Efficient Quantile Regression for Heteroscedastic Models

Yoonsuh Jung, *University of Waikato, New Zealand*
Yoonkyung Lee, *The Ohio State University*
Steven N. MacEachern, *The Ohio State University*

Technical Report No. 877

March, 2014

Department of Statistics
The Ohio State University
1958 Neil Avenue
Columbus, OH 43210-1247

Efficient Quantile Regression for Heteroscedastic Models

Yoonsuh Jung

Department of Statistics, University of Waikato, Hamilton 3240, New Zealand

yoonsuh@waikato.ac.nz

Yoonkyung Lee and Steven N. MacEachern*

Department of Statistics, The Ohio State University, Columbus, Ohio 43210

yklee@stat.osu.edu and snm@stat.osu.edu

Abstract

Quantile regression provides estimates of a range of conditional quantiles. This stands in contrast to traditional regression techniques, which focus on a single conditional mean function. Lee et al. (2012) proposed efficient quantile regression by rounding the sharp corner of the loss. The main modification generally involves an asymmetric ℓ_2 adjustment of the loss function around zero. We extend the idea of ℓ_2 adjusted quantile regression to linear heterogeneous models. The ℓ_2 adjustment is constructed to diminish as sample size grows. Conditions to retain consistency properties are also provided.

KEYWORDS: Check loss function; heteroscedasticity; quantile regression

1 Introduction

Quantile regression has emerged as a useful tool for providing conditional quantiles of a response variable Y given values of a predictor X. This allows us to estimate not only the center, but also the upper or lower tail of the conditional distribution of interest. Due to the ability of quantile regression to capture the full distributional aspects, rather than only the conditional mean, quantile regression has been widely applied. Koenker and Bassett (1978) pioneered quantile regression and proved consistency properties. Bassett and Koenker (1978) showed efficiency of median regression, when the median is more efficient than the mean in a location model. To

*This research was supported in part by National Security Agency grant H98230-10-1-0202 and National Science Foundation grant DMS-12-09194.

overcome the restriction of *iid* errors in linear quantile regression, Gutenbrunner and Jurečková (1992), Koenker and Zhao (1994), and He (1997) developed heterogeneous error models. A comprehensive review is provided in Koenker (2005).

Quantile regression employs asymmetric absolute loss, the so-called "check loss", as a criterion for minimization. Since the derivative of the check loss does not exist at its minimum, care must be taken with computation. Two main computational strategies exist: linear programming techniques can be used to minimize the criterion function, and hence to solve the problem exactly, (Koenker and D'Orey; 1987; Koenker and Park; 1996), or the cusp of the check loss can be smoothed to allow computational techniques that rely on differentiability. For the latter approach, Horowitz (1998) employed a smooth kernel under median regression, and Nychka et al. (1995) suggested a tiny, symmetric quadratic adjustment of the loss function over the interval $(-\epsilon, \epsilon)$ which is a small enough adjustment to be effectively zero relative to the scale of the data.

Lee et al. (2012) proposed a carefully tailored quadratic adjustment to the loss function that differs in intent from the computational adjustment, and that produces quite different results. Their modification is intended to make quantile regression more efficient, and the modification is asymmetric, except for median regression. The modification leads to superior finite sample performance by exploiting the bias-variance trade-off. To obtain the requisite reduction in variance, the interval of adjustment is not vanishingly small, as are the computationally-motivated adjustments, but is of substantial size. Jung et al. (2010) present a practical rule for choosing the magnitude of adjustment.

Many real applications of regression often call for heterogeneous error models. Although the modified quantile regression in Lee et al. (2012) allows for different local densities around regression quantiles, we expect further efficiency gain in quantile estimation when the differences in the local densities are accounted for. In this paper, we extend the efficient quantile regression method in Lee et al. (2012) to models with heterogeneous error distributions and demonstrate its effectiveness in data analysis. Subsequent sections provide a brief overview of efficient quantile regression, theoretical development for heterogeneous error models, practical strategies for implementation of the method, and documentation of its performance through simulations and data analysis.

2 Overview of Efficient Quantile Regression

To estimate the qth regression quantile, the check loss function ρ_q is employed:

$$\rho_q(r) = \begin{cases} qr & \text{for } r \geq 0 \\ -(1-q)r & \text{for } r < 0. \end{cases} \tag{1}$$

We first consider a linear model of the form $y_i = x_i^\top \beta + u_i$, where $x_i,\ \beta \in \mathbb{R}^p$ and the u_i's are *iid* from some distribution with qth quantile equal to zero. The quantile regression estimator $\hat{\beta}$ is the minimizer of

$$L(\beta) = \sum_{i=1}^{n} \rho_q(y_i - x_i^\top \beta). \tag{2}$$

To treat the observations in a systematic fashion, Lee et al. (2012) introduce case-specific parameters γ_i which change the linear model to $y_i = x_i^\top \beta + \gamma_i + u_i$. From the fact that this is a super-saturated model, $\gamma = (\gamma_1, ..., \gamma_n)^\top$ should be penalized. Together with the case-specific parameters and an additional penalty for γ, the objective function to minimize given in (2) is modified to be

$$L(\beta, \gamma) = \sum_{i=1}^{n} \rho_q(y_i - x_i^\top \beta - \gamma_i) + \frac{\lambda_\gamma}{2} J(\gamma), \tag{3}$$

where $J(\gamma)$ is the penalty for γ and λ_γ is a penalty parameter. Since the check loss function is piecewise linear, the quantile regression estimator is inherently robust. For improving efficiency, an ℓ_2 type penalty for the γ is considered. As detailed in Lee et al. (2012), a desired invariance suggests an asymmetric ℓ_2 penalty of the form $J(\gamma_i) \equiv \{q/(1-q)\}\gamma_{i+}^2 + \{(1-q)/q\}\gamma_{i-}^2$, where $\gamma_{i+} = \max(\gamma_i, 0)$ and $\gamma_{i-} = \max(-\gamma_i, 0)$. With the $J(\gamma_i)$, let us examine the values of the γ_i which minimizes (3), given β. First, note that $\min_\gamma L(\hat{\beta}, \gamma)$ decouples to minimization over individual γ_i. Hence, given $\hat{\beta}$ and a residual $r_i = y_i - x_i^\top \hat{\beta}$, $\hat{\gamma}_i$ is now defined to be

$$arg \min_{\gamma_i} \mathcal{L}_{\lambda_\gamma}(\hat{\beta}, \gamma_i) \equiv \rho_q(r_i - \gamma_i) + \frac{\lambda_\gamma}{2} J(\gamma_i), \tag{4}$$

and is explicitly given by

$$-\frac{q}{\lambda_\gamma} I\big(r_i < -\frac{q}{\lambda_\gamma}\big) + r_i I\big(-\frac{q}{\lambda_\gamma} \le r_i < \frac{1-q}{\lambda_\gamma}\big) + \frac{1-q}{\lambda_\gamma} I\big(r_i \ge \frac{1-q}{\lambda_\gamma}\big).$$

Plugging $\hat{\gamma}$ in (4) produces the ℓ_2 adjusted check loss,

$$\rho_q^M(r) = \begin{cases} (q-1)r - \frac{q(1-q)}{2\lambda_\gamma} & \text{for} \quad r < -\frac{q}{\lambda_\gamma} \\ \frac{\lambda_\gamma}{2}\frac{1-q}{q} r^2 & \text{for} \quad -\frac{q}{\lambda_\gamma} \le r < 0 \\ \frac{\lambda_\gamma}{2}\frac{q}{1-q} r^2 & \text{for} \quad 0 \le r < \frac{1-q}{\lambda_\gamma} \\ qr - \frac{q(1-q)}{2\lambda_\gamma} & \text{for} \quad r \ge \frac{1-q}{\lambda_\gamma}. \end{cases} \tag{5}$$

Figure 1 displays the quadratically adjusted check loss when $q > 0.5$. Note that the modified check loss is continuous and differentiable everywhere. The interval of quadratic adjustment is $(-q/\lambda_\gamma, (1-q)/\lambda_\gamma)$, and we refer to the length of this interval, $1/\lambda_\gamma$, as the "window width". With the loss in (5), we can define the efficient quantile regression estimator $\hat{\beta}^M$ as the minimizer of

$$L_{\lambda_\gamma}(\beta) = \sum_{i=1}^{n} \rho_q^M(y_i - x_i^\top \beta). \tag{6}$$

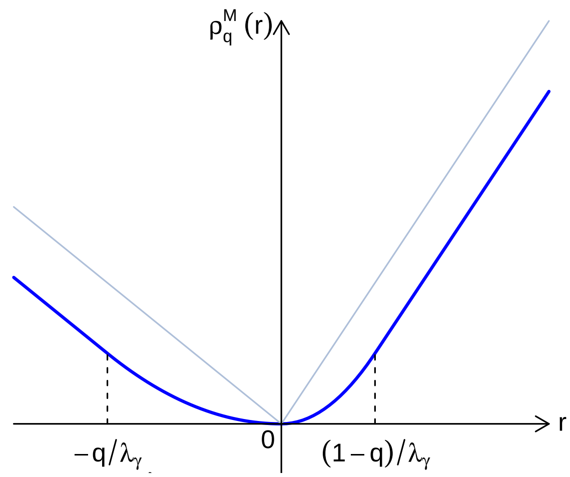

Figure 1: ℓ_2 adjusted check loss in (5) when $q > 0.5$.

When the window width of ρ_q^M is properly chosen, the modified procedure is expected to realize its advantage to the full. See Jung et al. (2010) for the details of the window width selection.

3 Efficient Quantile Regression under Heterogeneous Errors

The method described in Lee et al. (2012) relies on conditional independence of y_i given x_i, and on correct specification of the quantile function. It does not assume that the u_i are identically distributed. As a consequence, the results on consistency and asymptotic normality apply quite broadly. However, one might expect that an estimator that makes use of differences in the distributions of u_i would be more efficient. As we show in this section, this is indeed the case.

There are two main approaches to extending quantile regression from the homogeneous error model to the heterogeneous error model. The first approach directly scales the deviations from the quantile surface, so that one considers $\rho_q(w_i(y_i - x_i^\top\beta))$ in place of $\rho_q(y_i - x_i^\top\beta)$, where the values w_i are used to produce densities for scaled u_i which are locally constant over i. This is our primary focus in this work. A second approach plays off the duality between rescaling residuals and attaching weights to cases. This approach mimics the use of weights in weighted least squares. It also translates into approaches for traditional quantile regression where the linearity of the check loss function allows us to write $\rho_q(w_i(y_i - x_i^\top\beta)) = w_i\rho_q(y_i - x_i^\top\beta)$. However, this equality does not hold when ρ_q is replaced by ρ_q^M which combines both linear and quadratic terms. Nevertheless, there is some evidence that this second approach under ρ_q^M improves upon quantile regression, as shown in the sequel (also see Jung (2010)). We view this second approach mainly as an ad-hoc improvement on traditional techniques.

In the next two subsections, we first describe the scaled efficient QR when the scale factors are known, and then consider the case of unknown scales under a location-scale model.

3.1 Scaled efficient QR with *known* local densities

In this subsection, we provide a consistency result for efficient quantile regression based on the modified loss function in (5). We pursue an approach where the scaling occurs inside the loss function, and note the close connection to Koenker (2005). We retain his notation, describing the scale in terms of (its inverse) a weight.

Allowing a potentially different error distribution for each observation, let $Y_1, Y_2, \ldots$ be independent random variables with cdfs $F_1, F_2, \ldots$, and suppose that each F_i has continuous pdf f_i. Assume that the qth conditional quantile function of Y given x is linear in x and given by $x^\top \beta(q)$, and let ξ_i denote the true quantiles $x_i^\top \beta(q)$. First consider the following regularity conditions:

(C-1) $f_i(\xi)$, $i = 1, 2, \ldots$, are uniformly bounded away from 0 and ∞ at ξ_i.

(C-2) $f_i(\xi)$, $i = 1, 2, \ldots$, admit a first-order Taylor expansion at ξ_i, and $f_i'(\xi)$ are uniformly bounded at ξ_i.

(C-3) There exists a positive definite matrix D_0 such that $\lim_{n\to\infty} n^{-1} \sum x_i x_i^\top = D_0$.

(C-4) There exists a positive definite matrix D_1 such that $\lim_{n\to\infty} n^{-1} \sum f_i(\xi_i) x_i x_i^\top = D_1$.

(C-5) There exists a positive definite matrix D_2 such that $\lim_{n\to\infty} n^{-1} \sum f_i^2(\xi_i) x_i x_i^\top = D_2$.

(C-6) $\max_{i=1,\ldots,n} \|x_i\|/\sqrt{n} \to 0$.

(C-1), (C-3), (C-4), and (C-6) are the conditions considered for the limiting distribution of the standard regression quantile estimator, and (C-5) for the weighted (or scaled) version of it in Koenker (2005), while (C-2) is an additional assumption that we make.

Under conditions (C-1) through (C-6), Theorem 5.1 in Koenker (2005) states that for $\tilde\beta = \arg\min_{\beta\in\mathbb{R}^p} \sum_{i=1}^n w_i \rho_q(y_i - x_i^\top \beta) = \arg\min_{\beta\in\mathbb{R}^p} \sum_{i=1}^n \rho_q(w_i(y_i - x_i^\top \beta))$ with $w_i = f_i(\xi_i)$,

$$\sqrt{n}(\tilde\beta - \beta) \xrightarrow{d} N(0, q(1-q) D_2^{-1}).$$

Koenker (2005) proves the result by using the fact that the limiting behavior of

$$Z_n(\delta) \equiv \sum_{i=1}^n w_i\{\rho_q(u_i - x_i^\top \delta/\sqrt{n}) - \rho_q(u_i)\} = \sum_{i=1}^n \{\rho_q(w_i(u_i - x_i^\top \delta/\sqrt{n})) - \rho_q(w_i u_i)\}$$

determines the limiting distribution of $\tilde{\delta}_n = \sqrt{n}(\tilde{\beta} - \beta)$, where $\tilde{\delta}_n$ minimizes $Z_n(\delta)$ and $u_i = y_i - x_i^\top \beta(q)$.

We define the scaled efficient quantile regression estimator as

$$\tilde{\beta}^M \equiv \arg\min_{\beta \in \mathbb{R}^p} \sum_{i=1}^n \rho_q^M(w_i(y_i - x_i^\top \beta)).$$

Similarly, we will consider the limiting behavior of

$$Z_n^M(\delta) = \sum_{i=1}^n \{\rho_q^M(w_i(u_i - x_i^\top \delta/\sqrt{n})) - \rho_q(w_i u_i)\}. \tag{7}$$

To get a consistent estimator, we set the λ_γ in $\rho_q^M(\cdot)$ to be of the form $c \cdot n^\alpha$, where c is a constant, n is the sample size, and α is a positive constant. Details of the choice of parameters are in Jung et al. (2010) for efficient quantile regression with equal weights $w_i = 1$. Under the same condition on α as in the equal weight setting, we can show that the scaled quantile regression estimator under ρ_q^M is consistent and asymptotically normal when the w_i's are proportional to the local densities around the true quantiles $f_i(\xi_i)$.

Theorem 1. *Under the conditions (C-1) through (C-6), if $\alpha > 1/3$ and $w_i = f_i(\xi_i)$, then*

$$\sqrt{n}(\tilde{\beta}^M - \beta) \xrightarrow{d} N(0, q(1-q)D_2^{-1}).$$

The proof of the theorem is in the appendix. Lee et al. (2012) show that the unscaled version of $\hat{\beta}^M$ defined in (6) has an asymptotic variance of $q(1-q)D_1^{-1}D_0D_1^{-1}$. To compare $\tilde{\beta}^M$ with $\hat{\beta}^M$, we examine their asymptotic variances. Let

$$D = \begin{pmatrix} D_2 & D_1 \\ D_1 & D_0 \end{pmatrix} = \lim_{n\to\infty} n^{-1} \sum_{i=1}^n \begin{pmatrix} f_i^2 & f_i \\ f_i & 1 \end{pmatrix} \otimes x_i x_i^\top, \tag{8}$$

where $\otimes$ represents the Kronecker product. D is a non-negative definite matrix since the matrices in the right hand side of (8) are non-negative definite and the non-negative definiteness is preserved under the Kronecker product, summation and limit operations. Since all D_i in (8) are non-negative definite by (C-3), (C-4), and (C-5), there exists an orthogonal matrix P such that

$$P^\top D P = \begin{pmatrix} D_2 & 0 \\ 0 & D_0 - D_1 D_2^{-1} D_1 \end{pmatrix}.$$

The fact that $D_0 - D_1D_2^{-1}D_1$ is non-negative definite and that D_1 is nonsingular imply that $D_1^{-1}D_0D_1^{-1} - D_2^{-1}$ is also non-negative definite. Consequently, the scaled version, $\tilde{\beta}^M$, is more efficient than $\hat{\beta}^M$.

3.2 Scaled efficient QR with *unknown* linear scales

In practice, the true scales of errors are unlikely to be known, and we will be faced with the task of estimating them on the basis of the data. In many circumstances, estimation of the scales will not change the asymptotic properties of the estimators of β, although we do require a reasonably effective estimator of the scales. In this subsection, we present a result for a location-scale model where both the regression coefficients and the scale of the errors are to be estimated. The model is

$$y_i = x_i^\top \beta + (x_i^\top \tau) u_i, \tag{9}$$

where u_i's are *iid* from distribution F with finite density f. The scale for the ith error is given by $x_i^\top \tau$, leading to the weight $w_i = 1/(x_i^\top \tau)$. Note that this linear heteroscedastic model is a special case of the heterogeneous error model in the previous section when $f_i(y_i) = \frac{1}{\sigma_i} f(u_i + F^{-1}(q))$ with $\sigma_i \equiv x_i^\top \tau > 0$. Some of the regularity conditions are slightly modified and restated below for this special case:

(D-5) There exists a positive definite matrix D_2^* such that $\lim_{n\to\infty} n^{-1} \sum \frac{x_i x_i^\top}{(x_i^\top \tau)^2} = D_2^*$.

(D-6) $\max_{i=1,\dots,n} \|x_i/(x_i^\top \tau)\| = O(n^{1/4})$.

(D-6) strengthens (C-6), and it is used to establish a limiting result when τ is estimated from data to define the linear scales $x_i^\top \tau$. Koenker and Zhao (1994) define the weighted quantile regression estimator $\check{\beta}_\tau$ as $\arg\min_{\beta\in\mathbb{R}^p} \sum_{i=1}^n \rho_q((y_i - x_i^\top \beta)/x_i^\top \tau)$. With any $\sqrt{n}$-consistent estimator of τ up to a scale, $\hat{\tau} = \kappa\tau + O_p(n^{-1/2})$, they show that the asymptotic behavior of $\check{\beta}_\tau$ is the same as $\check{\beta}_{\hat{\tau}}$, where $\hat{\tau}$ is plugged in. That is, $\sqrt{n}(\check{\beta}_{\hat{\tau}} - \beta) \xrightarrow{d} N(0, \frac{q(1-q)}{f^2(F^{-1}(q))} D_2^{*-1})$, which has a smaller asymptotic variance than that of the unweighted version. Note that τ need only be estimated up to an arbitrary scale factor, as the value of κ does not effect the minimization procedure.

Now, with $\rho_q^M(\cdot)$, the modified quantile regression estimator $\check{\beta}_\tau^M$ is defined as

$$\check{\beta}_\tau^M = \arg\min_{\beta\in\mathbb{R}^p} \sum_{i=1}^n \rho_q^M((y_i - x_i^\top \beta)/x_i^\top \tau). \tag{10}$$

With a similar argument as above, a $\sqrt{n}$-consistent estimator of τ up to a scale will maintain the same consistency properties. Thus, we consider

$$\check{\beta}_{\hat{\tau}}^M = \arg\min_{\beta\in\mathbb{R}^p} \sum_{i=1}^n \rho_q^M((y_i - x_i^\top \beta)/x_i^\top \hat{\tau}).$$

Theorem 2. *Assume that $x_i^\top \tau > 0$ for all i and that $\hat{\tau} = \kappa\tau + O_p(n^{-1/2})$ for some scalar κ. Then, under (C-1) through (C-3), (D-5), and (D-6), if $\alpha > 1/3$, then*

$$\sqrt{n}(\check{\beta}_{\hat{\tau}}^M - \beta) \xrightarrow{d} N(0, \frac{q(1-q)}{f^2(F^{-1}(q))} D_2^{*-1}).$$

See the appendix for proof. Again, the above estimator is asymptotically more efficient than the unscaled version, and it achieves the same efficiency as weighted quantile regression estimates.

The analogy, that weighted least squares is to least squares as weighted quantile regression is to quantile regression, is telling. In addition to motivating both scaled and weighted modified quantile regression, our simulation reveals the type of improvement we expect when we incorporate more features of the model into the analysis. The analogy is farther reaching. If we have approximately correct scales, we expect to see improvement over the unscaled analysis. This suggests the use of a relatively simple model for the scales, perhaps coupled with a more complex model for the quantile function.

3.3 Estimation of scale factors

In this section, we outline two practical methods for estimation of the scale factors. Many other strategies are possible and merit future investigation. All such methods require some measure that can reveal heteroscedasticity in the data if it is present. With this in mind, and motivated by location-scale models we consider a robust estimate of scale, such as the interquartile range (conditional on x), $IQR(x)$. Quantile regression provides a means of estimating this scale function. The fitted IQR at ith observation is $\widehat{IQR}_i = x_i^\top(\hat{\beta}^{.75} - \hat{\beta}^{.25})$, with the $\hat{\beta}^q$ coming from a pair of quartile regression fits of the data. This allows us to define $\hat{w}_i = 1/\widehat{IQR}_i$ and to set $\hat{\tau} = \hat{\beta}^{.75} - \hat{\beta}^{.25}$ for a connection to the theoretical results in the previous subsection. Since $\hat{\beta}^{.25}$ and $\hat{\beta}^{.75}$ are $\sqrt{n}$-consistent, $\hat{\tau}$ will satisfy the required asymptotic rate in Theorem 2. We note that the quantile regression fit used to generate the $\hat{\beta}^q$ can be replaced with a modified quantile regression fit from $\rho_q^M(\cdot)$. Our subsequent investigations suggest the effectiveness of this modification.

A second approach focuses on departures of observations from the median regression. Specifically, scale factors can be obtained with the following steps. First, fit a median regression and store the absolute residuals, $|r_i|$. Second, again fit a median regression of $|r_i|$ on x_i, and obtain a fitted value, $\hat{m}_i$, for each case. Finally set the weight to $w_i = 1/\hat{m}_i$ and proceed to build a scaled quantile regression. In this example, we implicitly set $\hat{\tau}$ to be the estimate of regression coefficient at $q = 0.5$ for $(x_i, |r_i|)$. The intuitive explanation of this method is that, asymptotically, the median regression surface for $|r_i|$ given x_i captures half of the absolute residuals below the surface and half above the surface. If the model is correct, this holds for all x. Taking twice this spread gives an interval which (asymptotically) captures half of the probability density at each x_i. In the case of a symmetric error distribution, $2\hat{m}_i$ provides an estimate of IQR_i. The asymptotic properties of this method are more difficult to establish, as they depend upon a pair of quantile regressions which removes the independence of the $|r_i|$.

There are many other ways to estimate the scale factors for these procedures. A natural approach is to attempt to estimate each f_i directly, with a method that is local in nature, avoiding the assumption of a location-scale family for the y_i. While direct estimates of the local density–perhaps kernel density estimates–have great appeal, they also suffer from relatively poor rates of convergence. The theoretical treatment of the previous subsection which echoes results in Koenker and Zhao (1994) suggests the need for a $\sqrt{n}$-consistent estimator.

Our philosophical perspective is to seek a stable estimate of the scale (hence the weights) that is reasonably close to consistent, thereby obtaining the bulk of the improvement provided by adjusting for different scales. This pushes us in the direction of relatively robust estimates of the scale and toward approximations such as the location-scale model. It also suggests that we might tweak the estimates in the direction of stability. For example, we might use methods that force separation among the quantiles to prevent the estimated scale from being too close to 0: see He (1997). Or, we might begin with the estimated scales and then shrink them toward uniform scales, yielding a stable procedure which can be considered as midway between unscaled and scaled quantile regression.

Once we have estimated scales in hand, we can apply them to standard quantile regression (QR) and modified quantile regression (QR.M), to produce the corresponding scaled versions. The effectiveness of the scaled procedures is demonstrated in the simulations and real data analyses in the following sections. Our belief is that any appropriate scale estimator that is able to account for the variations in scale will be advantageous. The effect of adjusting for scales in some fashion is strong enough that the wide range of procedures we have investigated all show a benefit. The aforementioned methods are but two examples of how to produce estimates of scale factors.

4 Simulations

4.1 Univariate case

A simple simulation with a heterogeneous error model given below is considered, first with fixed x. The heterogeneous error model is given by

$$y_i = \beta_0 + \beta_1 x_i + x_i u_i,$$

where the u_i's are *iid* standard normal, $(\beta_0, \beta_1)^\top = (1, 2)^\top$, and x takes one of three values from the set $\{1, 2, 3\}$. 2000 data sets are generated from the model for each of two sample sizes, $n = 300$ and $n = 900$. In each case, 1/3 of the sample assumes each of the three possible covariate values.

Five different models are fit to the data; standard quantile regression (QR), weighted QR (WQR), modified quantile regression (QR.M), and two versions of QR.M (WQR.M and WQR.M2) that account for heterogeneous errors. For these versions, WQR.M2 follows the scaling approach and minimizes $\sum_{i=1}^n \rho_q^M(w_i(y_i - x_i^\top \beta))$, while WQR.M results from minimization of the weighted sum $\sum_{i=1}^n w_i \rho_q^M(y_i - x_i^\top \beta)$.

The estimate of the weight (or scale) for cases with covariate x is derived from an estimate of the IQR at x. The IQR can be estimated in a robust fashion, and, for the normal distributions, the IQR is about 1.35σ. Specifically, we find the fitted quartile surfaces ($q = 0.75$ and $q = 0.25$) with unweighted (or unscaled) linear quantile regressions (using either QR or QR.M), and then set $\widehat{IQR}(x)$ equal to the difference between these two surfaces at x. The fitted IQR values from QR are used to derive weights for WQR while those from QR.M are used for WQR.M. It is natural to take the scaling approach and minimize $\sum_{i=1}^{n} \rho_q^M((y_i - x_i^\top \beta)/\widehat{IQR}_i)$. Alternatively, the value $1/\widehat{IQR}_i$ can be used as a weight for the ith observation. Note that these two calculations become equivalent for $\rho_q(\cdot)$. In this way, we fit WQR, WQR.M, and WQR.M2 with $w_i = 1/\widehat{IQR}_i$ and compare them to their unweighted (or unscaled) counterparts. The true qth quantile regression line passes through the three points $(1, 3 + \Phi^{-1}(q))$, $(2, 5 + 2\Phi^{-1}(q))$, and $(3, 7 + 3\Phi^{-1}(q))$ for each $0 < q < 1$, where $\Phi(\cdot)$ is cumulative distribution function of standard normal distribution.

The mean squared error (MSE) between the true quantile line and each of the five fitted lines is computed for the 2000 data sets for each quantile from $q = 0.1$ to 0.5. Table 1 reveals an interesting pattern. The weighted (or scaled) versions (WQR, WQR.M, and WQR.M2) perform better than the unweighted (or unscaled) counterparts in nearly every case. We suspect that the lone reversal of this pattern for QR and WQR is due to simulation variation. Surprisingly, QR.M outperforms WQR when $q > 0.3$. In these instances, WQR.M performs even better than QR.M.

Table 1: Point estimate of MSE and standard error of the estimate in parentheses (multiplied by 1000), based on 2000 replicates with $n = 300$, and $n = 900$, at selected quantiles.

Method	$q = 0.1$	$q = 0.2$	$q = 0.3$	$q = 0.4$	$q = 0.5$
			$n = 300$		
QR	85.71 (2.45)	60.64 (1.74)	51.75 (1.43)	48.44 (1.33)	47.03 (1.30)
WQR	82.91 (2.40)	58.12 (1.69)	50.27 (1.43)	47.64 (1.31)	46.08 (1.27)
QR.M	88.81 (2.40)	58.63 (1.60)	45.60 (1.24)	41.07 (1.14)	39.05 (1.10)
WQR.M	80.11 (2.22)	54.61 (1.53)	43.65 (1.21)	39.75 (1.19)	37.98 (1.07)
WQR.M2	86.72 (2.35)	55.94 (1.59)	43.00 (1.18)	37.47 (1.03)	35.20 (0.99)
			$n = 900$		
QR	28.56 (0.76)	20.15 (0.56)	17.32 (0.47)	15.61 (0.42)	15.50 (0.41)
WQR	27.39 (0.75)	19.47 (0.55)	16.97 (0.46)	15.22 (0.41)	15.18 (0.41)
QR.M	28.85 (0.74)	19.33 (0.54)	15.84 (0.43)	13.87 (0.37)	13.19 (0.36)
WQR.M	27.49 (0.71)	18.25 (0.51)	15.25 (0.42)	13.54 (0.37)	12.95 (0.35)
WQR.M2	28.43 (0.80)	18.62 (0.54)	15.01 (0.42)	12.95 (0.35)	12.06 (0.32)

In many data analyses, the covariates are best viewed as arising at random. Our second simulation examines this case. We generate the x_i independently from the uniform distribution over the set $\{1, 2, 3\}$.

Under the heterogeneous model mentioned above, 2000 data sets are simulated with a sample size of 300. Standard normal and exponential with mean one error distributions are considered. Again, the MSE is employed to measure the accuracy of the five fitted models. Table 2 summarizes the simulation results and illustrates that all of the MSEs for the weighted (or scaled) methods (WQR, WQR.M, and WQR.M2) are smaller than their counterparts (QR, QR.M), by 3-5% on average. The MSE for WQR.M and WQR.M2 is less than that for WQR, except for $q = 0.1$, under the normal distribution. This assessment demonstrates a clear benefit from use of the weights or scales.

Table 2: Point estimate of MSE and standard error of the estimate in parentheses (multiplied by 1000), based on 2000 replicates with $n = 300$ under standard normal and Exponential(1) error distributions at selected quantiles.

Method	$q = 0.1$	$q = 0.2$	$q = 0.3$	$q = 0.4$	$q = 0.5$
			N(0,1)		
QR	77.70 (2.05)	55.19 (1.42)	46.53 (1.17)	44.57 (1.09)	43.74 (1.08)
WQR	75.89 (1.93)	54.07 (1.39)	45.08 (1.16)	42.62 (1.04)	41.91 (1.04)
QR.M	79.49 (1.95)	50.55 (1.26)	40.67 (1.00)	36.67 (0.91)	34.96 (0.85)
WQR.M	73.57 (1.81)	47.55 (1.20)	38.71 (0.97)	35.11 (0.87)	33.62 (0.83)
WQR.M2	79.23 (2.01)	49.33 (1.28)	38.49 (0.97)	34.00 (0.84)	32.29 (0.79)
			Exp(1)		
QR	3.26 (0.09)	6.80 (0.19)	11.34 (0.30)	17.70 (0.49)	27.34 (0.72)
WQR	3.14 (0.09)	6.52 (0.18)	10.96 (0.28)	17.06 (0.46)	26.36 (0.70)
QR.M	2.60 (0.08)	5.69 (0.16)	9.74 (0.27)	15.79 (0.43)	26.87 (0.71)
WQR.M	2.46 (0.08)	5.37 (0.15)	9.17 (0.26)	14.72 (0.40)	24.35 (0.64)
WQR.M2	2.42 (0.08)	5.36 (0.15)	9.23 (0.26)	15.22 (0.43)	26.70 (0.75)

4.2 Multivariate case

A more realistic case involves more than one covariate. To examine the performance of our method in this setting, we consider a model with 8 covariates ($p = 8$) and sample size 500:

$$y_i = \beta_0 + x_i^\top \beta + |x_i^\top \beta| u_i,$$

where $u_i \overset{iid}{\sim} N(0, \sigma^2)$. Let x_i be iid copies of X, where X follows the multivariate normal distribution with mean $5 \cdot 1_p$, and variance $\Sigma = [\ \rho^{|i-j|}\ ]$ with $\rho = 0.5$, and assume that X and u are independent. To measure the accuracy of the estimates of the

qth quantile, we use the MSE for the fitted surface, integrated over the distribution of X. The MSE can be expressed as

$$MSE(\hat{\beta}) = E^{\hat{\beta},X}\{(X^\top\hat{\beta}^q + \hat{\beta}_0^q) - (X^\top\beta + \beta_0 + |X^\top\beta| \cdot \sigma\Phi^{-1}(q)\}^2,$$

where $\hat{\beta}^q$ and $\hat{\beta}_0^q$ are estimates of β and β_0 at the qth quantile, respectively, and $\sigma\Phi^{-1}(q)$ (or, shortly $\Phi_{\sigma,q}^{-1}$) is the qth quantile of $N(0, \sigma^2)$. The MSE does not depend on the value of β_0 used in the simulation, and so, without loss of generality, we set $\beta_0 = 0$ and consider three scenarios of regression coefficients: i) *Sparse*: $\beta = (5, 0, 0, 0, 0, 0, 0, 0)^\top$, ii) *Intermediate*: $\beta = (3, 1.5, 0, 0, 2, 0, 0, 0)^\top$, and iii) *Dense*: $\beta = (2, 2, 2, 2, 2, 2, 2, 2)^\top$. To keep the signal-to-noise ratio comparable across the three scenarios, we set the value of σ such that $Var(|X^\top\beta|u) = 1$. In other words, $\sigma^2 = 1/\beta^\top(\Sigma + \mu\mu^\top)\beta$.

Under each scenario, 500 data sets are generated, each with sample size $n = 500$. When $X^\top\beta$ is substantial and positive, which is practically the case under our settings of design matrix and parameter values, we can drop the absolute value about $|X^\top\beta|$, which makes essentially no difference in the calculation of MSE. The following derivation shows that an estimate of the MSE can be computed from the estimates of β without involving the sampled value of X:

$$\begin{aligned} & MSE(\hat{\beta}) \\ = \; & E^{\hat{\beta},X}\{X^\top(\hat{\beta}^q - \beta - \beta\Phi_{\sigma,q}^{-1})\}^2 + 2E^{\hat{\beta},X}[X^\top(\hat{\beta}^q - \beta - \beta\Phi_{\sigma,q}^{-1})(\hat{\beta}_0^q - \beta_0)] + E^{\hat{\beta}}(\hat{\beta}_0^q - \beta_0)^2 \\ = \; & E^{\hat{\beta}}[(\hat{\beta}^q - \beta)^\top(\Sigma + \mu\mu^\top)(\hat{\beta}^q - \beta)] - 2E^{\hat{\beta}}[(\hat{\beta}^q - \beta)^\top(\Sigma + \mu\mu^\top)\beta\Phi_{\sigma,q}^{-1}] \\ & + (\Phi_{\sigma,q}^{-1})^2\beta^\top(\Sigma + \mu\mu^\top)\beta + 2E^{\hat{\beta}}[\mu^\top(\hat{\beta}^q - \beta - \beta\Phi_{\sigma,q}^{-1})(\hat{\beta}_0 - \beta_0)] + E^{\hat{\beta}}(\hat{\beta}_0 - \beta_0)^2, \end{aligned}$$

where $\mu = 5 \cdot 1_8$ in our simulation setting and thus $\mu\mu^\top = 25J$ with J denoting the matrix of ones.

For the estimates of weight, $w_i = 1/|x_i^\top\beta|$, first we fit a standard quantile regression at $q = 0.5$, then obtain residuals and fitted values. Then, treating the absolute value of the residuals as a response variable and the fitted values as an explanatory variable, we fit another standard median regression from which the fitted values are used as $1/\hat{w}_i$. Once we obtain the estimates of weight (and scale), they are used for WQR, WQR.M, and WQR.M2. Note that this approach for estimation of weight (or scale) is different from that in Section 4.1.

Table 3 shows the MSE values for five methods at selected quantiles. WQR shows uniform improvement over QR, and QR.M shows uniform improvement over WQR. In all cases, the two best performing methods are WQR.M and WQR.M2, with a relatively even split between the two for smallest MSE. The results demonstrate the value of both modification of the check loss and the use of scale or weight across the sparse, intermediate, and dense scenarios.

Table 3: Point estimate of integrated MSE and standard error of the estimate in parentheses (multiplied by 1000), based on 500 replicates with $n = 500$ at selected quantiles.

Method	$q = 0.1$	$q = 0.25$	$q = 0.5$	$q = 0.8$	$q = 0.9$
			Sparse		
QR	33.50 (1.016)	20.22 (0.620)	17.48 (0.529)	23.23 (0.754)	31.89 (1.074)
WQR	31.95 (1.001)	19.50 (0.611)	16.74 (0.499)	21.83 (0.713)	29.78 (1.009)
QR.M	31.60 (0.971)	17.10 (0.548)	13.92 (0.453)	20.33 (0.667)	30.19 (1.016)
WQR.M	29.51 (0.916)	16.10 (0.518)	12.89 (0.419)	18.93 (0.621)	27.62 (0.917)
WQR.M2	29.34 (0.974)	16.03 (0.517)	13.54 (0.433)	19.22 (0.655)	28.08 (0.950)
			Intermediate		
QR	39.79 (1.125)	23.86 (0.649)	21.08 (0.553)	27.63 (0.797)	38.14 (1.133)
WQR	39.03 (1.078)	23.44 (0.634)	20.40 (0.529)	27.00 (0.774)	37.06 (1.094)
QR.M	36.92 (1.052)	20.19 (0.561)	16.83 (0.479)	23.78 (0.693)	35.41 (1.054)
WQR.M	36.17 (1.055)	19.79 (0.547)	16.23 (0.455)	23.22 (0.677)	34.18 (1.029)
WQR.M2	35.78 (1.031)	19.82 (0.552)	16.33 (0.463)	23.07 (0.679)	34.55 (1.039)
			Dense		
QR	53.26 (1.235)	32.26 (0.704)	27.65 (0.554)	36.49 (0.825)	51.61 (1.186)
WQR	52.49 (1.208)	31.67 (0.676)	27.37 (0.544)	36.05 (0.817)	51.13 (1.170)
QR.M	47.40 (1.078)	27.00 (0.573)	22.26 (0.456)	31.40 (0.697)	46.62 (1.067)
WQR.M	46.81 (1.089)	26.70 (0.569)	21.86 (0.443)	31.11 (0.699)	46.20 (1.063)
WQR.M2	47.08 (1.082)	27.06 (0.574)	22.01 (0.468)	30.65 (0.682)	46.49 (1.058)

5 Empirical Examples

5.1 GDP growth data

In this section, we apply the heterogeneous error model developed in Section 3 to an economic growth data set. The data consist of 161 observations on determinants of cross-country gross domestic product (GDP) growth rates and were used in Koenker and Machado (1999). The first 71 observations are on the period 1965-75, the remainder on 1975-85. There are 13 covariates and one response variable, "Annual Change Per Capita GDP". The response variable indicates the rate of annual GDP growth with most of the observed values between -0.05 and 0.05. Our analysis focuses on the relationship between one of the covariates, "% of Female High School Graduate or More" and the response variable. The five methods (QR, WQR, QR.M, WQR.M, and WQR.M2) are applied to the data, and the weights/scales are estimated using IQR in Section 4.1. Before analyzing the data, we removed two extreme outliers because their inclusion resulted in negative weights under QR, as the fitted first quartile exceeded the fitted third quartile within the range of the data. There was no such problem with QR.M, reaffirming a pattern observed elsewhere, that QR.M reduces crossings of the fitted quantile surfaces.

The top panels in Figure 2 show the fitted lines from the four methods. First, QR and WQR produce quite different fits, especially for the lower and upper quantiles, while QR.M and WQR.M yield similar fits. We conjecture that this arises from the poor fit of QR (this view is supported by the simulation study). The $\widehat{IQR}$ or the fitted weights provide a clearer view of this. The rightmost observation in Figure 2 has the largest weight for both WQR and WQR.M. However, the QR fit produces a weight of 0.4883 for this observation (with the weights normalized to sum to one, this gives a mean weight of $1/159 \approx 0.0063$) which is about 78 times the mean weight and almost two hundred times larger than the smallest weight. The weight for this observation from QR.M is only 0.0265, approximately four times the mean weight. We believe that the more moderate swing in weights is preferable for this data set, and so prefer the weights generated by QR.M to those generated by QR.

The eventual fits of the models under WQR and WQR.M show only minor differences for small values of the covariate, but big differences for large values of the covariate. The estimated differences between the 10th and 90th percentiles when no females graduate from high school are 0.075927 and 0.075280, while the estimated difference when 50% of females have graduated from high school are 0.002497 and 0.025519. We find the modest spread estimated under the WQR.M method to be far more plausible than the near-degeneracy of 80% of the distribution estimated under the WQR method.

To further compare QR.M and WQR.M in terms of prediction accuracy, we used cross validation. Given the modest size of the data set, we used 1000 repetitions of 5-fold cross validation. For each repetition, the data were randomly partitioned into five sets (or folds) of near equal size. The model was fit to four of the folds and with the fifth fold reserved for validation. This was repeated with each fold serving as the validation data. To create the cross validation score, we used the defining loss function for a quantile, the check loss function. The parameters in the check loss were adjusted to allow examination of a number of quantiles. The mean discrepancy between the fitted and observed values appears as a single dot in Figure 3 with prespecified q, giving us 5000 points in each panel. The figure conveys a clear advantage for the weighted method, with the bulk of the points falling below the 45 degree line. A numerical summary of the cross validation score (CV score) appears in Table 4. In a similar way, WQR.M2 is also compared to QR.M in Figure 4.

As explained in Section 4, weights (or scales) are estimated from the fitted IQR with unweighted QR or QR.M for fitting the weighted (or scaled) counterparts at q = (0.25, 0.5, 0.75). For the more extreme quantiles $q = 0.1$ and 0.9, we have used the reciprocal of the fitted distance between $q = 0.9$ and 0.1 in place of the fitted IQR. The weights or scales are then used for WQR, WQR.M and WQR.M2 as before. Estimating the weights or scales with a range that extends beyond the upper and

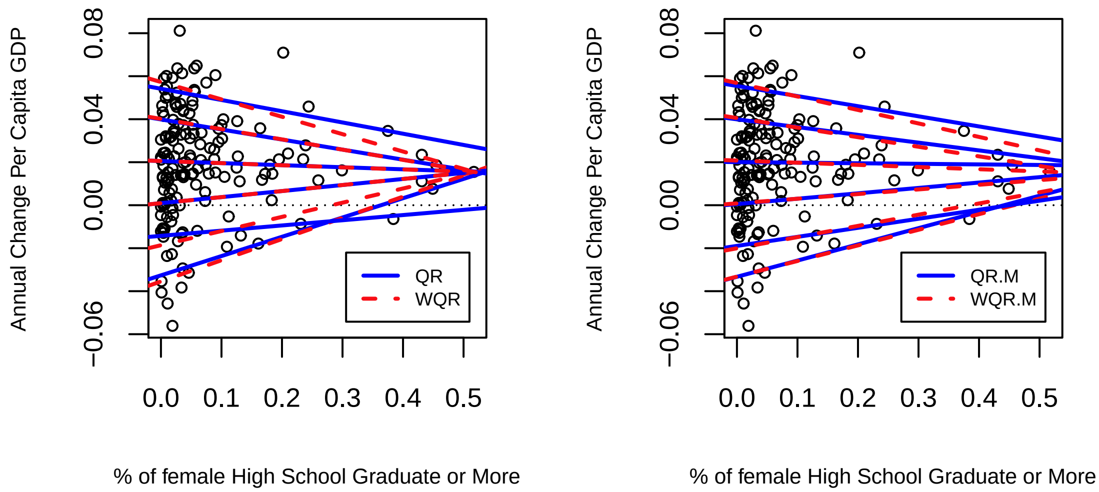

Figure 2: The two panels show quantile regression lines estimated by QR and WQR (left) and by QR.M and WQR.M (right) at $q = (0.05\ , 0.1, 0.25, 0.5, 0.75,$ and $0.9)$.

lower quartiles generally improves the fits, (equivalently reduces the CV score) at extreme quantiles such as $q = 0.1$ and 0.9.

5.2 Corrected Boston Housing data

The Boston Housing data set was originally examined by Rubinfeld (1978) to detect social and environmental factors that affect house prices in the Boston Metropolitan area. The data set is composed of 506 census tracts (observations) with 14 variables where median house value is considered as the response variable. Gilley and Pace (1996) provide details of corrections made to some of the data. The response variable and 13 explanatory variables are *CMEDV* (corrected median values of owner-occupied housing in USD 1000), *CRIM* (crimes per capita), *ZN* (proportion of residential land zoned for lots over 25,000 sqft), *INDUS* (proportion of non-retail business acres per town), *CHAS* (a factor with levels 1 if tract borders Charles River; 0 otherwise), *NOX* (nitric oxides concentration in parts per 10 million), *RM* (average number of rooms per dwelling), *AGE* (proportion of owner-occupied units built prior to 1940), *DIS* (weighted distance to five Boston employment centers), *RAD* (index of accessibility to radial highways), *TAX* (full-value property-tax rate per USD 10,000), *PTRATIO* (pupil-teacher ratio), B $(1000(AA - 0.63)^2$ where AA is the proportion of African American), and *LSTAT* (percentage values of lower status population).

To begin the analysis, we explored the data set by fitting weighted and unweighted

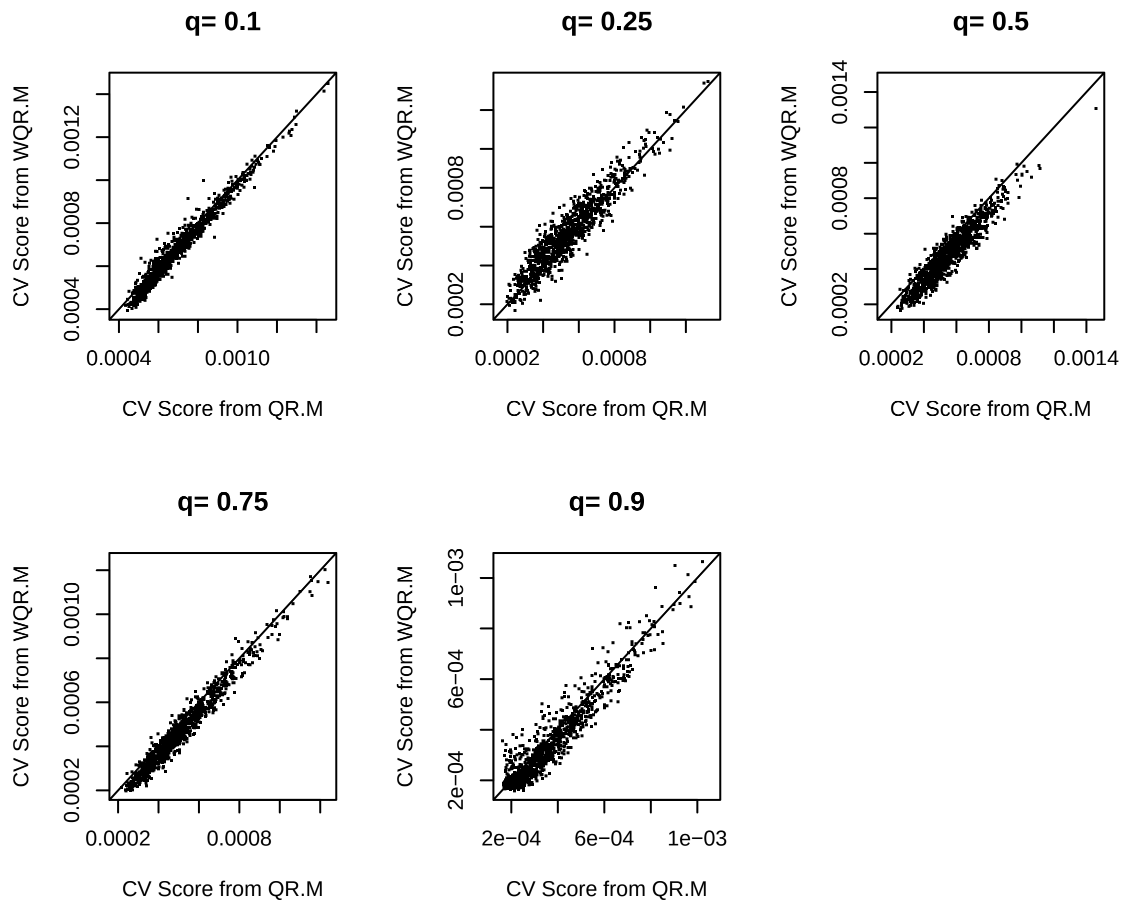

Figure 3: CV scores by QR.M and WQR.M at $q = (0.1, 0.25, 0.5, 0.75, \text{and } 0.9)$ with a line of slope 1 from "Annual Change Per Capita GDP" data. Weights are estimated from IQR at $q = (0.25, 0.5, 0.75)$, while the distance between 0.9 and 0.1 fitted lines is used to estimate the weights at $q = (0.1, 0.9)$.

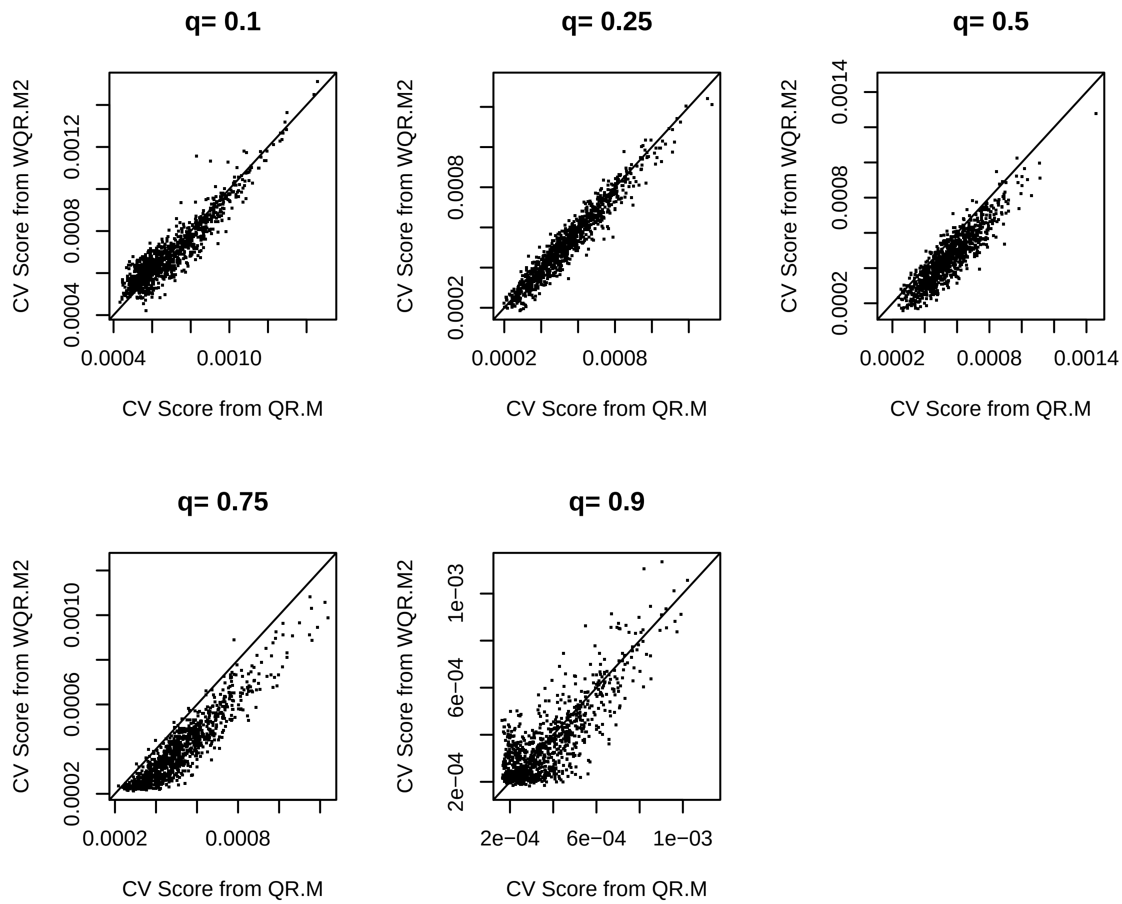

Figure 4: CV scores by QR.M and WQR.M2 at q = (0.1, 0.25, 0.5, 0.75, and 0.9) with a line of slope 1 from “Annual Change Per Capita GDP” data. Weights are estimated from IQR at q = (0.25, 0.5, 0.75), while the distance between 0.9 and 0.1 fitted lines is used to estimate the weights at q = (0.1, 0.9).

Table 4: Point estimate of mean cross validated score (CV score) and standard error of the estimate in parentheses (multiplied by 10,000), based on 1000 different splits of "Annual Change Per Capita GDP" data. CV score is measured by check loss function.

Method	$q = 0.1$	$q = 0.25$	$q = 0.5$	$q = 0.75$	$q = 0.9$
QR	7.147 (0.0678)	5.959 (0.0948)	6.082 (0.0621)	8.298 (0.1024)	7.801 (0.0970)
WQR	7.189 (0.0674)	6.034 (0.1009)	5.204 (0.0598)	5.347 (0.0902)	7.946 (0.0942)
QR.M	6.901 (0.0535)	5.403 (0.0574)	5.567 (0.0484)	5.193 (0.0523)	3.626 (0.0519)
WQR.M	6.689 (0.0545)	5.647 (0.0582)	4.847 (0.0491)	4.812 (0.0536)	3.429 (0.0521)
WQR.M2	6.911 (0.0482)	5.207 (0.0569)	4.599 (0.0479)	3.989 (0.0468)	3.515 (0.0504)

quantile regression models. This led us in the same direction as Rubinfeld (1978), and we transformed a number of variables for the subsequent analysis. We used $\log(CMEDV)$, $\log(DIS)$, $\log(RAD)$, $\log(LSTAT)$, NOX^2, and RM^2. Since ZN and $INDUS$ are rarely significant at various quantiles, they were excluded and one gigantic outlier was removed before the data analysis. In shorthand notation, the final model we build is,

$$\begin{aligned} \log(CMEDV) = \; & CRIM + CHAS + NOX^2 + RM^2 + AGE + \log(DIS) + \\ & \log(RAD) + TAX + PTRATIO + B + \log(LSTAT). \end{aligned} \tag{11}$$

The model is fit with five procedures, QR, WQR, QR.M, WQR.M, and WQR.M2, for q ranging from 0.1 to 0.9. Weights (or scales) are estimated in two steps. First, a standard median regression is fit, and the fits and residuals are obtained. Second comes another single-predictor median regression treating the fit as the independent variable and the absolute value of the residual as the response. Use of the fit as a single predictor stabilizes the second regression, leading to more stable weights (and scales). The weights are taken to be the reciprocals of the fitted values from the second regression.

To compare the performance of the five methods of fitting the model, prediction accuracy is examined through 5-fold cross validation as in Section 5.1. A summary of 500 different partitions of the data is given in Table 5.

In overall terms, the weighted and scaled versions perform better than their counterparts, although there are some exceptions. These exceptions appear to be driven by a departure of our implicit model for the scale of the error distribution. The actual data show a departure from the linear relationship we have used for the scale, and so a more flexible model for weights and scales is desirable. Additionally, the appropriate local scale seems to depend on q, the quantile under examination. This suggests the need for a more refined method for determining the weights and scales. We do not pursue such a method here, but do indicate directions to explore in the discussion

Table 5: Point estimate of mean cross validated score (CV score) and standard error of the estimate in parenthesis (multiplied by 10,000), based on 500 different splits of "Corrected Boston Housing" data. CV score is measured by check loss function.

Method	$q = 0.05$	$q = 0.1$	$q = 0.25$	$q = 0.5$	$q = 0.7$	$q = 0.95$
QR	187.8 (0.25)	304.2 (0.22)	537.5 (0.33)	717.5 (0.45)	699.4 (0.55)	374.1 (0.68)
WQR	186.6 (0.22)	304.0 (0.24)	538.2 (0.33)	715.9 (0.46)	697.8 (0.54)	373.5 (0.69)
QR.M	187.2 (0.23)	303.5 (0.22)	535.5 (0.33)	714.2 (0.45)	694.3 (0.53)	370.8 (0.65)
WQR.M	185.7 (0.21)	303.5 (0.23)	535.2 (0.31)	712.1 (0.42)	695.7 (0.52)	369.4 (0.67)
WQR.M2	185.2 (0.21)	302.6 (0.22)	536.0 (0.32)	714.0 (0.43)	695.1 (0.52)	370.4 (0.67)

section. In any event, even with misspecification of the weights (or scales), weighting (and scaling) produces an overall net benefit.

A second feature of Table 5 is even more striking. The fits from the l_2 adjusted check loss (QR.M and WQR.M) show uniformly smaller CV score values than that from the original check loss (QR and WQR, respectively). Modification of the check loss to enhance efficiency is undoubtedly worthwhile.

To further investigate the effect of weights and scales on the fit of the model, the estimated coefficients from the four methods are drawn in Figure 5 for a range of quantiles from 0.1 to 0.9. The (transformed) covariates are standardized before fitting the above model so that the coefficients are comparable. The coefficients under the various fits truly differ. Some of the more noticeable differences appear at the extreme quantiles of RM^2, $\log(RAD)$, and $\log(LSTAT)$. Another change is that the extent of negative effects on house price are alleviated in both WQR and WQR.M (compared to QR and QR.M, respectively) when the house price is high. These systematic differences that cut across both WQR and WQR.M suggest that these qualitative aspects of the analysis which are heavily tied to substantive interpretation of the data are directly related to the success of the models in capturing, however imperfectly, the heteroscedasticity which is present in the data.

6 Discussion

We have proposed a new quantile regression method for analyzing heteroscedastic data, which extends the methodology developed in Lee et al. (2012) to allow for heterogeneity in the error distributions. Asymptotic results establish the large-sample behavior of the method, placing it on a firm theoretical footing. Details of implementation have been considered, and a practical method to implement the technique has been proposed. The success of the method relative to standard quantile regression has been shown through simulation studies and an examination of a data set. We

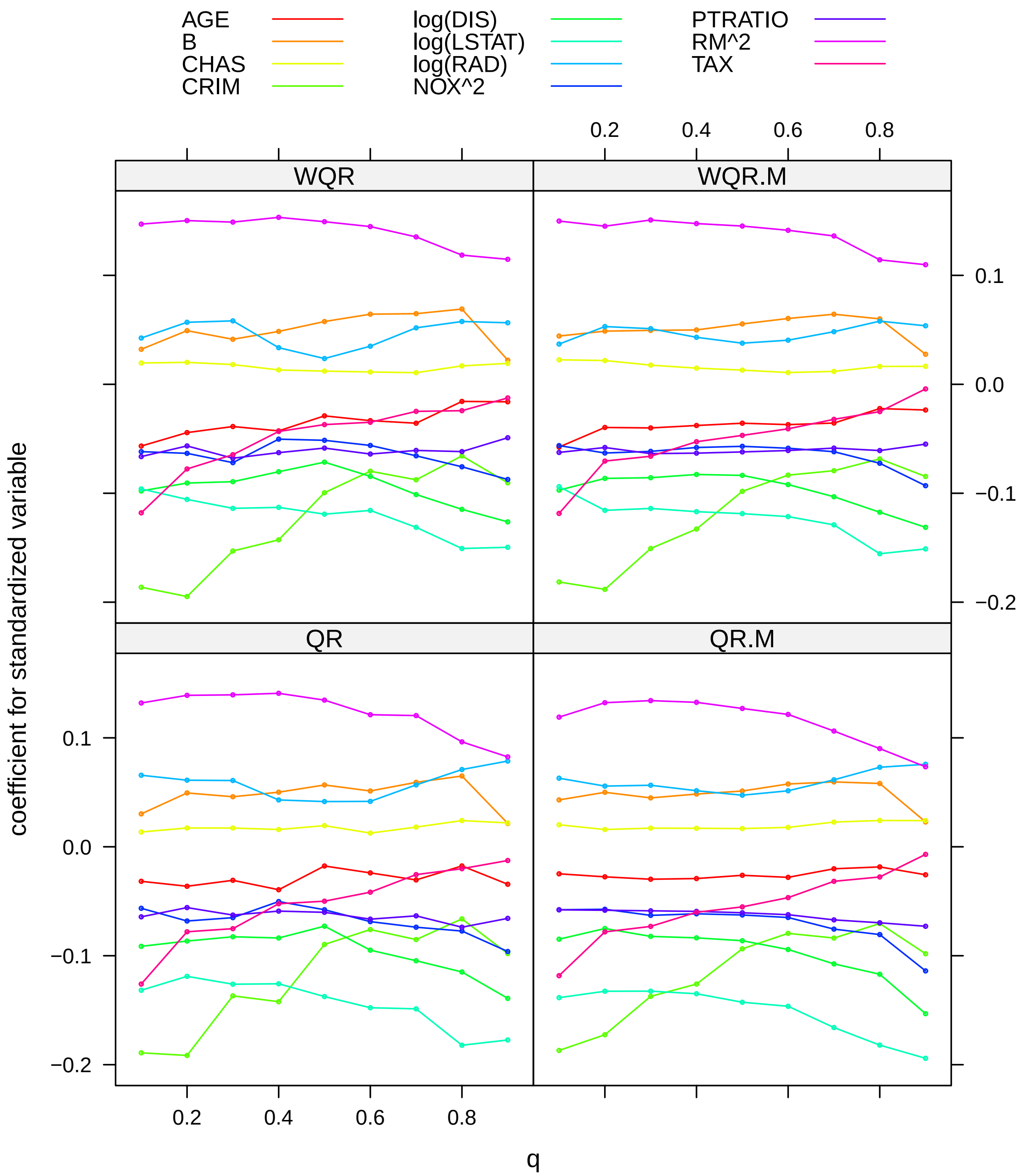

Figure 5: Estimated coefficients of the standardized covariates for Boston Housing data at various quantiles ($q = 0.1, ..., 0.9$) using QR, WQR, QR.M and WQR.M.

attribute the success of WQR.M and WQR.M2 relative to WQR partly to improved estimation of weights (and scales) with QR.M instead of QR and partly to the modification of the check loss function. There are many variations on the general theme presented here. We indicate a few promising directions.

The technique we have developed relies on a two-stage strategy, where the scale of the error distribution is estimated in a first stage, yielding scale factors used to estimate the quantile surface in the second stage. In our first scale-estimation scheme, an iterative algorithm can be used in the first stage, wherein scales are estimated, WQR.M is used to estimate the upper and lower quartile surfaces, with scales derived from these estimated surfaces. The second stage would then make use of these (more accurate) scales for final estimation of the quantile surfaces. In our second scale-estimation scheme, estimation of the median regression for the absolute residuals can be replaced by WQR.M estimation of the regression, again leading to more accurate estimates of the scales.

Throughout the simulations and data analysis, we have chosen to use a common scale estimate for estimation of the various quantile surfaces. This is in keeping with the notion of the errors following a scale family. Alternative approaches exist and would be appropriate when the assumption of a scale family is noticeably violated. Our methods can be focused directly on the scale near a particular quantile, say the upper quartile. For the first method, we can use, say, the difference between the estimated 90th and 60th percentile surfaces to obtain an estimate of the average density over this range of quantiles. With larger sample sizes, the band about the upper quartile can be shrunk, asymptotically focusing on the average density in a vanishingly small region about the upper quartile. The second method can also be focused on a quantile. Again considering a regression for the upper quartile, the two steps in the procedure can be altered as follows. First, fit the upper quartile surface to the data (x_i, y_i) with QR (or QR.M). Extracting the residuals, r_i, from this fit provides an estimate of local deviations from the quartile surface. Second, run QR (or QR.M) with the absolute residuals as the response, using $(x_i, |r_i|)$ as the data for the regression. The choice of quantile in this second regression determines how closely the scale estimate focuses on the density around the upper quartile. A regression for the 20th percentile extracts a scale estimate from one fifth of the distribution of the residuals; a regression for the 10th percentile extracts a scale estimate from one tenth of the distribution of residuals, and so on. Larger sample sizes support stable estimates of low percentile surfaces in this second regression, allowing us to better capture the densities near the upper quartile surface.

In this work, we have focused exclusively on linear quantile regression, but the proposed method can be readily extended to nonlinear quantile regression with heterogeneous errors. To do so, we need only modify the stage where the scale (as a function of covariates) is estimated. Following the path laid out here, we could estimate the upper and lower quartile surfaces through nonlinear quantile regression and obtain scale factors that vary locally, using these scales in the second stage.

On the implementation side, we fit the standard quantile regression by `rq` in the R package `quantreg` and its weighted version with an additional input argument for the weights. The modified quantile regression method was directly implemented by specifying the derivative of (5) in the `rlm` function in the `MASS` package. The `rlm` function allows both scaling and weighting to treat cases differently: one through case weights (`wt.method="case"`) and the other through scales (`wt.method ="inv.var"`). Again, simply adding one argument for the weights (or scales) produces the weighted (or scaled) counterpart which we have found enhances the accuracy of estimation.

Appendix

Proof of Theorem 1.
$Z_n^M(\delta) = \sum_{i=1}^n \{\rho_q^M(w_i(u_i - x_i^\top \delta/\sqrt{n})) - \rho_q(w_i u_i)\}$ can be decomposed into

$$\sum_{i=1}^n \{\rho_q^M(w_i(u_i - x_i^\top \delta/\sqrt{n})) - \rho_q(w_i(u_i - x_i^\top \delta/\sqrt{n}))\} + Z_n(\delta).$$

Using similar arguments as in Lee et al. (2012) for consistency of modified quantile regression, we show that the first term becomes asymptotically negligible in determining the minimizer of $Z_n^M(\delta)$. First, we consider the expectation of the first term of the above decomposition.

$$\begin{aligned}
& E\Big(\sum_{i=1}^n \{\rho_q^M(w_i(u_i - x_i^\top \delta/\sqrt{n})) - \rho_q(w_i(u_i - x_i^\top \delta/\sqrt{n}))\}\Big) + \frac{nq(1-q)}{2\lambda_\gamma} \\
= & \sum_{i=1}^n \int_{x_i^\top \delta/\sqrt{n}}^{\frac{1-q}{\lambda_\gamma w_i} + x_i^\top \delta/\sqrt{n}} \Big(\frac{\lambda_\gamma}{2}\frac{q w_i^2}{1-q}\Big(u - \frac{x_i^\top \delta}{\sqrt{n}}\Big)^2 - q w_i\Big(u - \frac{x_i^\top \delta}{\sqrt{n}}\Big) + \frac{q(1-q)}{2\lambda_\gamma}\Big) f_i(\xi_i + u) du \\
+ & \sum_{i=1}^n \int_{-\frac{q}{\lambda_\gamma w_i} + x_i^\top \delta/\sqrt{n}}^{x_i^\top \delta/\sqrt{n}} \Big(\frac{\lambda_\gamma}{2}\frac{(1-q) w_i^2}{q}\Big(u - \frac{x_i^\top \delta}{\sqrt{n}}\Big)^2 - (q-1) w_i\Big(u - \frac{x_i^\top \delta}{\sqrt{n}}\Big) + \frac{q(1-q)}{2\lambda_\gamma}\Big) f_i(\xi_i + u) du \\
= & \sum_{i=1}^n \int_{x_i^\top \delta/\sqrt{n}}^{\frac{1-q}{\lambda_\gamma w_i} + x_i^\top \delta/\sqrt{n}} \frac{\lambda_\gamma}{2}\frac{q w_i^2}{1-q}\Big(u - \frac{x_i^\top \delta}{\sqrt{n}} - \frac{1-q}{\lambda_\gamma w_i}\Big)^2 f_i(\xi_i + u) du \\
+ & \sum_{i=1}^n \int_{-\frac{q}{\lambda_\gamma w_i} + x_i^\top \delta/\sqrt{n}}^{x_i^\top \delta/\sqrt{n}} \frac{\lambda_\gamma}{2}\frac{(1-q) w_i^2}{q}\Big(u - \frac{x_i^\top \delta}{\sqrt{n}} + \frac{q}{\lambda_\gamma w_i}\Big)^2 f_i(\xi_i + u) du.
\end{aligned}$$

Making use of a Taylor series expansion of f_i at ξ_i from (C-2), we can show that for $\lambda_\gamma = cn^\alpha$, the above expression is given by

$$\frac{q(1-q)}{6c^2 n^{2\alpha}} \sum_{i=1}^n \frac{f_i(\xi_i)}{w_i} + \frac{q(1-q)}{6c^2 n^{2\alpha}} \sum_{i=1}^n \frac{f_i'(\xi_i) x_i^\top \delta}{w_i \sqrt{n}} + o(n^{-2\alpha+1/2}).$$

Note that $\sum_{i=1}^n f_i(\xi_i)/w_i = O(n)$ for $w_i = f_i(\xi_i)$, and $\sum_{i=1}^n \{f_i'(\xi_i)/w_i\}(x_i^\top \delta/\sqrt{n}) = O(\sqrt{n})$ as $f_i'(\xi_i)/w_i$, $i = 1, \ldots, n$ are uniformly bounded from the condition (C-2), and (C-1), and $|x_i^\top \delta| \le \|x_i\|_2 \|\delta\|_2 \le (\|x_i\|_2^2 + \|\delta\|_2^2)/2$ while $\sum_{i=1}^n \|x_i\|_2^2 = O(n)$ from the condition (C-3). Thus, we have

$$E\sum_{i=1}^n \{(\rho_q^M(w_i(u_i - x_i^\top \delta/\sqrt{n})) - (\rho_q(w_i(u_i - x_i^\top \delta/\sqrt{n}))\} - C_n \to 0 \quad \text{if } \alpha > 1/4,$$

where $C_n \equiv -q(1-q)/(2cn^{\alpha-1}) + q(1-q)/(6c^2 n^{2\alpha})\sum_{i=1}^n f_i(\xi_i)/w_i$. And similarly,

$$\begin{aligned} & Var\Big(\sum_{i=1}^n \{\rho_q^M(w_i(u_i - x_i^\top \delta/\sqrt{n})) - \rho_q(w_i(u_i - x_i^\top \delta/\sqrt{n}))\}\Big) \\ = \; & \sum_{i=1}^n \frac{q^2(1-q)^2 f_i(\xi_i)}{20c^3 n^{3\alpha} w_i} + o(n^{-3\alpha+1}) \to 0 \quad for\ \alpha > 1/3 \end{aligned}$$

Thus, under the condition that α >1/3,

$$Z_n^M(\delta) - Z_n(\delta) - C_n = \sum_{i=1}^n \{\rho_q^M(w_i(u_i - x_i^\top \delta/\sqrt{n})) - \rho_q(w_i(u_i - x_i^\top \delta/\sqrt{n}))\} - C_n \xrightarrow{p} 0.$$

Finally, Theorem 5.1 of Koenker (2005) shows that $Z_n(\delta) \xrightarrow{d} -\delta \check{W} + \frac{1}{2}\delta^\top D_2 \delta$ where $\check{W} \sim N(0, q(1-q)D_2)$, which completes the proof. □

Proof of Theorem 2.
The proof of convergence is similar to the proof of Theorem 1, except that u_i is now expressed as $(y_i - x_i^\top \beta)/(x_i^\top \tau)$. The behavior of $\sqrt{n}(\check{\beta}_\tau^M - \beta)$ follows from consideration of $\sum_{i=1}^n \{\rho_q^M(u_i - \frac{x_i^\top \delta}{x_i^\top \tau}\frac{1}{\sqrt{n}}) - \rho_q(u_i)\}$. First, we decompose the above expression.

$$\begin{aligned} Z_n^{M^*}(\delta) &= \sum_{i=1}^n \{\rho_q^M(u_i - \frac{x_i^\top \delta}{x_i^\top \tau}\frac{1}{\sqrt{n}}) - \rho_q(u_i)\} \\ &= \sum_{i=1}^n \{\rho_q^M(u_i - \frac{x_i^\top \delta}{x_i^\top \tau}\frac{1}{\sqrt{n}}) - \rho_q(u_i - \frac{x_i^\top \delta}{x_i^\top \tau}\frac{1}{\sqrt{n}})\} + Z_n^*(\delta), \end{aligned}$$

where $Z_n^*(\delta) = \sum_{i=1}^n \{\rho_q(u_i - \frac{x_i^\top \delta}{x_i^\top \tau}\frac{1}{\sqrt{n}}) - \rho_q(u_i)\}$. Similar to the proof in Theorem 1, for $\alpha > 1/3$, it can be observed that $Z_n^{M*}(\delta) - Z_n^*(\delta) - C_n^* \xrightarrow{p} 0$, where

$$\begin{aligned} C_n^* &\equiv -q(1-q)/(2cn^{\alpha-1}) + q(1-q)/(6c^2 n^{2\alpha})\sum_{i=1}^n f_i(\xi_i)(x_i^\top \tau) \\ &= -q(1-q)/(2cn^{\alpha-1}) + q(1-q)/(6c^2 n^{2\alpha})(n \cdot f(F^{-1}(q))). \end{aligned}$$

Thus, asymptotic behavior of $\check{\beta}_\tau^M$ is equivalent to that of $\check{\beta}_\tau$. Koenker and Zhao (1994) show that $\sqrt{n}(\check{\beta}_\tau - \beta) \xrightarrow{d} N(0, \frac{q(1-q)}{f^2(F^{-1}(q))} D_2^{*-1})$. Now, $\check{\beta}_{\hat{\tau}}$ with a $\sqrt{n}$-consistent estimator of τ up to scale will have the same asymptotic behavior as $\check{\beta}_\tau$ as shown in Theorem 2.1 of Koenker and Zhao (1994). □

References

Bassett, G. and Koenker, R. (1978). Asymptotic theory of least absolute error regression, *Journal of the American Statistical Association* **73**(363): 618–622.

Gilley, O. and Pace, R. K. (1996). On the harrison and rubinfeld data, *Journal of Environmental Economics and Management* **31**: 403–405.

Gutenbrunner, C. and Jurečková, J. (1992). Regression rank scores and regression quantiles, *The Annals of Statistics* **20**(1): 305–330.

He, X. (1997). Quantile curves without crossing, *The American Statistician* **51**(2): 186–192.

Horowitz, J. (1998). Bootstrap methods for median regression models, *Econometrica* **66**: 1327–1352.

Jung, Y. (2010). *Regularization of Case Specific Parameters: A New Approach for Improving Robustness and/or Efficiency of Statistical Methods*, PhD thesis, The Ohio State University.

Jung, Y., MacEachern, S. N. and Lee, Y. (2010). Window width selection for l_2 adjusted quantile regression, *Technical Report 835*, Department of Statistics, The Ohio State University.

Koenker, R. (2005). *Quantile Regression*, Cambridge University Press.

Koenker, R. and Bassett, G. (1978). Regression quantiles, *Econometrica* **46**(1): 33–50.

Koenker, R. and Machado, J. A. F. (1999). Goodness of fit and related inference processes for quantile regression, *Journal of the American Statistical Association* **94**(488): 1296–1310.

Koenker, R. and Park, B. J. (1996). An interior point algorithm for nonlinear quantile regression, *Journal of Econometrics* **71**: 265–283.

Koenker, R. W. and D'Orey, V. (1987). Algorithm as 229: Computing regression quantiles, *Journal of the Royal Statistical Society. Series C (Applied Statistics)* **36**(3): 383–393.

Koenker, R. and Zhao, Q. (1994). L-estimation for linear heteroscedastic models, *Journal of Nonparametric Statistics* **3**(3): 223–235.

Lee, Y., MacEachern, S. N. and Jung, Y. (2012). Regularization of case-specific parameters for robustness and efficiency, *Statistical Science* **27**(3): 350–372.

Nychka, D., Gray, G., Haaland, P., Martin, D. and O'Connell, M. (1995). A nonparametric regression approach to syringe grading for quality improvement, *Journal of the American Statistical Association* **90**(432): 1171–1178.

Rubinfeld, D. H. D. L. (1978). Hedonic housing prices and the demand for clean air, *Journal of Environmental Economics and Management* **5**: 81–102.