implement
the methodology. Theoretically, we will establish asympto-
totic normality of the resulting estimators and show their
optimality, no matter whether the error variance is finite
or not. Practically, we will develop an algorithm for fast
implementation of the proposed methodology. This algorithm
solves a succession of linearized weighted QR problems,
each of whose dual problems is derived. We will use the
“interior point algorithm” [see Jiang, X., et.cl.(2012), V an-
derbei, et.cl.(1986), Koenker and Park (1996)] to solve these
dual problems. The resulting algorithm is easy to implement.
Simulations endorse our discovery.

The rest of the article is organized as follows. In Section II
we introduce the weighted CQR for model (1.1). In section
III we consider the computation method for the proposed
methodology and conduct simulations to demonstrate the per-
formance of it. In Section IV we apply the proposed methods
to analyse a real dataset. Finally, Proofs for theorems are put
in supplementary materials to save space.

II. WEIGHTED COMPOSITE QUANTILE REGRESSION

Our idea can be well motivated from the linear model,

\[ y_i = x_i' \beta + \varepsilon_i, \quad i = 1, \ldots, n, \]  

(II.2)

where \( \varepsilon_i \)'s are i.i.d. noise with unknown distribution
\( G(\cdot) \) and density \( g(\cdot) \).

By Koenker and Bassett (1978), the \( \tau \)-th QR estimate of
\( \beta \) can be obtained via minimizing

\[ \sum_{i=1}^{n} |G(\tau) - G(\beta x_i' + \varepsilon_i)|. \]
over $\beta$ and $b_{r_{k}}$, where $\{\tau_{k}\}_{k=1}^{K}$ are predetermined over (0, 1). This is the aforementioned CQR in the introduction section.

Note that the CQR method uses the same weight for different QR models. Intuitively, it is more effective if different weights are used. Applying the weighting scheme to model (I.1), one can estimate $\beta$ by minimizing

$$L_{n}(\beta, b) = \sum_{k=1}^{K} \sum_{i=1}^{n} \omega_{k} \rho_{\tau_{k}}(y_{i} - f(\mathbf{x}_{i}; \beta) - b_{r_{k}}),$$

over $\beta$ and $b_{r_{k}}$, where $\{\tau_{k}\}_{k=1}^{K}$ are predetermined over (0, 1). This is the aforementioned CQR in the introduction section.

In order to derive asymptotic normality of the proposed estimator, in the following we introduce some notations and conditions. Let $\beta^{*}$ be the true value of $\beta$, $b_{\tau_{k}}$ be the $\tau_{k}$-th quantile of $\varepsilon$, and $b^{*} = (b_{\tau_{1}^{*}}, \ldots, b_{\tau_{K}^{*}})'$. Denote by $f^{*}_{i} = f(\mathbf{x}_{i}; \beta^{*})$, $\nabla f^{*}_{i} = [\partial f(\mathbf{x}_{i}; \beta^{*})/\partial \beta]_{\beta=\beta^{*}}$, and $\nabla^{2} f^{*}_{i} = [\partial^{2} f(\mathbf{x}_{i}; \beta^{*})/\partial \beta \partial \beta]_{\beta=\beta^{*}}$. Assume that

(a) $\mathbf{G} = \text{var}(\nabla f^{*}_{i}) > 0$.
(b) The error $\varepsilon_{i}$ has the distribution function $G(\cdot)$ and density function $g(\cdot)$ which is positive and continuous at the $\tau_{k}$-th quantiles $b_{\tau_{k}^{*}}$.
(c) There is a large enough open subset $\Omega \subset \mathbb{R}^{p}$ which contains the true parameter point $\beta^{*}$, such that for all $x_{k}$, the second derivative matrix $\nabla^{2} f(\mathbf{x}_{i}; \beta)$ of $f(\mathbf{x}_{i}; \beta)$ with respect to $\beta$ satisfies that

$$\|\nabla^{2} f(\mathbf{x}_{i}; \beta_{1}) - \nabla^{2} f(\mathbf{x}_{i}; \beta_{2})\| \leq M(\mathbf{x}_{i})\|\beta_{1} - \beta_{2}\|$$

$$|\partial^{2} f(\mathbf{x}_{i}; \beta)/\partial \beta \partial \beta| \leq N_{jk}(x_{i})$$

for all $\beta_{1}, \beta_{2} \in \Omega$, where $E[M^{2}(x_{i})] < \infty$ and $E[N_{jk}(x_{i})] < C_{1} < \infty$ for all $j, k$.

Under these mild conditions, we have the following asymptotic normality result.

**Theorem 1:** Let

$$S(\omega) = \sum_{k=1}^{K} \omega_{k} \omega_{k'} \min(\tau_{k}, \tau_{k'}) (1 - \max(\tau_{k}, \tau_{k'})).$$

Under the conditions (a)-(c),

$$\sqrt{n}(\beta - \beta^{*}) \Rightarrow N(0, \sigma^{2}(\omega)\mathbf{G}^{-1}),$$

where

$$\sigma^{2}(\omega) = S(\omega)\{\sum_{k=1}^{K} \omega_{k} g(b_{\tau_{k}^{*}})\}^{-2}$$

For model (II.2), $\mathbf{G} = \text{var}(x_{1})$. If all $\omega_{k}$ are equal, then Theorem 1 reduces to the asymptotic normality of the CQR estimators in Zou and Yuan (2008). When $K = 1$ and $\tau_{1} = \tau$, it follows from the above theorem that the $\tau$-th QR estimate of $\beta$ is $\sqrt{n}$-consistent and asymptotically normal with mean zero and variance $g^{-2}(b_{\tau}^{*})(1 - \tau)\mathbf{G}^{-1}$.

Since $\mathbf{G}$ does not involve $\omega$, the weights should be selected to minimize $\sigma^{2}(\omega)$. Let $g = (g(b_{\tau_{1}^{*}}), \ldots, g(b_{\tau_{K}^{*}}))'$, and let $\Omega$ be a $K \times K$ matrix with the $(k, k')$ element being $\Omega_{kk'} = \min(\tau_{k}, \tau_{k'}) (1 - \max(\tau_{k}, \tau_{k'}))$. Then the optimal weight $\omega_{opt}$, which minimizes $\sigma^{2}(\omega)$, can be shown as

$$\omega_{opt} = (g'\Omega^{-2}g)^{-1/2}\Omega^{-1}g.$$

The optimal weight components can be very different, and some of them may even be negative. In fact, in our simulations we also experience such a scenario. This reflects the necessity to use a data-driven weighting scheme. The usual nonparametric density estimation methods such as kernel smoothing based on estimated residuals $\hat{\varepsilon}_{i}$ can provide a consistent estimation $\hat{g}(\cdot)$ of $g(\cdot)$. Let the resulting estimate of $g$ be $\hat{g}$. Then $\hat{\omega} = (\hat{g}'\Omega^{-2}\hat{g})^{-1/2}\Omega^{-1}\hat{g}$ provides a nonparametric estimator of $\omega$. This leads to an adaptive estimator of $\beta$ by minimizing

$$\sum_{k=1}^{K} \sum_{i=1}^{n} \rho_{\tau_{k}}(y_{i} - f(\mathbf{x}_{i}; \beta) - b_{r_{k}})$$

over $\beta$ and $\tau_{k}$, where $\omega_{k}$ is the $k$-th component of $\omega$. Let the resulting estimator of $\beta$ be $\hat{\beta}_{2}$. Then $\hat{\beta}_{2}$ is asymptotically normal from the following theorem.

**Theorem 2:** Under the same conditions as in Theorem 1,

$$\sqrt{n}(\hat{\beta}_{2} - \beta^{*}) \Rightarrow N(0, \sigma^{2}(\omega_{opt})\mathbf{G}^{-1}).$$

Since $\sigma^{2}(\omega_{opt}) = (g'\Omega^{-2}g)^{-1}$, $\hat{\beta}_{2}$ has the same asymptotic variance matrix as $\beta_{1}$ as if $\omega_{opt}$ were known. That is, the estimator $\hat{\beta}_{2}$ is adaptive. Therefore, $\omega$ is called the adaptive weight vector. By Theorem 2, the asymptotic relative efficiency (ARE) of the adaptive WCQR estimation with respect to the least squares (LS) estimation is $e(WCQR, LS) = \sigma^{2}(\omega_{opt})^{-1}g$. Since the oracle maximum likelihood (ML) estimator of $\beta$ has asymptotic variance matrix $I_{g}^{-1}g^{-1}$, we have

$$e(WCQR, ML) = I_{g}^{-1}g\Omega^{-1}g.$$

For equally spaced $\{\tau_{k}\}_{k=1}^{K}$, the adaptive estimator $\hat{\beta}_{2}$ is nearly efficient as the oracle MLEs for various error distributions (see Theorem 4 in Jiang,X., et al. (2012)), which is a great advantage of the proposed methodology.

For each $K$, the AREs of the adaptive estimator $\hat{\beta}_{2}$ with respect to some common estimators can be calculated. To appreciate how much efficiency is gained in practice, we investigate the performance of common estimators. Table I reports those AREs for linear models with various error distributions, which demonstrates that $\hat{\beta}_{2}$ is highly efficient for all the distributions under consideration. For linear models, Leng (2009) demonstrated that his regularized rank regression estimator ($R^{2}$) was quite efficient and robust. Table I indicates that the proposed adaptive estimate dominates the
\[ R^2 \] for all error distributions and is much more efficient than it when the error follows the Cauchy or chi-squared distribution. It also suggests that typically one could choose \( K = 10 \) in practice and efficiency is much gained, as shown in simulations.

### III. Numerical Implementation

#### A. Extended interior algorithm

Minimization in (II.4) involves a complicate nonlinear optimization problem. We use the interior point algorithm (see Jiang, X., et al. (2012)) to solve the problem. Matlab codes are available upon request for the proposed methods.

Consider the equivalent problem of (II.4):

\[
\min_{\theta} \sum_{k=1}^{K} \omega_k \sum_{i=1}^{n} \rho_{\tau_k}(y_i - l_k(\theta)),
\]

where \( l_k(\theta) = f(x_k, \beta) + b_k \), and \( \theta = (b_1, \ldots, b_{\tau_\max}, \beta') \).

Following Osborne and Watson (1971), we solve the problem (III.6) using the following algorithm:

1. Given the current value, \( \theta^{(j)} \), of \( \theta \), calculate \( t \) to minimize

\[
\sum_{k=1}^{K} \omega_k \sum_{i=1}^{n} \rho_{\tau_k}(y_i - l_k(\theta^{(j)}) - \nabla l_k(\theta^{(j)}))^2,
\]

where \( \nabla l_k(\theta^{(j)}) = (d_{ik}(\theta^{(j)}) \rho_i(\theta^{(j)}) \theta^{(j)}) \theta = \theta^{(j)} \).

Let the minimum be \( s(\theta) \), and let the minimizer be \( t = t(\theta) \).

2. Calculate \( \lambda \) to minimize

\[
\sum_{k=1}^{K} \omega_k \sum_{i=1}^{n} \rho_{\tau_k}(y_i - l_k(\theta^{(j)} + \lambda^{(j)})).
\]

Let the minimum be \( s^{(j+1)} \) with \( \lambda = \lambda^{(j)} \).

3. Set \( \theta^{(j+1)} = \theta^{(j)} + \lambda^{(j)} t^{(j)} \). Update the current value of \( \theta \) by \( \theta^{(j+1)} \), and repeat the above procedure until the new iterate fails to improve the objective function by a specified tolerance such as \( 10^{-4} \).

In the above method, the problem (III.8) can be easily solved by line search in the resulting direction \( t = t^{(j)} \), but one has to solve a succession of linearized weighted quantile regression problems in (III.7). Let \( y_{ik} = y_i - l_k(\theta^{(j)}) \) and \( x_{ik} = \nabla l_k(\theta^{(j)}) \). Then the process (III.7) becomes

\[
\min_{t} \sum_{k=1}^{K} \omega_k \sum_{i=1}^{n} \rho_{\tau_k}(y_{ik} - x_{ik}^t),
\]

whose dual problem can be shown as

\[
\max \{ \tilde{y}^t d | d = (d_1^t, \ldots, d_K^t) \},
\]

where \( \tilde{y} = (y_1^t, \ldots, y_K^t) \), \( X = (X_1^t, \ldots, X_K^t) \), \( X_k = (x_{1k}, \ldots, x_{nk}) \), and \( y_k = (y_{1k}, \ldots, y_{nk}) \). The interior point method can be used to solve the dual problem (III.10), the algorithm details are listed as follows:

1. For any initial feasible point \( \theta \), e.g., \( d = 0 \), following Meketon (1986), set \( D_k = \text{diag}(\min(\omega_k \tau_k - d_{ik}, d_{ik} - \omega_k (\tau_k - 1))) \), \( D = \text{diag}(D_1, \ldots, D_K) \), \( s = D^{2} (I - X(X'D^2 X)^{-1}X'D^2 \tilde{y}) \), and \( t = (X'D^2 X)^{-1}X'D^2 \tilde{y} \).

2. Set \( d_k = d_k + (\eta/\gamma_k) s_k \), where \( \gamma_k = \max(\sum_{i=1}^{K} s_{ik}/\omega_k \tau_k - d_{ik}), -s_{ik}/(d_{ik} - \omega_k (\tau_k - 1)) \), and \( \eta \in (0,1) \) is the constant chosen to ensure feasibility. As suggested by Koekner and Park, \( \eta = 0.97 \).

3. Set \( d = d^* \). Updating \( D, s, \) and the new \( d^* \) continues the iteration.

After solving (III.10) using the above interior point algorithm, we arrive at the next loop which uses the current value \( \theta = \theta^{(j+1)} \) for the primal problem in (III.9). This leads to the updated dual problem (III.10) with \( y_{ik} = y_i - f_k(\theta^{(j+1)}) \) and \( \theta = \theta^{(j+1)} \). The current \( d \) should be adjusted to ensure that it is feasible for the new value of \( \theta \). Similar to Koekner and Park (1996), we project the current \( d \) onto the null space of the new \( X \), i.e., \( d = (I - X(X'X)^{-1}X')d \) and then shrinking it to insure that \( d_k \) lies in \( [\omega_k \tau_k - 1], \omega_k \tau_k \) \( \in \mathbb{N} \), so the adjusted \( d \) becomes

\[
d_k = \left( \max \left\{ \frac{d_{ik}}{\omega_k (\tau_k - 1)}, \frac{d_{ik}}{\omega_k \tau_k} \right\} + \delta \right)^{-1} d_k
\]

for some tolerance parameter \( \delta > 0 \).

#### B. Simulations

In this section, we report on simulations to investigate the advantages of the WQCR estimation. An exponential regression model was used:

\[
y = 1 + b \exp(c'x) + \varepsilon,
\]

where \( b = (b_1, b_2, b_3)' \) and \( \varepsilon \) is the error. The true values of parameters were set as \( b = 1.5, \) and \( \varepsilon = (-0.6, -0.8, -0.7)' \). We draw from the working model 400 samples of sizes 200 and 400. In each simulation, the components of \( X \) were jointly normal distributed with the pairwise correlation coefficient 0.5 and standard normal as marginals. We considered three sets of errors: \( N(0, 1), t(4), \) and \( \chi^2(4) \). All of them were centralized and scaled so that the medians of the absolute errors were ones. As before, we used \( K = 10 \) and equally spaced \( \tau_k \) on \( (0,1) \).

We compared four estimation methods: the L1, CQR, WQCR estimation, and the oracle MLE (maximum likelihood estimation) of known error distribution. In each simulation the “root of mean squared errors (RMSE)” for different coefficient estimators were calculated, and their average over simulations is reported in Tables II-IV, where \( \Sigma \) denotes the sum of RMSE for all components in \( \beta \). Therefore, better estimators should have smaller \( \Sigma \) values. As expected, the
oracle MLE performs the best, the WCQR performs better than the CQR and $L_1$, and the $L_1$ is the worst. In terms of overall performance (the value of $\Sigma$), WCQR estimators uniformly dominate the $L_1$ and CQR estimators and are comparable to the oracle MLE under all kind of errors we considered. This is in accordance with our previous theoretical results.

### IV. A REAL EXAMPLE

We study the daily S&P 500 Composite index from January 3, 2000 to July 27, 2011. This index represents the bulk of the daily value in the US equity market. The return series $y_t$ is defined as the difference of the log-price. We are interested in the asymmetry of the conditional mean and conditional variance. The double-threshold ARCH (DTARCH) model [see Li and Li (1996), Hui, Y. V. and Jiang, J. (2005), Jiang, J. and et.al. (2013)] can be used to describe this kind of asymmetry.

The proposed estimation approaches are applied to the data set with 2908 observations. We set the threshold parameter $r = 0$ and the delay parameter $d = 1$ in DTARCH models, which is consistent with observations in the stock market. Therefore, we consider the following DTARCH model for the return series $y_t$:

$$y_t = \begin{cases} 
\alpha_1^{(1)} y_{t-1} + \epsilon_t, & \text{if } y_{t-1} \leq 0, \\
\alpha_2^{(1)} y_{t-1} + \epsilon_t, & \text{if } y_{t-1} > 0,
\end{cases}$$

where $\epsilon_t = h_t u_t$, with

$$h_t = \begin{cases} 
\beta_0^{(1)} + \beta_1^{(1)} |\epsilon_{t-1}|, & \text{if } y_{t-1} \leq 0, \\
\beta_0^{(2)} + \beta_1^{(2)} |\epsilon_{t-1}|, & \text{if } y_{t-1} > 0.
\end{cases}$$

For comparison, we apply the WCQR estimate with equal weights $\omega_k$, labeled as “CQR estimate”, and the QR estimates.

### TABLE IV

RMSE (multiplied by $10^3$) for different QR estimators under chi-square error.

<table>
<thead>
<tr>
<th>Estimates</th>
<th>$b$</th>
<th>$\hat{c}_1$</th>
<th>$\hat{c}_2$</th>
<th>$\hat{c}_3$</th>
<th>$\Sigma$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$L_1$ QC</td>
<td>147</td>
<td>42</td>
<td>34</td>
<td>30</td>
<td>254</td>
</tr>
<tr>
<td>$L_1$ CQR</td>
<td>145</td>
<td>48</td>
<td>28</td>
<td>24</td>
<td>214</td>
</tr>
<tr>
<td>WCQR</td>
<td>124</td>
<td>38</td>
<td>28</td>
<td>24</td>
<td>214</td>
</tr>
<tr>
<td>OML</td>
<td>126</td>
<td>37</td>
<td>29</td>
<td>27</td>
<td>221</td>
</tr>
</tbody>
</table>

(i.e. $K = 1$) at $\tau_K = 0.5$ and 0.75 to fit the model. We calculate the estimated parameters and their standard errors. The results are reported in Tables V. For $\tau_K = 0.5$, it corresponds to the $L_1$-estimate. Since the result from $L_1$-estimation is poor, we omit it there.

It is observed that all the estimation approaches identify positive beta’s values. This is desired because the volatility coefficients, $\beta$’s, are nonnegative. Negative $\alpha_1^{(1)}$ and $\alpha_2^{(1)}$ indicate that the future mean return will be forecasted as positive (negative) if the current return is negative (positive). This is expected in an efficient market. On the other hand, negative returns in this model have about 3 times the effect of positive returns on future conditional scales. This indicates that the volatility is significantly higher when prices are falling. That is, volatility tends to be higher in bear markets, an asymmetric volatility effect described by Nelson’s EGARCH model (Nelson (1991)).

Among the three estimation methods, the WCQR estimator is the best for all estimators because it has the smallest standard deviation.

### ACKNOWLEDGMENT

This research was supported by NSF Grant 11101432 and NSF Grant 11361013. It was also supported in part by the Fundamental Research Funds for the Central Universities of China (No: 31541111215) for Xuejun Jiang at Zhongnan University of Economics and Law, Wuhan, China. Correspondence should be addressed to Xuejun Jiang, Department of Financial Mathematics and Financial Engineering, South University of Science and Technology, Shenzhen, China, E-mail: jiang.xj@sustc.edu.cn.

### REFERENCES


