Journal of Multivariate Analysis **72**, 78–99 (2000)
Article ID jmva.1999.1849, available online at http://www.idealibrary.com on IDEAL®

Restricted Regression Quantiles

Quanshui Zhao[1]

City University of Hong Kong, Kowloon, Hong Kong, China

Received September 10, 1998

Regression quantiles can be used as prediction intervals for the response variable. But such applications are often hampered by the problem of quantile crossing in finite sample cases. This article examines the efficiency properties of restricted regression quantiles that are proposed by X. He (1997, *Amer. Statist.* **51**, 186–192) to overcome the crossing problem of the usual regression quantiles of R. Koenker

AMS 1991 subject classifications: 62J05.
Key words and phrases: linear model, heteroscedasticity, regression quantile, efficiency.

1. INTRODUCTION

Regression quantile (RQ) planes for linear models, analogous to the sample quantiles for the location model, provide distribution-free prediction intervals for the response variable (cf. Koenker and Bassett (1978) and Zhou and Portnoy (1996)). As an extension of the sample quantiles for the location model, RQ planes, however, do not inherit all the characteristics of the sample quantiles in its first place. Quantile plane crossing is a common problem in finite sample case. For example, the 80% quantile plane may lie above the 90% quantile plane for a given design point. The crossing of the two quantile planes causes a prediction interval with a high confidence level to be narrower than that with a low confidence level and contradicts the common probability theory. To avoid the quantile plane crossing problem, Koenker (1984) considers restricted regression quantiles (RRQ) by computing multiple parallel regression quantiles simultaneously for linear models. He (1997) uses a multi-step strategy and defines restricted

[1] Address correspondence to Department of Management Sciences, City University of Hong Kong, Kowloon, Hong Kong. The author is indebted to Xuming He, Roger Koenker, Stephen Portnoy, Jan De Leeuw, and two referees for many helpful comments. This research was partially supported by a research grant from the City University of Hong Kong.

0047-259X/00 $35.00

regression quantile for a broad class of models including linear heteroscedastic models and nonlinear regression quantile models. In this article, we adopt the definitions of He (1997) for both linear models and linear heteroscedastic models. We will study the asymptotic properties of the RRQs with comparison to the RQs.

For linear models with i.i.d. innovations, restricting RQs to a subclass, for instance the parallel planes as suggested by Koenker (1984), can avoid the quantile plane crossing problem. For linear models, we will show that the RRQs can be more efficient than RQs for certain type of unimode error distributions. Consider the linear model

$$y_i = x_i'\beta + \varepsilon_i, \qquad i = 1, 2, ..., n, \tag{1.1}$$

where the ε_i's are i.i.d. random errors with a c.d.f. F and the x_i's are either nonrandom or random but independent of ε_i. Without loss of generality, we assume that the model includes an intercept term. Therefore the vector x_i has its first component to be a constant one, i.e., $x_i' = (1, x_{1i}')$ and the parameter vector is $\beta' = (\beta_0, \beta_1')$. The theoretical τth $(0 < \tau < 1)$ regression quantile is therefore $\beta(\tau) = (\beta_0 + F^{-1}(\tau), \beta_1')'$, of which only the first component depends on τ and the slope part is independent of τ, implying that the slope can be estimated separately. Koenker and Bassett (1978) show that the asymptotic covariance of the RQ estimator is proportional to $\tau(1-\tau)/f^2(F^{-1}(\tau))$, which is large when τ is either close to zero or one for certain error distributions. The slope parameter is often estimated poorly at the lower or upper tails, where the regression quantile planes are often used to construct prediction intervals.

The inefficient slope estimates at the tail quantiles seem to be the main cause of the quantile plane crossing problem for linear models. Restricting conditional quantiles to a subclass as suggested by Koenker (1984) and He (1997) is a straightforward solution. In this article, we adopt the two-step approach of He (1997). First, we estimate the slope parameter by the least absolute deviation (LAD) estimator, $\hat{\beta}_1$. This common slope estimate guarantees that all the quantile planes will be parallel with no crossing. Second, we estimate the intercept at different quantiles by the sample quantile $\hat{\beta}_0(\tau)$ of the residuals obtained from the first step

$$\hat{r}_i = y_i - x_{1i}'\hat{\beta}_1 \simeq \beta_0 + \varepsilon_i. \tag{1.2}$$

The combined estimate $\hat{\beta}(\tau) = (\hat{\beta}_0(\tau), \hat{\beta}_1')'$ serves as a consistent estimator of the theoretical regression quantile $\beta(\tau)$, and is so-called the restricted regression quantile (RRQ) in this article. The definition of the RRQ assures that it may be more efficient than the RQ for certain type of error distribution such as Gaussian and Student's t.

The theoretical quantile planes for a linear model with heteroscedastic errors are not necessarily parallel, but they should retain the noncrossing property at observed design points. For the linear heteroscedastic model

$$y = x'\beta + (x'\gamma)\,\varepsilon \tag{1.3}$$

studied by Gutenbrunner and Jurečková (1992) and Koenker and Zhao (1994), the RQs are defined as

$$\tilde{\beta}(\tau) = \arg\min_{b} \sum_{1}^{n} \rho_\tau(y_i - x_i' b), \tag{1.4}$$

for observations $\{(x_i, y_i), i = 1, 2, ..., n\}$, where $\rho_\tau(u) = \tau u + \max\{-u, 0\}$, $0 < \tau < 1$. The estimated RQ planes $y = x'\hat{\beta}(\tau)$ are not parallel in nature, but often cross at some design points x, with $x'\gamma > 0$ where the quantile planes should not cross in theory (cf. He (1997)).

To eliminate the troublesome quantile plane crossing problem, He (1997) proposes a three-step RRQ definition. Suppose $x'\gamma > 0$ for all x values in the interior of its domain, ε has median zero and $|\varepsilon|$ is normalized to have a median of one. The three step definition is explained as follows:

(i) Regress $\{y_i\}$ on $\{x_i\}$ to obtain coefficient estimate $\hat{\beta}$ and residuals $\hat{r}_i = y_i - x_i'\hat{\beta}$;

(ii) Regress the absolute residuals $\{|\hat{r}_i|\}$ on $\{x_i\}$'s to obtain the LAD regression coefficient $\hat{\gamma}$ with the fitted values being $s_i = x_i'\hat{\gamma}$;

(iii) Find quantile factor $\hat{c}_\tau$ by minimizing $\sum_i \rho_\tau(\hat{r}_i - c\, s_i)$ over c, and obtain the τth quantile plane $y = x'\hat{\beta} + \hat{c}_\tau x'\hat{\gamma} = x'(\hat{\beta} + \hat{c}_\tau\hat{\gamma})$. The vector $\hat{\beta}(\tau) = \hat{\beta} + \hat{c}_\tau\hat{\gamma}$ is the so-called RRQ estimator.

He (1997) shows that $\hat{c}_\tau$ is monotonously nondecreasing in τ. Therefore, the quantile planes do not cross at any design point x_i with $x_i'\hat{\gamma} > 0$, since the monotone property of $\hat{c}_\tau$ assures that $x_i'(\hat{\beta} + \hat{c}_{\tau_1}\hat{\gamma}) \leqslant x_i'(\hat{\beta} + \hat{c}_{\tau_2}\hat{\gamma})$ for any τ_1 and τ_2 $(0 < \tau_1 \leqslant \tau_2 < 1)$. Similar to the linear model case, the RRQ estimator may be more efficient than the unrestricted RQs for the linear heteroscedastic model with unimode symmetric error distributions, while it may be less efficient for certain other types of error distribution. The problem we are interested in is the magnitude of efficiency gain or loss in avoiding quantile plane crossing.

In this article, we investigate the efficiency properties of the RRQ estimators in comparison with the unrestricted RQ estimators. In Section 2, we present a Bahadur-type representation of the RRQs for both linear and linear heteroscedastic models under mild regularity conditions. Asymptotic

normality and asymptotic covariance matrices of the RRQs are obtained. In Section 3, a relative efficiency measure based on the mean square prediction error (MSPE) of quantile planes is defined and used to compare the performance of the RRQ and RQ estimators. For linear models, the relative efficiency measure leads to an explicit expression that shows clearly the outstanding performance of the RRQs. For linear heteroscedastic models, since the relative efficiency depends on design matrix, performance is assessed by Monte Carlo simulation. An example is presented to illustrate how RRQs are used in constructing calibration confidence intervals in biometrical experiment. A short conclusion is given in Section 4 and the proofs of the theorems are given in the Appendix.

2. ASYMPTOTIC DISTRIBUTIONS

In this section, we investigate the asymptotic distributions of the RRQ estimators for linear and linear heteroscedastic models. The asymptotic results will be used in Section 3 to evaluate the performance of RRQ estimators in comparison with the RQ estimators.

2.1. *Linear Model*

Suppose $\{(x_i, y_i), i = 1, 2, ..., n\}$ are observations from the linear model

$$y_i = x_i'\beta + \varepsilon_i, \tag{2.1}$$

where the ε_i are i.i.d. random variables with a density function f and a c.d.f. F, and x_i is a design vector in k-dimensional space. Without loss of generality, we assume that the model contains an intercept term.

Following the discussion in the Introduction, we assume that $\hat{\beta}_1$ is a $\sqrt{n}$-consistent estimator of the slope vector β_1 used in the first stage of the RRQ definition, i.e., $\sqrt{n}(\hat{\beta}_1 - \beta_1) = O_p(1)$. Define

$$\hat{\beta}_0(\tau) = \arg\min_c \sum_1^n \rho_\tau(\hat{r}_i - c),$$

as the sample quantile of the residuals $\{\hat{r}_i\}$, where $\hat{r}_i = y_i - x_{1i}'\hat{\beta}_1$ and x_{1i} is a $k-1$ dimensional vector. Then the RRQ estimator is in the form $\hat{\beta}(\tau) = (\hat{\beta}_0(\tau), \hat{\beta}_1')'$, which is an estimator of the theoretical regression quantile $\beta(\tau) = (\beta_0 + F^{-1}(\tau), \beta_1')'$. Compared with the unrestricted RQ estimator of Koenker and Bassett (1978), defined as

$$\tilde{\beta}(\tau) = \arg\min_b \sum_1^n \rho_\tau(y_i - x_i'b),$$

the RRQ estimator has the merit of obtaining parallel, noncrossing quantile planes. The following theorem gives the asymptotics of the RRQ estimator.

THEOREM 2.1. *Suppose that* $f(x)$ *is continuous and finite at* $F^{-1}(\tau)$, $0<\tau<1$, *and*

(i) $\sqrt{n}\,(\hat{\beta}_1-\beta_1)=O_p(1)$;

(ii) $n^{-1}\sum_1^n x_{1i}-\mu_1=o_p(1)$ *and* $n^{-1/2}\max_i \|x_i\|=o_p(1)$;

(iii) $E\|x_i\|^2<\infty$.

Then

$$\sqrt{n}\,(\hat{\beta}_0(\tau)-\beta_0-F^{-1}(\tau))=\frac{1}{f(F^{-1}(\tau))}\,n^{-1/2}\sum_1^n \psi_\tau(\varepsilon_i-F^{-1}(\tau)) -\mu_1'\sqrt{n}\,(\hat{\beta}_1-\beta_1)+o_p(1),$$

where $\psi_\tau(u)=\tau-I_{\{u<0\}}$.

Remark. The design vectors x_i's can be random or nonrandom. When they are random, we basically assume they are i.i.d. distributed and satisfy conditions (ii) and (iii). When the design matrix is nonrandom, condition (ii) should be $n^{-1}\sum_1^n x_{1i}-\mu_1=o(1)$ and $n^{-1/2}\max_i\|x_i\|=o(1)$, and condition (iii) should be $n^{-1}\sum_1^n\|x_i\|^2<\infty$.

The theorem shows that the intercept estimator at the second step, $\hat{\beta}_0(\tau)$, depends on the initial estimator $\hat{\beta}_1$ through the linear relationship shown in the theorem. In fact, if $\mu_1=0$, the dependence would disappear and the distribution of $\hat{\beta}_0(\tau)$ would be easy to obtain. In fact, if one centers the predictor x_{1i}'s before estimating β_1, the above-mentioned condition can be achieved. Then we have the following result.

COROLLARY 2.1. *Suppose* $\mu_1=0$. *Then*

$$\sqrt{n}\,(\hat{\beta}_0(\tau)-\beta_0-F^{-1}(\tau))\longrightarrow_{\mathscr{L}} N\left(0,\frac{\tau(1-\tau)}{f^2(F^{-1}(\tau))}\right).$$

The intercept estimator is as efficient as that of the unrestricted RQ estimator (cf. Koenker and Bassett, 1978) and is asymptotically normal distributed when x_{1i}'s are centered. The asymptotic distribution does not depend on the first step slope estimator, which can be chosen as efficient as possible (e.g., an L-estimator). The RRQ estimator can in general be designed to be more efficient than its unrestricted counterpart. This point will be strengthened in Section 3 by using an overall efficiency measure.

In this subsection, we have used a model with the strong assumption that the errors are i.i.d. distributed, what happens when the model admits different slope parameters for distinct quantiles?

2.2. *Linear Heteroscedastic Model*

For the linear heteroscedastic model (1.3), the RRQ estimator is defined by the three-step method proposed by He (1997) as described in the introduction section. First, the median regression quantile estimator $\hat{\beta}(0.5)$, as shown in Koenker and Zhao (1994), consistently estimates β and has asymptotic normal distribution under mild regularity conditions. This estimator is a natural choice for estimating β. If the error distribution type is known, L-estimators, such as the average of several regression quantiles, can be used to estimate β more efficiently.

In general, suppose we have a $\sqrt{n}$-consistent initial estimator $\hat{\beta}$ such that $\sqrt{n}\,(\hat{\beta}-\beta)=O_p(1)$. Denote $\hat{r}_i=y_i-x_i'\hat{\beta}$ $(i=1, 2, ..., n)$ and define

$$\hat{\gamma}=\arg\min_b \sum_1^n |\,|\hat{r}_i|-x_i'b|,$$

following He (1997). The heteroscedastic scale components can be estimated by $\hat{\sigma}_i=x_i'\hat{\gamma}$, $(i=1, 2, ..., n)$. Finally, the quantile factor can be estimated by $\hat{c}_\tau$, as solved by the minimization problem,

$$\hat{c}_\tau=\arg\min_c \sum_1^n \rho_\tau(r_i-c\hat{\sigma}_i).$$

He (1997) shows that $\hat{c}_\tau$ is a monotonously nondecreasing function of τ, which assures that the quantile plane $y=x'(\hat{\beta}+\hat{c}_\tau\hat{\gamma})$ is monotonous in τ at any design point x with $x'\hat{\gamma}>0$. Quantile plane crossings at such design points can therefore be avoided.

For the RRQ estimator defined as, $\hat{\beta}(\tau)=\hat{\beta}+\hat{c}_\tau\hat{\gamma}$, we have the following result.

THEOREM 2.2. *Suppose*

(i) *$f(x)$ is continuous at $F^{-1}(\tau)$ for $0<\tau<1$ and is normalized such that $med(\varepsilon)=0$ and $med(|\varepsilon|)=1$.*

(ii) $n^{-1}\sum_1^n x_i-\mu=o_p(1)$ *and* $n^{-1/2}\max_i \|x_i\|=o_p(1)$.

(iii) $E\|x_i\|^2<\infty$;

(iv) $n^{-1}\sum_1^n \sigma_i^{-1}x_ix_i'-Q_1=o_p(1)$, *where $\sigma_i=x_i'\gamma>0$ for all i and Q_1 is positive definite.*

Then

$$\begin{aligned}\sqrt{n}\,(\hat{\gamma}-\gamma) &= \frac{Q_1^{-1}}{f(1)+f(-1)}\,n^{-1/2}\sum_1^n x_i\psi_{1/2}(|\varepsilon_i|-1)\\ &\quad -\frac{f(1)-f(-1)}{f(1)+f(-1)}\sqrt{n}\,(\hat{\beta}-\beta)+o_p(1),\\ \sqrt{n}\,(\hat{c}_\tau-F^{-1}(\tau)) &= \frac{\gamma'}{f(F^{-1}(\tau))(\mu'\gamma)}\,n^{-1/2}\sum_1^n x_i\psi_\tau(\varepsilon_i-F^{-1}(\tau))\\ &\quad -\frac{\mu'}{\mu'\gamma}\,(\sqrt{n}\,(\hat{\beta}-\beta)+F^{-1}(\tau)\sqrt{n}\,(\hat{\gamma}-\gamma))+o_p(1)\end{aligned} \tag{2.2}$$

and

$$\begin{aligned}&\sqrt{n}\,(\hat{\beta}(\tau)-\beta(\tau))\\ &\quad=\frac{\gamma\gamma'}{f(F^{-1}(\tau))(\mu'\gamma)}\,n^{-1/2}\sum_1^n x_i\psi_\tau(\varepsilon_i-F^{-1}(\tau))\\ &\qquad+\left(I-\frac{\gamma\mu'}{\mu'\gamma}\right)\frac{F^{-1}(\tau)\,Q_1^{-1}}{f(1)+f(-1)}\,n^{-1/2}\sum_1^n x_i\psi_{1/2}(|\varepsilon_i|-1)\\ &\qquad+\left(I-\frac{\gamma\mu'}{\mu'\gamma}\right)\left(1-\frac{f(1)-f(-1)}{f(1)+f(-1)}F^{-1}(\tau)\right)\sqrt{n}\,(\hat{\beta}-\beta)+o_p(1).\end{aligned} \tag{2.3}$$

COROLLARY 2.2. *Under the conditions of Theorem* 2.2,

$$\sqrt{n}\,(\hat{\beta}(\tau)-\beta(\tau))\longrightarrow_{\mathscr{L}} N(0,\Sigma),$$

where

$$\begin{aligned}\Sigma &= \tau(1-\tau)\,AQA'+\tfrac{1}{4}(b^2+c^2)\,BQB'\\ &\quad+[\,b(F(1)\wedge\tau-F(-1)\wedge\tau-\tau/2)+c(\tau\wedge 0.5-\tau/2)\,](AQB'+BQA')\\ A &= \frac{\gamma\gamma'}{f(F^{-1}(\tau))(\gamma'\mu)}, \qquad B=\left(I-\frac{\gamma\mu'}{\gamma'\mu}\right)Q_1^{-1}\\ b &= \frac{F^{-1}(\tau)}{f(1)+f(-1)}, \qquad c=\frac{1}{f(0)}\left(1-\frac{f(1)-f(-1)}{f(1)+f(-1)}F^{-1}(\tau)\right)\end{aligned}$$

and $Q=\lim n^{-1}\sum_1^n x_i x_i'$ *in probability.*

The asymptotic covariance of the RRQ estimator is in a complicated form, and may be simplified when the density function $f(x)$ is symmetric about zero. As both the RQ and RRQ estimators are consistent and

asymptotically normal distributed, it is plausible to find out which estimator is more efficient. A detailed discussion based on an efficiency measure is given in the next section.

3. ASYMPTOTIC EFFICIENCY

Asymptotic variance is commonly used to assess one-dimensional estimators. For vector estimators such as regression quantiles, it may be plausible to compare the asymptotic covariance matrices if there are any ways to do so. In this section, a universal measure, the mean square prediction error (MSPE), is proposed. The efficiency of the RRQ estimator relative to the RQ estimator can be evaluated based on this measure.

The MSPE is a natural performance measure for vector estimators that emphasizes the predictability of the model, rather than the estimation accuracy of the estimator itself. For both linear and linear heteroscedastic models, RRQ and RQ estimators are mainly used to estimate quantile planes, $y(\tau) = x'\beta(\tau)$, with which prediction intervals or calibration intervals can be constructed. The overall estimation errors of the quantile planes over all the observed design points seems to be a natural performance measure for the estimation method.

We define the mean square prediction error of a quantile plane as

$$MSPE = n^{-1} \sum_{1}^{n} (\hat{y}_i(\tau) - y_i(\tau))^2,$$

where $\hat{y}_i(\tau) = x_i'\hat{\beta}(\tau)$ and $y_i(\tau) = x_i'\beta(\tau)$. The $MSPE$ is the average square prediction errors over all the observed design points of the regression quantile plane, a measure also used by He (1997). Since $\beta(\tau)$ is not observable, $MSPE$ can not be assessed unless Monte Carlo simulation approach is used. In fact, the MSPE can be approximated by the large sample properties developed in Section 2 as

$$\begin{aligned} MSPE &= (\hat{\beta}(\tau) - \beta(\tau))' \, n^{-1} \sum_{1}^{n} x_i x_i' (\hat{\beta}(\tau) - \beta(\tau)) \\ &\simeq (\hat{\beta}(\tau) - \beta(\tau))' \, Q(\hat{\beta}(\tau) - \beta(\tau)), \end{aligned}$$

where $Q = \lim_n n^{-1} \sum_1^n x_i x_i'$ in probability. Since $\hat{\beta}(\tau) - \beta(\tau)$ has an asymptotic normal distribution, the MSPE defined above is a quadratic form of a normal vector. Suppose $\hat{\beta}(\tau) - \beta(\tau) \sim N(0, n^{-1}\Sigma)$, the the expected value of MSPE is easily obtained to be

$$E(MSPE) \sim n^{-1} \operatorname{trace}(Q\Sigma),$$

which converges to zero as n goes to infinity. But trace(ΣQ) is a constant, which depends only on the asymptotic covariance of the estimator and Q, and measures the accuracy of the quantile plane estimation.

In this section, we are specially interested in the gain or loss in efficiency of the RRQ estimators relative to the RQ estimators. This relative efficiency measure is defined as

$$RE(RRQ, RQ) = \frac{\text{trace}(Q\Sigma_{RRQ})}{\text{trace}(Q\Sigma_{RQ})}, \tag{3.1}$$

where Σ_{RRQ} and Σ_{RQ} are the asymptotic covariances of the RRQ and the RQ estimators, respectively. For a particular model, if $RE(RRQ, RQ)$ is less than one, the RRQ is more efficient; otherwise, the RQ is more efficient.

3.1. *Linear Models*

The relative efficiency measure has an explicit formula for linear models, and makes evaluating the performance of the RRQs straightforward. In this subsection we compare the performance of the RRQ and RQ estimators for several error distributions.

For model (2.1), Theorem 2.1 shows that the RRQ estimator has an asymptotic normal distribution. Further, if $\mu_1 = 0$, which is realized by centering the regressors, the intercept and the slope estimators are uncorrelated. The asymptotic covariance of the RRQ estimator is

$$\Sigma_{RRQ} = \begin{pmatrix} \tau(1-\tau)\, f^{-2}(F^{-1}(\tau)) & 0 \\ 0 & (2f(0))^{-2}\, Q_{11}^{-1} \end{pmatrix},$$

where $Q_{11} = \lim_n n^{-1} \sum_1^n x_{1i} x_{1i}'$ is the diagonal element of

$$Q = \begin{pmatrix} 1 & 0 \\ 0 & Q_{11} \end{pmatrix}.$$

Thus, for k-dimensional x,

$$\text{trace}(Q\Sigma_{RRQ}) = \frac{\tau(1-\tau)}{f^2(F^{-1}(\tau))} + \frac{k-1}{4f^2(0)}.$$

The asymptotic distribution of the unrestricted RQs defined by Koenker and Bassett (1978) is known to satisfy

$$\sqrt{n}\,(\tilde{\beta}(\tau) - \beta(\tau)) \to_{\mathscr{L}} N(0, \Sigma_{RQ}),$$

under similar regularity conditions, where

$$\Sigma_{RQ} = \frac{\tau(1-\tau)}{f^2(F^{-1}(\tau))} Q^{-1}.$$

It follows that

$$\text{trace}(Q\Sigma_{RQ}) = \frac{k\tau(1-\tau)}{f^2(F^{-1}(\tau))}.$$

The relative efficiency of the RRQ estimator with respect to the RQ estimator follows as

$$RE(RRQ, RQ) = \frac{\text{trace}(Q\Sigma_{RRQ})}{\text{trace}(Q\Sigma_{RQ})} = \frac{1}{k} + \frac{(k-1) f^2(F^{-1}(\tau))}{4k\tau(1-\tau) f^2(0)}. \tag{3.2}$$

This explicit expression shows that the RQ and RRQ estimators are equally efficient at the median, $F^{-1}(0.5) = 0$. However, there is no clear answer for other quantiles, where the relative efficiency depends on error distribution.

The relative efficiency of the RRQ estimators with respect to the RQ estimators varies for different error distributions and different quantile levels. Table I tabulates the relative efficiency for three unimode symmetric distributions, normal, double exponential and Cauchy, and two asymmetric distributions, exponential and lognormal, where the asymmetric distributions are rescaled to meet the assumptions in Theorem 2.1. In the case of symmetric distributions, since the median regression estimator is more efficient than the quantile regression estimator for the slope, the RRQ is more efficient than the RQ, as shown in Table I.

When asymmetric error distributions are concerned, the RRQ estimator cann't dominate for all quantile levels. Since the first step estimator, the LAD, may not be the most efficient one in the RQ family. The combined RRQ estimator should in general inherit this property. The results in Table I for the exponential and lognormal error distributions show exactly what is expected. It can be seen that the RRQ estimator is more efficient than the RQ estimator for quantile levels at one tail, but less efficient at the other, depending on error distribution. To achieve a high efficiency for the RRQ estimator, L-estimators, which may incorporate more information of the error distribution, can be used in the first step.

3.2. *Linear Heteroscedastic Models*

The relative efficiency measure of the RRQ with respect to the RQ for linear heteroscedastic models is in the form of a complicated formula. Evaluation of the efficiency of the RRQs is not straightforward. We will use

TABLE I

The Relative Efficiency RE(RRQ, RQ) for Linear Models ($k=1$)

τ	Normal	D-Exp[a]	Cauchy	Exponential	Lognormal
0.05	0.6759	0.5263	0.5016	10.0000	5.2197
0.10	0.7688	0.5556	0.5127	5.0000	3.9877
0.20	0.8847	0.6250	0.5933	2.5000	2.5711
0.30	0.9521	0.7143	0.7550	1.6667	1.7905
0.40	0.9885	0.8333	0.9261	1.2500	1.3107
0.50	1.0000	1.0000	1.0000	1.0000	1.0000
0.60	0.9885	0.8333	0.9261	0.8333	0.7943
0.70	0.9521	0.7143	0.7550	0.7143	0.6584
0.80	0.8847	0.6250	0.5933	0.6250	0.5715
0.90	0.7688	0.5556	0.5127	0.5555	0.5207
0.95	0.6759	0.5263	0.5016	0.5263	0.5066

[a] D-Exp means double exponential distribution.

both a special case and Monte Carlo simulation in this subsection to assess the relative performance.

For linear heteroscedastic models (1.3), Koenker and Zhao (1994) show that the RQ estimator $\tilde{\beta}(\tau)$ has asymptotic normal distribution

$$\sqrt{n}\,(\tilde{\beta}(\tau)-\beta(\tau)) \to_{\mathscr{L}} N(0, \Sigma_{RQ}),$$

where

$$\Sigma_{RQ}=\frac{\tau(1-\tau)}{f^2(F^{-1}(\tau))}\,Q_1^{-1}QQ_1^{-1}.$$

It follows that

$$\operatorname{trace}(Q\Sigma_{RQ})=\frac{\tau(1-\tau)}{f^2(F^{-1}(\tau))}\operatorname{trace}(Q_1^{-1}QQ_1^{-1}Q). \tag{3.3}$$

The MSPE of the RQ estimator can be obtained. Corollary 2.2 shows that the RRQ estimator $\hat{\beta}(\tau)$ also has an asymptotic normal distribution. As a matter of fact, the asymptotic covariance depends on the design matrix X, tabulating the relative efficiency for all different designs is impossible.

We first consider a case when heteroscedasticity is not in existence. Since the MSPE is a continuous function of the heteroscedasticity parameter γ_1, we hope that the case $\gamma_1=0$ is representative for cases where γ_1 is close to zero. Without loss of generality, we assume that the design vectors are

centered before regression is conducted such that $\mu = (1, 0)'$. Under these assumptions, it is known that $Q_1 = \gamma_0^{-1} Q$, and

$$\text{trace}(Q\Sigma_{RRQ}) = \gamma_0^2 \left[\frac{\tau(1-\tau)}{f^2(F^{-1}(\tau))} + \frac{(k-1)(F^{-1}(\tau))^2}{4(f(1)+f(-1))^2} + \frac{k-1}{4f^2(0)} \left(1 - \frac{f(1)-f(-1)}{f(1)+f(-1)} F^{-1}(\tau) \right)^2 \right]. \quad (3.4)$$

In this case, we have an explicit formula of the relative efficiency, which is independent of the design matrix. Table II lists the relative efficiencies for the same error distributions considered in Table I at various quantile levels. we expect that the RRQ estimator would still be more efficient than the RQ estimator for unimode symmetric distributions such as the linear models. While the efficiency gain should be decreased because the three step method introduces more variations to the estimates.

Table II reports the relative efficiency of the RRQ estimator with respect to the RQ estimator for the special case of the linear heteroscedasticity model when $\gamma_1 = 0$. The relative efficiency measure is computed using (3.3) and (3.4). Similar to the linear models, the relative efficiency in this case is independent of the design matrix and has exactly the same feature shown in Table I. Since the three-step approach introduces extra variation to the RRQ estimator, the relative efficiency of the RRQ estimator is affected as can be clearly seen from a comparison of Table I and Table II. Since the relative efficiency measure is a continuous function of γ_1, we expect that the RRQ estimator has similar performance when γ_1 is close to zero.

TABLE II

The Relative Efficiency RE(RRQ, RQ) for Linear Heteroscedastic Models ($k = 1, \gamma_1 = 0$)

τ	Normal	D-Exp	Cauchy	Exponential	Lognormal
0.05	1.0880	0.8167	0.5644	5.4259	2.8006
0.10	1.1511	0.8551	0.6326	2.7693	2.1222
0.20	1.1208	0.8434	0.7699	1.5161	1.4736
0.30	1.0598	0.8307	0.8896	1.1652	1.1946
0.40	1.0156	0.8679	0.9711	1.0389	1.0389
0.50	1.0000	1.0000	1.0000	1.0000	1.0000
0.60	1.0156	0.8679	0.9711	1.0046	0.9638
0.70	1.0598	0.8307	0.8896	1.0143	0.9312
0.80	1.1208	0.8434	0.7699	1.0192	0.8812
0.90	1.1511	0.8551	0.6326	0.9703	0.7836
0.95	1.0880	0.8167	0.5644	1.0910	0.6940

Efficiency comparison becomes difficult when heteroscedasticity is in existence, because the relative efficiency formula depends on the design matrix. Table III lists the relative efficiency of the RRQ estimator with respect to the RQ estimator for a small-scale Monte Carlo simulation. In the simulation, we used the linear heteroscedastic model

$$y_i = x_i + (1 + \gamma_1 x_i)\, e_i,$$

where x_i is sampled from N(3,2) distribution and γ_1 takes five values, i.e., 0, 0.1, 0.3, 0.5, 0.8. Two error distributions, i.e., normal distribution and double exponential distribution, are used in the simulation. For each error type, the e_i's are rescaled such that the median is zero and the median of their absolute value is one. The Monte Carlo simulation replicates 1000 times for relative small sample sizes, $N = 20$ and $N = 50$, respectively. Table III shows that for most of the parameter combinations, the RRQ estimator is less efficient than the RQ estimator, although the efficiency loss is only a few percent and therefore minor. When sample size increases from 20 to 50, the relative efficiency of the RRQ estimator improves sharply, showing the sensitivity of efficiency to sample sizes. Overall, the simulation results show that the cost of avoiding quantile plane crossing is low in these simulated cases.

3.3. *Application*

In this subsection, we present an application of the RRQ estimators in constructing calibration intervals using the esterase assay data from Carroll and Ruppert (1988, pp. 46–47). The data set has two variables, the concentration of an enzyme esterase and the number of bindings counted in binding experiments. A total of 113 observations were collected in the data set and are plotted in Fig. 1, with two quantile lines (75th and 90th) crossed.

Carroll and Ruppert (1988) studied this data set using a simple linear model

$$E(count \mid esterase) = \beta_0 + \beta_1 [esterase].$$

Since the counts have large variations when the concentration of esterase is large, the spread of the error term is assumed to be linear in concentration. The data set may be modeled into the linear heteroscedastic model

$$count_i = \beta_0 + \beta_1 [esterase]_i + (\gamma_0 + \gamma_1 [esterase]_i)\, \varepsilon_i,$$

where the ε_i's are treated as i.i.d. random variables.

The main objective of fitting this data set is to construct calibration confidence intervals, i.e., finding the confidence interval for the regression

TABLE III

The Relative Efficiency RE(RRQ, RQ) for Linear Heteroscedastic Models ($k=1, \gamma_1>0$)

N	τ	γ_1 0.0	0.1	0.3	0.5	0.8
		I. Normal error distribution				
20	0.60	1.0752	1.0709	1.0855	1.0901	1.0678
20	0.70	1.0218	1.0281	1.0539	1.0477	1.0817
20	0.80	1.0378	1.0422	1.0678	1.0653	1.0756
20	0.90	1.0323	1.0462	1.0520	1.0539	1.0782
20	0.95	1.0114	1.0280	1.0105	1.0167	1.0050
50	0.60	1.0316	1.0392	1.0456	1.0482	1.0530
50	0.70	1.0057	1.0131	1.0078	1.0272	1.0187
50	0.80	1.0184	1.0198	1.0324	1.0304	1.0379
50	0.90	1.0098	1.0105	1.0121	1.0306	1.0203
50	0.95	0.9973	1.0026	1.0093	1.0056	1.0075
		II. Double exponential error distribution				
20	0.60	1.0314	1.0490	1.0526	1.0606	1.0654
20	0.70	1.0108	1.0186	1.0174	1.0126	1.0252
20	0.80	1.0204	1.0298	1.0277	1.0388	1.0383
20	0.90	1.0166	1.0205	1.0325	1.0360	1.0389
20	0.95	1.0172	1.0200	1.0168	1.0148	1.0137
50	0.60	1.0125	1.0200	1.0187	1.0197	1.0239
50	0.70	1.0012	0.9982	1.0043	0.9986	1.0014
50	0.80	1.0085	1.0082	1.0097	1.0094	1.0164
50	0.90	1.0075	1.0104	1.0086	1.0148	1.0138
50	0.95	1.0024	1.0025	1.0024	1.0061	1.0031

variable given the response variable. The calibration interval can be used to estimate the concentration of a new enzyme esterase. To construct the calibration confidence intervals, one needs to estimate the quantile lines (or curves). Because of the heteroscedasticity in the response variable, the ordinary least squares method does not provide satisfactory results. Carroll and Ruppert (1988) used the generalized least squares method under the assumption of normal errors. In this section, the RRQ and RQ estimators are used to build the calibration intervals without assumption of specific error distributions.

First, we compute a median regression of the count on concentration, including an intercept term. The estimator is $\hat{\beta}=(-27.4893, 17.8997)'$. Then we compute the residuals $\hat{r}_i$ and regress the absolute residuals on the concentration. The estimated $\hat{\gamma}$ is $\hat{\gamma}=(14.0057, 1.9287)'$. The third step is to

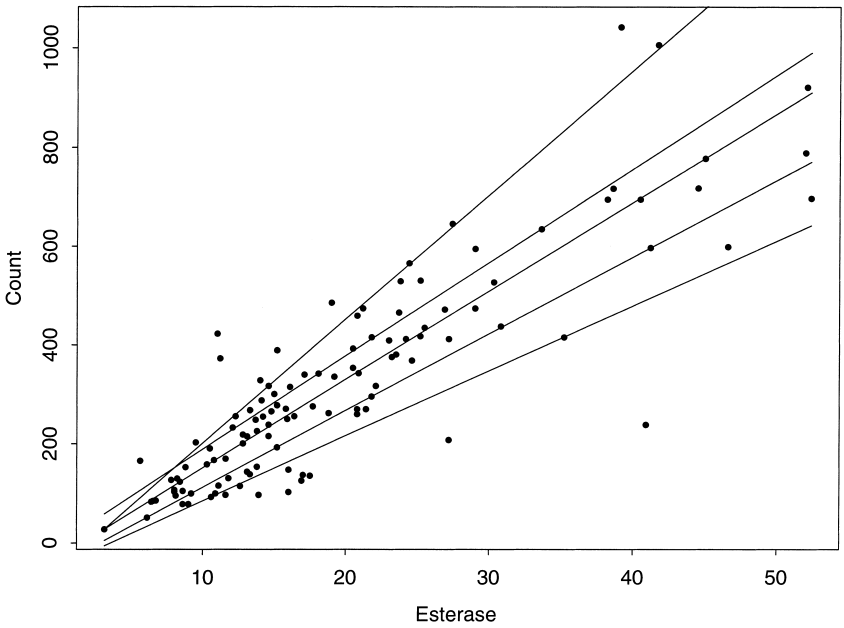

FIG. 1. The unrestricted 0.10, 0.25, 0.5, 0.75, and 0.90th regression quantile lines for the esterase count data from Carroll and Ruppert (1988). The 0.75th and 0.90th regression quantiles cross once.

estimate the $\hat{c}_\tau$'s for different τ's as suggested by He (1997). The estimates and the final RRQs, $\hat{\beta}(\tau)$, are listed in Table IV. RQ estimators, $\tilde{\beta}(\tau)$, are also tabulated in the same table for comparison purposes.

Table IV shows the difference between the RQ an RRQ estimators. The RRQ estimator $\hat{\beta}(\tau)$ is increasing in τ, which assures that the regression quantile lines do not cross each other at all observed x_i's, as is shown in Fig. 2. These quantile lines can be used to construct calibration intervals in the whole range responses. In contrast, the unrestricted RQ lines cross at least once occurs at the point when the esterase concentration is about 9

TABLE IV

RRQ and RQ Estimators for the Esterase Assay Data

τ	$\hat{c}_\tau$	$\hat{\beta}(\tau)'$	$\tilde{\beta}(\tau)'$
0.10	−2.1679	(−57.8527, 13.7185)	(−46.1789, 13.1301)
0.25	−1.1904	(−44.1622, 15.6038)	(−42.7151, 15.5271)
0.50	0.0000	(−27.4893, 17.8997)	(−27.4893, 17.8997)
0.75	0.7747	(−16.6385, 19.3940)	(−0.1485, 18.9032)
0.90	2.2227	(3.6415, 22.1867)	(−49.9043, 25.1304)

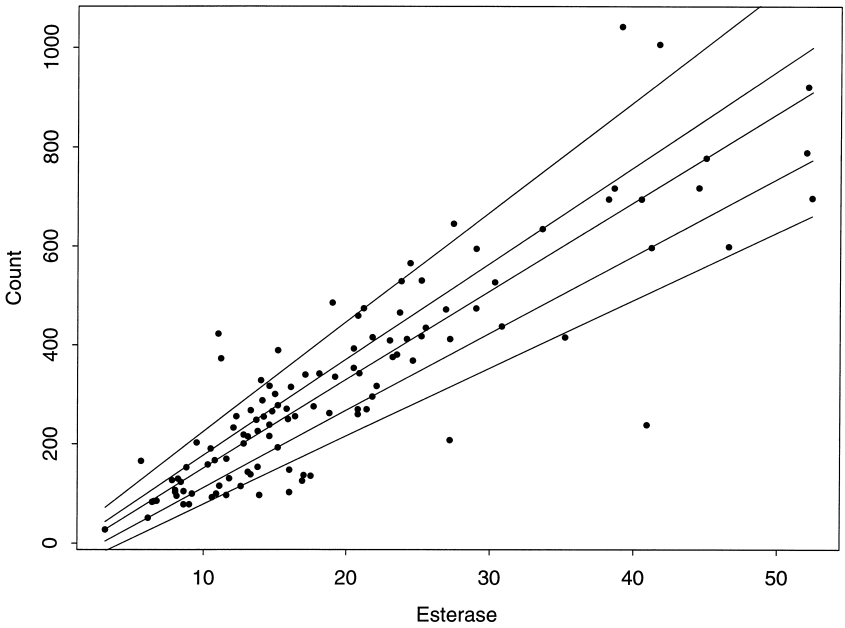

FIG. 2. The restricted 0.10, 0.25, 0.5, 0.75, and 0.90th regression quantile lines for the esterase count data from Carrol and Ruppert (1988). No quantile line crossing in the interior domain of the design points is observed.

units, which corresponds to the 25th quantile of the response variable. This makes the concentration prediction inaccurate for all bindings counts that are less than 200. Weighted regression quantiles may help to shift the crossing point, but cann't completely eliminate it. The RRQ estimation shows as a natural solution for the quantile line crossing problem.

4. CONCLUSION

In this article, we examine the properties of the RRQ estimators proposed by He (1997) in comparison with the RQ estimators of Koenker and Bassett (1978). As a natural extension of the sample quantiles of the location model, the regression quantiles somehow do not inherit the properties owned by the sample quantile in the sense that the predicted quantile planes are likely to cross each other in finite sample cases. The RRQs are designed to overcome this troublesome quantile plane crossing problem.

We studied the RRQ estimator for two models, the linear model and the linear heteroscedastic model. For the first model, we found that RRQs have

an important characteristic, in addition to the ability to overcome quantile plane crossing. They are in general more efficient than the RQ estimators for unimode distributions, such as normal, double exponential, Cauchy and Student's t-distributions, although this property does not hold for asymmetric error distributions. For heteroscedastic linear models, we show that the RRQ estimators perform quite well for small sample cases, even for small sample size of 20 or 50. An application of the RRQs to construct calibration intervals is illustrated by an example using the esterase assay data from Carroll and Ruppert (1988).

5. APPENDIX. PROOFS

The proof of all the theorems are based on the following two lemmas organized by Koenker and Zhao (1996). Details for showing each theorem will be given later.

LEMMA A.1. *Let*

$$V_n(\Delta) = n^{-1/2} \sum_1^n z_i \psi_\tau(u_i - F^{-1}(\tau) - n^{-1/2} x_i' \Delta),$$

where $\psi_\tau(u) = \tau - I_{\{u<0\}}$, $0 < \tau < 1$, z_i *and* x_i *are i.i.d. and independent of* u_i. *Suppose*

(i) u_i *are i.i.d. random variables with continuous density*;

(ii) $n^{-1/2} \max_{i \leqslant n} \|x_i\| = o_p(1)$.

(iii) $E\|z_1\|^2 < \infty$ *and* $E\|x_1\|^2 < \infty$.

Then

$$\sup_{\|\Delta\| \leqslant L} \|V_n(\Delta) - V_n(0) - EV_n(\Delta) + EV_n(0)\| = o_p(1),$$

for any $0 < L < \infty$.

The proof of this lemma adopts ideas from Bickel (1976) and Ruppert and Carroll (1980). The details of the proof are given in Koenker and Zhao (1996).

LEMMA A.2. *Suppose* $V_n(\Delta)$ *satisfies*

(i) $\Delta' V_n(\lambda\Delta) \leqslant \Delta' V_n(\Delta)$, *For* $\lambda \geqslant 1$;

(ii) $\sup_{\|\Delta\| \leqslant L} \|V_n(\Delta) - A_n + G\Delta\| = o_p(1)$ *for a random vector* A_n, *nonrandom matrix* G *and any* $0 < L < \infty$;

(iii) *Δ_n is a random vector satisfying* $\|V_n(\Delta_n)\| = o_p(1)$.

Then $\|\Delta_n\| = O_p(1)$ *and* Δ_n *has the representation*

$$\Delta_n = G^{-1} A_n + o_p(1).$$

The proof of this lemma uses ideas from Jurečková (1975) and is refined by Koenker and Zhao (1994, 1996), where details of the proof can be found. Now by using the above two lemmas, we will prove Theorem 2.1 and Theorem 2.2.

Proof of Theorem 2.1. First of all, let's construct a function $V_n(\Delta)$ which satisfies the conditions of Lemma A.1. Given

$$\hat{r}_i - \hat{\beta}_0(\tau) = \varepsilon_i - F^{-1}(\tau) - (\hat{\beta}_0(\tau) - \beta_0 - F^{-1}(\tau)) - x'_{1i}(\hat{\beta}_1 - \beta_1).$$

Denote $\Delta_{1n} = \sqrt{n}\,(\hat{\beta}_1 - \beta_1)$, $\Delta_{2n} = \sqrt{n}\,(\hat{\beta}_0(\tau) - \beta_0 - F^{-1}(\tau))$ and define

$$V_n(\Delta) = n^{-1/2} \sum_1^n \psi_\tau(\varepsilon_i - F^{-1}(\tau) - n^{-1/2}(x'_{1i}\Delta_1 + \Delta_2)),$$

which satisfies the conditions of Lemma A.1. Hence

$$\sup_{\|\Delta\| \leqslant L} \|V_n(\Delta) - V_n(0) - EV_n(\Delta) + EV_n(0)\| = o_p(1).$$

Now, it is easy to verify that $EV_n(0) = 0$ and

$$\begin{aligned} EV_n(\Delta) &= n^{-1/2} \sum_1^n (\tau - F(F^{-1}(\tau) + n^{-1/2}(x_i\Delta_1 + \Delta_2)) \\ &= -f(F^{-1}(\tau))(\mu'_1\,\Delta_1 + \Delta_2) + o(1). \end{aligned}$$

Combined with condition (i) of Theorem 2.1, we have

$$\sup_{\|\Delta\| \leqslant L} \|V_n(\Delta_1, \Delta_2) - V_n(0) + f(F^{-1}(\tau))(\mu'_1\,\Delta_1 + \Delta_2)\| = o_p(1).$$

By the definition of $\hat{\beta}_0(\tau)$, it follows the Ruppert and Carroll (1980) and Koenker and Zhao (1996) that $V_n(\Delta_{1n}, \Delta_{2n}) = o_p(1)$. Here the argument is that

$$n^{-1/2} \sum_1^n x_i \psi_\tau(y_i - x_i\hat{\beta}) = o_p(1)$$

if $\hat{\beta}$ is the minimizer of $\sum_1^n \rho_\tau(y_i - x'_i b)$.

Further, using the monotone property of $\psi_\tau(u)$, one can easily show that the first condition of Lemma A.2 is valid. Therefore by Lemma A.2

$$f(F^{-1}(\tau))(\mu_1' \Delta_{1n} + \Delta_{2n}) = V_n(0) + o_p(1),$$

or

$$\Delta_{2n} = \frac{1}{f(F^{-1}(\tau))} n^{-1/2} \sum_1^n \psi_\tau(\varepsilon_i - F^{-1}(\tau)) - \mu_1' \Delta_{1n} + o_p(1).$$

Theorem 2.1 is thus proved. ∎

Proof of Corollary 2.1. Suppose $\mu_1 = 0$, then $\hat{\beta}_0(\tau)$ has a Bahadur-type representation which has nothing to do with the preliminary estimator $\hat{\beta}_1$. By using the central limit theorem, one can easily obtain the asymptotic distribution. ∎

Proof of Theorem 2.2. The proof is similar to that of Theorem 2.1. First we are going to define a function $V_n(\Delta)$ as done in the proof of Theorem 2.1. let $\sigma_i = x_i'\gamma > 0$, $\Delta_{1n} = \sqrt{n}\,(\hat{\beta} - \beta)$ and $\Delta_{2n} = \sqrt{n}\,(\hat{\gamma} - \gamma)$. Thus

$$\begin{aligned} |\hat{r}_i| - x_i'\hat{\gamma} &= |\sigma_i \varepsilon_i - n^{-1/2} x_i' \Delta_{1n}| - x_i'\hat{\gamma} \\ &= \sigma_i(|\varepsilon_i| - 1) - n^{-1/2} x_i'(sgn(\varepsilon_i)\,\Delta_{1n} + \Delta_{2n}) + o_p(n^{-1/2}), \end{aligned}$$

where the $o_p(n^{-1/2})$ part can be combined into $O_p(n^{-1/2})$. Define

$$V_n(\Delta) = n^{-1/2} \sum_1^n x_i \psi_{1/2}(|\varepsilon_i| - 1 - n^{-1/2}\sigma_i^{-1} x_i'(\mathrm{sgn}(\varepsilon_i)\,\Delta_1 + \Delta_2)),$$

which satisfies the conditions of Lemma A.1. Now

$$\begin{aligned} &E[\psi_{1/2}(|\varepsilon_i| - 1 - n^{-1/2}\sigma_i^{-1} x_i'(\mathrm{sgn}(\varepsilon_i)\,\Delta_1 + \Delta_2)) \mid x_i] \\ &\quad = \tfrac{1}{2} - \int_{-1 - n^{-1/2}\sigma_i^{-1} x_i'(-\Delta_1 + \Delta_2)}^{1 + n^{-1/2}\sigma_i^{-1} x_i'(\Delta_1 + \Delta_2)} f(u)\,du \\ &\quad = -n^{-1/2}\sigma_i^{-1} x_i'(f(1)(\Delta_1 + \Delta_2) + f(-1)(-\Delta_1 + \Delta_2)) + o(n^{-1/2}). \end{aligned}$$

Thus

$$\begin{aligned} E[V_n(\Delta) \mid x_i] &= -n^{-1} \sum_1^n \frac{x_i x_i'}{\sigma_i} (f(1)(\Delta_1 + \Delta_2) + f(-1)(-\Delta_1 + \Delta_2)) + o(1) \\ &= -Q_1((f(1) - f(-1))\,\Delta_1 + (f(1) + f(-1))\,\Delta_2) + o(1). \end{aligned}$$

By Lemma A.1 and the first condition (i) of Theorem 2.2, we have

$$\sup_{\|\Delta_2\| \leqslant L} \| Vn(\Delta_{1n}, \Delta_2) - V_n(0) + Q_1(f(1) - f(-1))\, \Delta_{1n} + (f(1) + f(-1))\, \Delta_2) \| = o_p(1).$$

Similarly, by using Lemma A.2, we have the representation

$$\Delta_{2n} = \frac{Q_1^{-1}}{f(1) + f(-1)}\, n^{-1/2} \sum_1^n x_i \psi_{1/2}(|\varepsilon_i| - 1) - \frac{f(1) - f(-1)}{f(1) + f(-1)}\, \Delta_{1n} + o_p(1).$$

This proves the first part of the theorem.

Now let's show the second part, i.e., the representation of $\hat{c}_\tau$. Set $\Delta_{1n} = \sqrt{n}\,(\hat{\beta} - \beta)$, $\Delta_{2n} = \sqrt{n}\,(\hat{\gamma} - \gamma)$ and $\Delta_{3n} = \sqrt{n}\,(\hat{c}_\tau - F^{-1}(\tau))\,\hat{\gamma}$. Then

$$\hat{r}_i - \hat{c}_\tau x_i' \hat{\gamma} = (x_i' \gamma)(\varepsilon_i - F^{-1}(\tau)) - n^{-1/2} x_i' (\Delta_{1n} + F^{-1}(\tau)\, \Delta_{2n} + \Delta_{3n}).$$

Define

$$V_n(\Delta) = n^{-1/2} \sum_1^n (x_i' \gamma)\, \psi_\tau(\varepsilon_i - F^{-1}(\tau) - n^{-1/2} \sigma_i^{-1} x_i' (\Delta_1 + F^{-1}(\tau)\, \Delta_2 + \Delta_3)).$$

We can get

$$EV_n(\Delta) = -f(F^{-1}(\tau))\, \mu'(\Delta_1 + F^{-1}(\tau)\, \Delta_2 + \Delta_3) + o(1).$$

By Lemma A.1 we have

$$\sup_{\|\Delta_3\| \leqslant L} \| V_n(\Delta_{1n}, \Delta_{2n}, \Delta_3) - V_n(0) + f(F^{-1}(\tau))\, \mu'(\Delta_{1n} + F^{-1}(\tau)\, \Delta_{2n} + \Delta_3) \| = o_p(1).$$

Then by applying Lemma A.2, we get the Bahadur-type representation

$$\mu'\, \Delta_{3n} = \frac{\gamma'}{f(F^{-1}(\tau))}\, n^{-1/2} \sum_1^n x_i \psi_\tau(\varepsilon_i - F^{-1}(\tau)) - \mu'(\Delta_{1n} + F^{-1}(\tau)\, \Delta_{2n}) + o_p(1).$$

If we note that $\Delta_{2n} = O_p(1)$, i.e., $\hat{\gamma} = \gamma + O_p(n^{-1/2})$ and the definition of Δ_{3n}, it follows that

$$\sqrt{n}\,(\hat{c}_\tau - F^{-1}(\tau)) = \frac{\gamma'}{f(F^{-1}(\tau))(\mu'\gamma)}\, n^{-1/2} \sum_1^n x_i \psi_\tau(\varepsilon_i - F^{-1}(\tau))$$
$$- \frac{\mu'}{\mu'\gamma}(\sqrt{n}\,(\hat{\beta} - \beta) + F^{-1}(\tau)\sqrt{n}\,(\hat{\gamma} - \gamma)) + o_p(1).$$

Finally, for the RRQ estimator $\hat{\beta}(\tau)$, if the initial estimator $\hat{\beta}$ is the least median estimator as suggested by He (1997), it should have a Bahadur representation

$$\sqrt{n}\,(\hat{\beta} - \beta) = \frac{1}{f(0)}\, Q_1^{-1} n^{-1/2} \sum_1^n x_i \psi_{1/2}(\varepsilon_i) + o_p(1).$$

The representation of $\hat{\beta}(\tau)$ thus follows from the previous results. This proves Theorem 2.2. ∎

Proof of Corollary 2.2. The proof can be obtained using the central limit theorem. ∎

REFERENCES

1. P. J. Bickel, One-step Huber estimates in the linear model, *J. Amer. Statist. Assoc.* **70** (1975), 428–434.
2. R. J. Carroll and D. Ruppert, "Transformation and Weighting in Regression," Chapman & Hall, London, 1988.
3. B. Efron, Regression percentiles using asymmetric squared error loss, *Statist. Sinica* **1** (1991), 93–125.
4. G. Gutenbrunner and J. Jurečková, Regression rank scores and regression quantiles, *Ann. Statist.* **20** (1992), 305–330.
5. G. Gutenbrunner, J. Jurečková, R. Koenker, and S. Portnoy, Tests of linear hypotheses based on regression rank scores, *J. Nonparametr. Statist.* **2** (1993), 307–331.
6. X. He, Quantile curves without crossing, *Amer. Statist.* **51** (1997), 186–192.
7. J. Jurečková, Asymptotic relations of M-estimates and R-estimates in linear regression models, *Ann. Statist.* **5** (1977), 464–472.
8. R. Koenker, A note on L-estimates for linear models, *Statist. Probab. Lett.* **2** (1984), 323–325.
9. R. Koenker and G. Bassett, Regression quantiles, *Econometrica* **46** (1978), 33–50.
10. R. Koenker and G. Bassett, Robust tests for heteroscedasticity based on regression quantiles, *Econometrica* **50** (1982), 43–61.
11. R. Koenker and S. Portnoy, L-estimation for linear models, *J. Amer. Statist. Assoc.* **82** (1987), 851–857.
12. R. Koenker and Q. Zhao, L-estimation for linear heteroscedastic models, *J. Nonparametr. Statist.* **3** (1994), 223–235.

13. R. Koenker and Q. Zhao, Conditional quantile estimation and inference for ARCH models, *Econometric Theory* **12** (1996), 793–813.
14. S. Portnoy, Asymptotic behavior of regression quantiles in non-stationary, dependent cases, *J. Multivariate Anal.* **38** (1992), 100–113.
15. D. Ruppert and R. J. Carroll, Trimmed least square estimation in the linear models, *J. Amer. Statist. Assoc.* **75** (1980), 828–838.
16. A. H. Welsh, The trimmed mean in the linear model, *Ann. Statist.* **85** (1987), 729–743.
17. K. Q. Zhou and S. Portnoy, Direct use of regression quantiles to construct confidence sets in linear models, *Ann. Statist.* **24** (1996), 287–306.