

EMPIRICAL REGRESSION QUANTILE

Toshiyuki Sueyoshi
The Ohio State University

(Received April 2, 1990; Revised December 3, 1990)

Abstract This study proposes a new use of goal programming for empirically estimating a regression quantile hyperplane. The approach can yield regression quantile estimates that are less sensitive to not only non-Gaussian error distributions but also a small sample size than conventional regression quantile methods. The performance of regression quantile estimates is compared with least absolute value estimates in a simulation study.

1. Introduction

An underlying assumption on the use of least squares (LS) method is that errors distribute normally with a common variance. The assumption seems to be a compound of truth, assumption and convenience in use. It is already well known that the normal error distribution is useful over a wide variety of research areas (e.g., gambling, physics and biology). However, it is also widely accepted that even a few outliers occurring with low probability can cause a serious departure from normality. In addition, the same variance is not maintained by all independent variables.

In the last twenty years we have witnessed an increasing interest in robust estimation among researchers who have recognized that the distribution of a real data set is different from the normal and has often longer tails than the normal in most cases. [See, for instance, Hogg [13] and Huber [14] for their descriptions regarding robust estimation.]

In many cases these researchers are really interested in estimates of various quantiles (percentiles) than LS estimates because the former estimates can reduce the influence of outliers. Thus, quantile estimates can be replaced for the use of the LS method in regression analysis when a data set is contaminated by outliers. [See, for instance, Amemiya [1] and [13] for their detailed discussions concerning robustness of quantile estimates.]

In order to specify the research purpose more clearly, this article starts with fitting a linear regression hyperplane that is mathematically defined as

$$y_i = X_i\beta + \epsilon_i, \quad i = 1, \dots, n,$$

where y_i is the i th observed dependent variable, $\beta = (\beta_1, \beta_2, \dots, \beta_m)^T$ is a column vector indicating parameter coefficients to be measured, $X_i = (x_{i1}, x_{i2}, \dots, x_{im})$ is the i th row vector of an observed design matrix and ϵ_i is an error associated with the i th observation.

Usually, ϵ_i is assumed to be normally distributed with a common variance. In order to drop the assumption regarding the error distribution, this study will develop an algorithmic framework that produces a linear regression hyperplane on the $(100p)$ th percentile of the y distribution. That is, this study attempts to find parameter estimates $\check{\beta} = (\check{\beta}_1, \check{\beta}_2, \dots, \check{\beta}_m)$ that can satisfy

$$N(\check{\beta})/n = p \tag{1}$$

where $N(\tilde{\beta})$ indicates the number of observations (i) such that $y_i < X_i\tilde{\beta}$.

In the history of statistics, Koenker and Bassett [4, 15] have first proposed a linear programming (LP) method referred to as “regression quantile” that can measure parameter estimates of the regression quantile hyperplane. They also investigated conditions for a unique solution of the regression quantile and established asymptotical theory concerning the variance-covariance matrix of quantile estimates.

Clearly, the most important quantile value is $p = 50\%$, where fifty percent sample observations on y are above and the remaining fifty percent are below the regression quantile hyperplane. Koenker and Bassett termed specially the regression quantile with $p = 50\%$ as “regression median.” The regression median is also often referred to as “least absolute value (LAV) estimation” because it minimizes the sum of absolute value errors. [Hogg [13] has also proposed another method for estimating percentile regression.]

A drawback of the regression quantile proposed by Koenker and Bassett [4, 15] is that quantile estimates satisfy (1) asymptotically (i.e., a large sample size) and approximately satisfy (1) when a data set has a small sample size. The estimation property due to the small sample can be explained by the fact that any LP solution exists on an extreme point (s). Since the extreme point can be determined by the combination of sample observations, the regression quantile always needs several data points on its regression hyperplane. For instance, when a linear regression model with three parameters is fitted to a data set, at least three sample observations are often required to be on the regression quantile hyperplane. That is the rationale as to why (1) cannot be maintained by the regression quantile when a sample size is small.

This study will present an empirical use of regression quantile that can yield a linear regression hyperplane exactly on the $(100p)$ th percentile of the y distribution. Here, the term “empirical” is used because the regression quantile proposed in this study can be applicable to any real data set, maintaining the condition of (1). The approach does not require the assumptions on normality and/or a large sample size. This study admits that there are several alternatives (e.g. [5] and [13]) to produce the regression quantile estimates. However, the approach of this study is computationally efficient and contains less assumptions than these conventional methods. [It is hoped that the new method will be one small step in the study of robust estimation.]

The remainder of this article is organized as follows. Section 2 presents the mathematical notation of regression quantile and extend it to the LP formulation originally proposed in [15]. Section 3 describes an empirical use of LP for measuring exactly the regression quantile estimates. Section 4 documents the new result of a Monte Carlo simulation study in which the performance of the approach is compared with the original regression quantile. Conclusion and future extensions are summarized in Section 5.

2. Regression Quantile

Koenker and Bassett [15] formally defined the p th regression quantile as any solution to the following minimization problem:

$$\underset{\beta}{\text{minimize}} \left[\sum_{i \in \{i: y_i \geq X_i\beta\}} p|y_i - X_i\beta| + \sum_{i \in \{i: y_i < X_i\beta\}} (1-p)|y_i - X_i\beta| \right]. \quad (2)$$

By the result of Charnes *et al.* [9], (2) can be transformed into the following goal programming (GP) problem:

$$\underset{\delta}{\text{minimize}} \quad \sum_{i=1}^n [p\delta_i^+ + (1-p)\delta_i^-]$$

$$\begin{aligned} \text{subject to } & X_i\beta + \delta_i^+ - \delta_i^- = y_i, \quad i = 1, \dots, n, \\ & \delta_i^+ \geq 0 \quad \text{and} \quad \delta_i^- \geq 0, \quad i = 1, \dots, n. \end{aligned} \quad (3)$$

Here, δ_i^+ and δ_i^- are positive and negative deviations related to the i th observation, respectively.

[See Charnes and Cooper [6, 7] for a detailed discussion on the nonlinear condition $\delta_i^+ \cdot \delta_i^- = 0, i = 1, \dots, n$.]

2.1 Property of Dual Variables

The dual form of (3) becomes

$$\begin{aligned} & \text{maximize } \sum_{i=1}^n w_i y_i \\ & \text{subject to } \sum_{i=1}^n w_i x_{ij} = 0, \quad j = 1, \dots, m, \\ & \quad - (1 - p) \leq w_i \leq p, \quad i = 1, \dots, n, \end{aligned} \quad (4)$$

where w_i indicates the i th dual variable and x_{ij} is the j th component of the i th row vector $X_i = (x_{i1}, \dots, x_{ij}, \dots, x_{im})$.

An important property of the dual variable is that it indicates the locational relationship between sample observations and the p th regression quantile hyperplane estimated by (3). That is, if $w_i = -(1 - p)$, $-(1 - p) < w_i < p$, and $w_i = p$; then the corresponding i th observation is located below, on and above the regression hyperplane, respectively. The property follows from the theory of complementary slackness condition of linear programming. [See Sueyoshi and Chang [16] for their detailed discussion on the property of dual variables derived from (4) with $p = 50\%$ (i.e., LAV estimation).]

2.2 Degeneracy and Multiple Optimal Solutions

A common assumption concerning the p th regression quantile is that the problem of degeneracy does not occur at an optimal solution of (3). Hence, (3) can avoid multiple optimal solutions with a same objective value. The degeneracy occurs when optimal δ_i^+ or δ_i^- for some i becomes a basic variable and equals zero in primal form (3). Meanwhile, in dual form (4), a degenerated extreme point is a sample observation with $w_i = p$ or $w_i = -(1 - p)$ and is on the p th regression quantile hyperplane. As described before, when one measures m unknown parameters of a linear regression model, at least m sample observations are required to consist of a regression hyperplane. If more than m sample observations are on the regression hyperplane, the degeneracy often occurs. In order to deal with the occurrence of degeneracy, when it happens, one needs to develop an algorithmic scheme which can select the location of a regression quantile hyperplane in a sample space more flexibly than (3). That is, the regression hyperplane can be determined on not only an extreme point but also other points in the sample space.

3. Empirical Regression Quantile

As first presented in [5], a difficulty of (3) is that it yields a regression quantile hyperplane approximately on the $(100p)$ th percentile of the y distribution. That is, although one expects that there are exactly np sample observations below and $n(1 - p)$ observations above the p th

regression quantile hyperplane, (3) frequently yields a result that differs from the expectation, particularly when a data set has a small sample size. For instance, (3) with $p = 50\%$ fitted to a data set with 50 sample observations produces an estimation result in which 26, 5 and 19 data points are above, on and below the 50%th regression quantile hyperplane. The regression quantile is approximately acceptable. The estimation problem often occurs to any regression models with or without a parameter representing a constant. [Hereafter, this study uses β_0 to express the constant parameter.]

This study proposes a two-stage approach to estimate exactly the p th regression quantile. Hereafter, this article refers to problem (3) proposed by Koenker and Bassett [15] as “original regression quantile” while referring the estimation method proposed in this study as “empirical regression quantile (ERQ).” The first stage of ERQ starts with the original regression quantile (3). Using dual variables produced by (3), a data set is classified into two groups of sample observations (i.e., observations above and below the p th regression hyperplane). The second stage consists of two GP models that determine the location of the p th regression quantile hyperplane between the dichotomously classified sample observations.

3.1 Classification of Sample Observations

In order to describe the use of dual variables for dichotomizing sample observations, ERQ returns to the optimal objective value of (3) and (4) that can be expressed by

$$z^* = \sum_{i=1}^n [p\delta_i^+ + (1-p)\delta_i^-] = \sum_{i=1}^n w_i y_i. \quad (5)$$

where the symbol “*” indicates optimality. The derivative of (5) yields

$$w_i = \partial z^* / \partial y_i. \quad (6)$$

The optimal dual variable w_i thus indicates the rate of change of the objective value z^* with one unit increase in y_i . Assuming that $w = (w_1, w_2, \dots, w_n)$ is a unique, nondegenerate solution to (4), the interpretation on the rate of change of z^* can be classified as follows:

- (a) if $0 \leq w_i \leq p$, then each unit increase (decrease) in y_i produces w_i increment (decrement) in z^* , and
- (b) if $-(1-p) \leq w_i < 0$, then each unit increase (decrease) in y_i yields w_i decrement (increment) in z^* .

Thus, the optimal w_i is associated with the change of δ_i^+ or δ_i^- . Therefore, the examination on the dual variable can provide information on the optimal sum of a new regression hyperplane that can be determined by slightly changing the previously optimal p th regression quantile hyperplane. [The dual variable associated with each observation is not available in the statistical test based on LS regression. The approach proposed in this study is similar to the sensitivity analysis of linear programming in spirit.]

Now, to express the data classification in a formal manner, this article uses a new symbol $k(k = 1, \dots, n)$ that indicates the descending order of the dual variables, i.e.,

$$w_1 \geq w_2 \geq \dots \geq w_k \geq \dots \geq w_n. \quad (7)$$

Note that the symbol “i” denotes an observed order of sample observations while the “k” indicates the descending order of the dual variables of (4).

Following the order of (7), sample observations are classified into the two groups as follows:

$G_A = \{i \mid \text{the } i\text{th observation has } w_i \text{ that belongs to the top } (1-p)\text{th percentile in (7)}\}$
 and
 $G_B = \{i \mid \text{the } i\text{th observation has } w_i \text{ that belongs to the bottom } p\text{th percentile in (7)}\}.$

The sample size of G_A is $n(1-p)$ while that of G_B is np . It is important to note that for some cases the approach cannot clearly separate a data set into G_A and G_B . In this case, one needs to round off fractions to an integral value. For instance, the combination between $n = 26$ and $p = .3$ yields $np = 7.8$ and $n(1-p) = 18.2$. The round-off method determines $np \cong 8$ for G_B and $n(1-p) \cong 18$ for G_A . Furthermore, the combination between $n = 25$ and $p = .5$ yields $np = 12.5$ and $n(1-p) = 12.5$. In the case both G_A and G_B contain 13 data points. A sample observation, whose dual variable is the middle (i.e., $(n+1)/2$) on the order of (7), belongs to both G_A and G_B .

3.2 Goal Programming Models

After dichotomizing a data set into G_A and G_B , ERQ uses the following GP models:

$$\text{minimize } \sum_{i \in G_A} \delta_i^+ + L \sum_{i \in G_B} \delta_i^+ \quad (8)$$

$$\text{subject to } X_i\beta + \delta_i^+ = y_i, i \in G_A \quad (8a)$$

$$X_i\beta + \delta_i^+ - \delta_i^- = y_i, i \in G_B \quad (8b)$$

$$\delta_i^+ \geq 0, \text{ and } \delta_i^- \geq 0, \quad (8c)$$

and

$$\text{minimize } L \sum_{i \in G_A} \delta_i^- + \sum_{i \in G_B} \delta_i^- \quad (9)$$

$$\text{subject to } X_i\beta + \delta_i^+ - \delta_i^- = y_i, i \in G_A \quad (9a)$$

$$X_i\beta - \delta_i^- = y_i, i \in G_B \quad (9b)$$

$$\delta_i^+ \geq 0, \text{ and } \delta_i^- \geq 0. \quad (9c)$$

Here all the symbols used in (8) and (9) are the same as described in (2), except L indicating a non-Archimedean large number.

Model (8) is used to yield a bottom hyperplane for G_A where “bottom” indicates that all the sample observations in G_A are above or on the regression hyperplane. The first set of constraints (8a) produce the bottom hyperplane because only positive deviations $\delta_i^+, i \in G_A$ are minimized in (8). Meanwhile, the second set of constraints (8b) prevent any observation in G_B from being above the bottom hyperplane of G_A because $\delta_i^+, i \in G_B$ are weighted by L in (8).

Conversely, (9) can estimate an upper regression hyperplane for G_B where “upper” indicates that all the sample observations in G_B are below or on the regression hyperplane. The second set of constraints (9b) can yield the upper regression hyperplane for G_B because only negative deviations $\delta_i^-, i \in G_B$ are minimized in (9). On the other hand, the first set of constraints (9a) prevent any sample observation in G_A from being below the upper hyperplane of G_B because $\delta_i^-, i \in G_A$ are weighted by L in (9).

3.3 Algorithm

The algorithm for the p th ERQ can be formally defined as the following steps:

1. Solve problem (3). Then, classify all sample observations into G_A and G_B using dual variables produced by (3).

2. Solve problem (8) to estimate a bottom hyperplane of G_A that is represented by $y_A = X\beta_A$.
3. Solve problem (9) to estimate an upper hyperplane of G_B that is represented by $y_B = X\beta_B$.
4. Determine the p th ERQ hyperplane $y = X\beta$ where $\beta = (1 - \alpha)\beta_A + \alpha\beta_B$ where α is a constant in the interval $[0,1]$.

The algorithm creates a family of the p th ERQ hyperplane by varying α . Choosing $\alpha = 0$ and $\alpha = 1$ yield upper and lower bounds on the family, respectively. Thus, different values on α may result in distinct hyperplanes.

3.4 Choice of α

An important question is how to select the α value. The ERQ uses the following quadratic programming problem so as to uniquely determine the optimal α value:

$$\text{minimize } f(\beta) = \sum_{i \in G_A} p(y_i - X_i\beta)^2 + \sum_{i \in G_B} (1-p)(y_i - X_i\beta)^2 \quad (10)$$

$$\text{subject to } \beta = (1 - \alpha)\beta_A + \alpha\beta_B \quad (10a)$$

$$0 \leq \alpha \leq 1, \quad (10b)$$

which can be transformed to the following problem by inserting (10a) to (10):

$$\begin{aligned} \text{minimize } f(\alpha) &= \sum_{i \in G_A} p[y_i - (1 - \alpha)X_i\beta_A - \alpha X_i\beta_B]^2 \\ &\quad + \sum_{i \in G_B} (1-p)[y_i - (1 - \alpha)X_i\beta_A - \alpha X_i\beta_B]^2 \\ \text{subject to } &0 \leq \alpha \leq 1. \end{aligned} \quad (11)$$

Here, this study has introduced a new criterion for the objective function (10) that is in the form of minimizing the sum of weighted squared deviations (WSD). The advantage of the new criterion is that it can provide $\beta = (1 - \alpha)\beta_A + \alpha\beta_B$ with the property of WSD estimation. As a result, β can maintain the property of estimates that are obtained by minimizing the sum of both WSD and weighted absolute value deviations. Furthermore, the α value can be uniquely selected because the objective of (11) is a convex function.

The derivative of $f(\alpha)$, i.e. $df(\alpha)/d\alpha$ can produce the following optimal condition:

$$\sum_{i \in G_A} p[-\eta_i + \alpha\lambda_i] + \sum_{i \in G_B} (1-p)[- \eta_i + \alpha\lambda_i] = 0, \quad (12)$$

where $\eta_i = (y_i - X_i\beta_A)(X_i\beta_B - X_i\beta_A)$ and

$$\lambda_i = (X_i\beta_A - X_i\beta_B)^2.$$

Furthermore, the α value, which is the solution to (12), needs to satisfy the condition $0 \leq \alpha \leq 1$. Hence, the optimal α^* value for minimizing (11) is determined by the following way:

(a) if $0 \leq \alpha \leq 1$, then the optimal α^* becomes

$$\alpha^* = \alpha = \frac{\sum_{i \in G_A} p\eta_i + \sum_{i \in G_B} (1-p)\eta_i}{\sum_{i \in G_A} p\lambda_i + \sum_{i \in G_B} (1-p)\lambda_i}, \quad (13)$$

- (b) if $\alpha > 1$, then optimal $\alpha^* = 1$, and
- (c) if $\alpha < 0$, then optimal $\alpha^* = 0$.

The selection of α^* stops the algorithm for measuring the p th ERQ.

3.5 Degeneracy in Empirical Regression Quantile

As described before, the degeneracy problem may often occur when many data points are located on the p th regression hyperplane. The algorithmic scheme of ERQ may potentially deal with the problem when it happens in (3). That is, sample observations on the regression hyperplane are classified into two distinct subgroups. Therefore, the occurrence of degeneracy is reduced with the dichotomized sample observations.

Of course, the approach also needs to assume the absence of degeneracy for two regression hyperplanes for G_A and G_B . The multiple solutions occur on the bottom hyperplane of G_A when at least one optimal δ_i^+ , $i \in G_A$ becomes a basic variable and equals zero in (8). Similarly, the degeneracy can occur on the upper hyperplane of G_B when at least one optimal δ_i^- , $i \in G_B$ becomes a basic variable and equals zero in (9). The absence of degeneracy in (8) and (9) is required to uniquely determine an ERQ hyperplane.

4. Computational Results

4.1 Illustrative Example

This study applies first the algorithm of ERQ to an illustrative data set so as to confirm the difference between original regression quantile and ERQ. The resulting ERQ hyperplane describes empirically the level of improvement in terms of (1). The illustrative data set used in this study contains ten sample observations $[(x, y) : (1, 5.0)(2, 3.0)(3, 5.5)(4, 2.0)(5, 8.0)(6, 4.5)(7, 7.0)(8, 3.5)(9, 9.0)(10, 5.0)]$. The linear regression model, i.e., $y = \beta_0 + \beta_1 x$ is fitted to the data set. The two coefficients are measured by both original regression quantile and ERQ with nine different quantile values. The comparison between the two approaches is summarized in Table 1. In Table 1, the fifth column indicates the ERQ lines with nine different p values while the sixth column denotes the corresponding original regression quantile lines. All the regression lines presented at the second and third columns are the bottom line of G_A measured by (8) and the upper line of G_B measured by (9), respectively. The optimal α scores at the fourth column are derived from (13).

Table 1: Comparison between Original and Empirical-Regression Quantile Lines

p	Bottom Line of G_A	Upper Line of G_B	Optimal α^*	Empirical- Regression Quantile	Original Regression Quantile
0.1	$2.833 + 0.083x$	$0.500 + 0.375x$	0.040861	$2.738 + 0.095x$	$0.500 + 0.375x$
0.2	$2.500 + 0.250x$	$0.500 + 0.375x$	0	$2.500 + 0.250x$	$2.833 + 0.083x$
0.3	$5.100 - 0.100x$	$2.833 + 0.083x$	0.287327	$4.449 - 0.047x$	$2.500 + 0.250x$
0.4	5.000	$3.750 + 0.125x$	0.695652	$4.130 + 0.087x$	$3.750 + 0.150x$
0.5	$4.375 + 0.375x$	$3.750 + 0.125x$	0.576817	$4.014 + 0.231x$	5.000
0.6	$4.667 + 0.333x$	$4.750 + 0.250x$	1	$4.750 + 0.250x$	$4.750 + 0.250x$
0.7	$4.500 + 0.500x$	$4.375 + 0.375x$	1	$4.375 + 0.375x$	$4.670 + 0.330x$
0.8	$6.750 + 0.250x$	$4.667 + 0.333x$	1	$4.667 + 0.333x$	$4.500 + 0.500x$
0.9	$6.750 + 0.250x$	$4.500 + 0.500x$	1	$4.500 + 0.500x$	$6.750 + 0.250x$

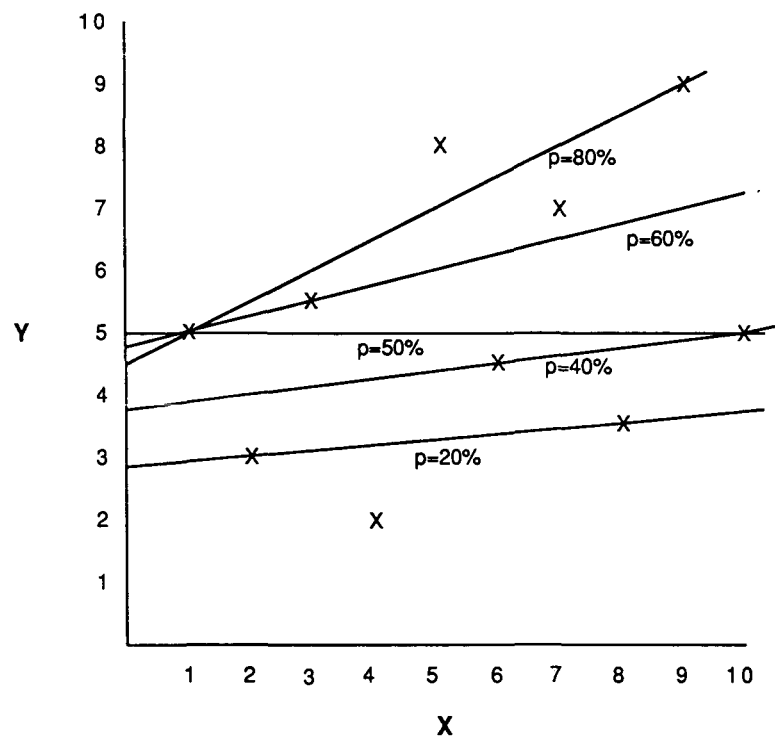


FIGURE 1: Regression Quantile Lines

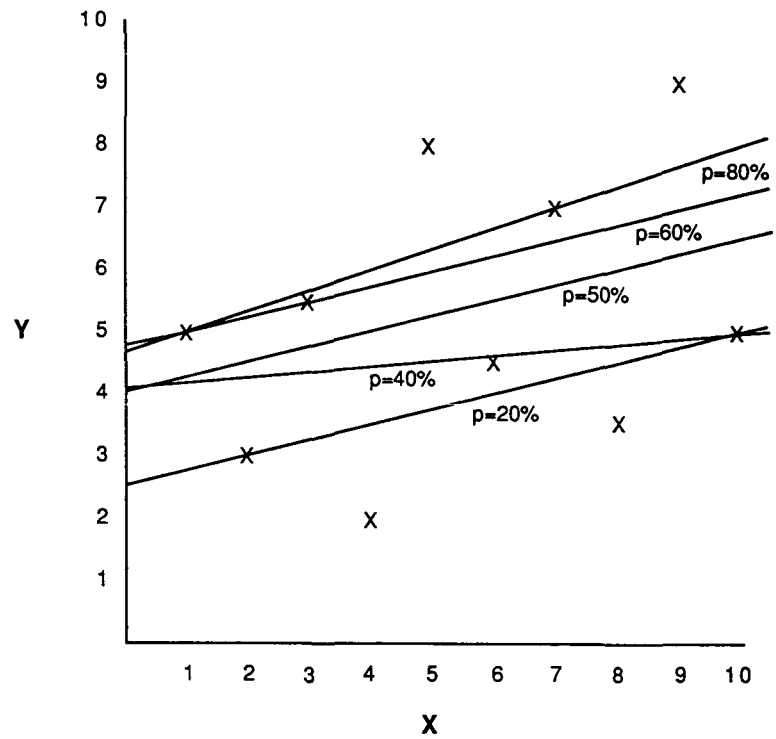


FIGURE 2: GP - Regression Quantile Lines

As presented in Table 1, the two regression quantile methods yield distinct regression lines at all the p values except $p = 60\%$. When $p = 60\%$, both the original regression quantile line and the upper line of G_B become $y = 4.750 + 0.25x$. Furthermore, the optimal α^* value is determined as 1 by (13). Thus, the ERQ line equals the original regression quantile line.

Figure 1 depicts five original regression quantile lines with $p = 20, 40, 50, 60$ and 80% in Table 1. Similarly, Figure 2 presents the corresponding five regression lines using the ERQ approach. Two data points are always required to be on each regression line in Figure 1 while the ERQ line does not require the estimation property as presented in Figure 2. See, for instance, the line with $p = 50\%$. The ERQ approach can measure the p th regression quantile hyperplane more flexibly than (3) in the sense that it can select an optimal solution on not only an extreme point but also other points in a sample space.

4.2 Simulation Study

A Monte Carlo simulation study was conducted to assess the behavior of regression quantile estimates measured by the ERQ approach. The quantile $p = 50\%$ (i.e., regression median) is selected because it can yield the most important regression hyperplane among the family of regression quantiles. LAV estimates were employed as a means of comparing the performance of ERQ estimates so as to investigate how much the latter can improve the statistical efficiency of the former. [It is well known that LAV estimates outperform LS estimates to any error distributions except normal. Hence, LS estimates was not compared in the simulation study.] This study conducted a $6 \times 4 \times 5$ factorial experiment in which each treatment had 10 replications. Factors used in the simulation study were

- (a) Sample size (n): 10, 20, 30, 40, 50 and 100
- (b) Error distribution (ϵ)
 - i) Normal: $N(0, \sigma^2)$ where $\sigma = 1, 1.5, 2, 2.5$ and 3
 - ii) Uniform: $U(-d, d)$ where $d = 0.5, 1, 1.5, 2$ and 3
 - iii) Contaminated Normal: $95\%N(0, 1) + 5\%N(0, \sigma^2)$
where $\sigma = 1, 5, 10, 15, 20$

$$\text{iv) Laplace : } F(x) = \begin{cases} \exp(kx)/2 & (x < 0), \\ 1/2 & (x = 0), \\ 1 - \exp(-kx)/2 & (x > 0), \end{cases}$$

where $k = 0.1, 0.5, 0.8, 1$ and 2.

- (c) Regression model: $y = 5 + 2x_1 + 3x_2 + \epsilon$, x_1 and $x_2 \sim U[1, 50]$.

$F(x)$ is a distribution function of a random variable X . $N(\eta, \sigma^2)$ denotes a normal distribution with a mean (η) and a standard deviation (σ). $U(a, b)$ indicates a uniform distribution on the interval $[a, b]$. The constants, d and k , determine the shape of error distributions.

It is important to note that [16] provides a similar simulation result of a simple regression model, i.e., $y = \beta_0 + \beta_1 x$, while this study explores the simulation with a multiple regression model, i.e., $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2$.

All the error distributions were generated by the inversion of these distribution functions in a personal computer, IBM-AT. For each error distribution with five different shapes and for each sample size, the estimates of β_0, β_1 and β_2 were measured by LAV estimation and ERQ with $p = 50\%$. This study has developed two computer codes for the two estimation methods using the revised simplex method.

Table 2: Ratio of Mean Squares Errors (ERQ/LAV Estimation)

Sample Size Distribution		10			20			30			40			50			100			Number (≥1)	Number (<1)	
		β_0	β_1	β_2	β_0	β_1	β_2	β_0	β_1	β_2	β_0	β_1	β_2	β_0	β_1	β_2	β_0	β_1	β_2			
Normal	$\alpha =$	0.5	.826	1.244	.586	1.666	1.500	.982	1.642	.928	.800	.998	.500	1.000	.960	.950	.800	.666	.500	2.000	32	58
		1.0	1.575	1.300	.964	.667	.667	.611	.954	4.222	.627	1.785	1.187	1.555	1.469	1.272	1.194	.998	.900	1.000		
		1.5	.927	.612	1.117	.847	.934	.891	.847	.979	1.227	1.031	.891	1.153	.971	.894	1.521	.840	.846	.884		
		2.0	.955	.841	.506	.691	.547	1.333	1.313	1.112	1.525	.694	.961	.571	.981	.991	1.121	1.050	.750	1.382		
		2.5	.908	.912	.945	.778	.810	1.124	.767	.840	.676	1.568	1.205	1.714	.986	1.024	.786	.853	.894	.951		
Uniform	$\alpha =$	0.5	.736	1.000	.648	.500	1.000	.555	.833	1.166	.800	.750	.800	.998	.500	.833	1.000	.666	1.021	.988	31	59
		1.0	.806	.733	.569	1.684	.619	1.888	.739	1.058	.500	.998	.999	1.222	.866	.769	.952	1.000	.500	.833		
		1.5	1.111	.597	.987	.302	.450	.490	.956	2.600	.906	.884	.681	.997	.705	1.238	.653	.925	.846	1.125		
		2.0	.619	.469	.830	1.147	1.545	1.073	.772	1.142	.666	1.067	1.473	1.080	1.000	1.046	1.296	1.333	.764	2.000		
		2.5	.689	.758	.564	.983	.941	.762	.966	1.195	.858	1.085	.611	1.297	.895	.590	1.529	.833	1.150	.906		
Contaminated Normal	$\alpha =$	1.0	.448	1.148	.709	.592	.681	1.000	.981	.850	.675	.979	1.111	1.080	1.090	.800	.571	.750	.500	.833	43	47
		5.0	.711	1.041	2.090	1.142	.614	1.090	.947	1.161	.734	.993	.795	1.500	.673	1.357	.550	.857	1.000	.800		
		10.0	2.191	.406	6.627	.962	.711	1.815	.588	.846	1.291	.945	1.562	.576	1.285	1.900	1.086	1.181	1.444	.785		
		15.0	3.203	1.871	1.406	1.372	.615	1.700	.879	1.769	1.000	1.212	1.107	.956	.588	.548	.875	1.038	1.357	.909		
		20.0	2.375	2.745	.500	.856	.923	.890	1.613	1.206	1.046	.847	.802	1.312	.794	2.181	1.184	1.046	.978	.375		
Laplace	$\lambda =$	0.1	1.394	1.388	1.356	1.237	.999	1.140	.968	2.040	1.158	.994	.954	1.112	1.119	.765	1.110	.922	.945	.841	48	42
		0.5	.611	.554	.729	.814	1.263	.796	1.067	.967	1.216	1.242	1.396	1.148	1.294	.708	1.204	.939	.833	1.250		
		0.8	.893	.574	1.126	1.129	1.677	.552	1.351	1.391	1.051	.940	.809	.966	1.307	.857	1.166	1.026	.750	1.000		
		1.0	1.207	.864	1.345	1.128	.909	.666	1.114	.960	.809	1.045	1.666	2.000	1.020	.866	.909	1.027	.999	.800		
		2.0	1.666	2.000	.924	1.342	1.000	.600	1.166	.998	.777	1.208	.753	1.250	1.140	.500	.999	1.000	1.010	.981		
Number (≥ 1)		24			24			27			31			27			22			154	206	
Number (< 1)		36			36			33			29			33			38					

Empirical Regression Quantile

Table 2 presents the ratio of mean squared errors $[MSE(ERQ)/MSE(LAV)]$ for β_0, β_1 and β_2 where MSE of each parameter is computed by $MSE = \sum_{k=1}^{10} (\hat{\beta}_{jk} - \beta_j)^2 / 10, j = 0, 1, \text{ and } 2$.

Here, the term k indicates the number of the replications and $\hat{\beta}_{jk}$ is the estimate of β_j ($\beta_0 = 5, \beta_1 = 2$ and $\beta_2 = 3$) for the k th replicate. The study generated 10 repetitions for the specific combination among its error, shape and sample size. The notation of $\sum_k (\hat{\beta}_{jk} - \beta_j)^2$ indicates the sum of squared deviations between the j th true parameter (β_j) and its estimate ($\hat{\beta}_j$). Thus, the ratio of MSE indicates MSE of the ERQ estimates divided by that of the LAV estimates. [See [19] for the rationality on the MSE.]

In Table 2 the ratio of MSE is presented with a decimal number. If the ratio is less than 1, ERQ estimate is more efficient than LAV estimate; conversely, if the ratio is larger than 1, the opposite is indicated. The last two columns of the table, when aggregated by the distribution, show the frequency count of MSE ratios that were ≥ 1 and < 1 for all simulation runs, respectively.

The results indicate that ERQ outperforms LAV estimation for normal, uniform and contaminated normal distributions. For the Laplace distribution, ERQ could not improve the performance of the LAV regression, indicating 46% (42/90) improvement and 54% (48/90) inferiority. This may be explained by the fact that the LAV estimator becomes the maximum likelihood estimate under the Laplace distribution.

The last two rows at the bottom, when aggregated by the sample size, report the frequency count of MSE ratios that were ≥ 1 and < 1 , respectively. ERQ could improve the performance of LAV estimation at all sample sizes (except 40) examined in the simulation study. One might expect that the frequency count for MSE ratios less than one decreases as the sample size increases, because LAV estimation can asymptotically provide regression quantile estimates. However, the expectation was not confirmed by the simulation study.

Finally, this study needs to describe a shortcoming related to the simulation study exhibited in Table 2. That is, the simulation study measures the performance of ERQ using only three parameter coefficients (β_0, β_1 and β_2) of a multiple regression model. The simulation evidence contradicting the result in Table 2 may occur when the number of parameters of the regression model is increased. Hence, we cannot conclude immediately that ERQ is always much better than LAV, depending upon only the simulation result in Table 2. It is needed to conduct a more intensive simulation study incorporating different perspectives related to different error distributions and different structures of the regression model. This is a future task.

5. Conclusion and Future Extensions

This article has presented an empirical use of regression quantile that can produce a linear regression hyperplane exactly on the $(100p)$ th quantile of the y distribution. The ERQ is designed to utilize information regarding dual variables, that are produced by original regression quantile of Koenker and Bassett [15], so as to classify a data set into two groups of sample observations. The classified data sets determine the upper and lower bounds of the p th regression quantile hyperplane. The criterion of minimizing the sum of weighted squared deviations uniquely selects the best regression quantile hyperplane. As an important property, the resulting estimates of the ERQ were found to be less sensitive to a small sample size than the original form. The evidence was confirmed by an illustrative data set and a simulation study in which this study selected $p = 50\%$ and compared the performance of ERQ with LAV estimation.

Besides the property of robustness, the ERQ quantile can become a practical tool to deal

with managerial issues such as EEO (equal employment opportunity). For instance, consider an EEO evaluation for hospital management in which the amount of salary is measured by each employee's rank, gender, race, education background and experience, using a linear regression model. Usually, medical doctors receive much higher salaries than nurses and administrative employees. The wages of doctors often become outliers. A common practice of LS regression analysis is identification of outliers to be deleted from estimation, using an assumption of a normal error distribution. Deletion of outliers (i.e., medical doctors in the case) may be statistically important but often produce managerially insignificant conclusions. [For instance, Atkinson [3] has proposed a way of deleting outliers based on LS regression median.] On the other hand, the ERQ can be a practical method to deal with various evaluation problems of employees in situations where one needs to avoid pronounced outliers in order to obtain consensual evaluations. [See Sueyoshi [17] that describes the use of ERQ for examining possible salary discrimination against a protected (female or minority) group. His study transforms the ERQ result into a 2×2 contingency table and thereby presents an exact way of computing type I and type II errors concerning discrimination. Furthermore, Sueyoshi [17] proposes a new measure of the level of discrimination based upon the ERQ result, which is expected to be easily understood by lawyers, judges and other individuals involved in the EEO issue. See [18], as well.]

This study needs to explore two research topics so as to provide ERQ with more practicability. First, as described by Charnes *et al.* [8, 9, 10], the ERQ can easily incorporate prior information concerning estimation results in various forms of side constraints. The ERQ can also maintain the methodological capability to include *a priori* requirements in its estimation process. Second, the ERQ needs a computational framework for testing the null hypothesis regarding parameter estimates. As an analytical method, this study proposes the use of ERQ in combination with the statistical test proposed by Hogg [13]. Since it is a straightforward matter to extend the ERQ result into Hogg's method, this study omits a detailed discussion on how to conduct a statistical test using the ERQ result. [See Hogg [13] for his description on the statistical test and regression quantile.] Furthermore, as a promising approach, this study proposes the use of the bootstrap method for testing the null hypothesis. As presented in [12], the computer intensive simulation study (i.e., the bootstrap method) can numerically measure the variance-covariance matrix of parameter estimates. The bridge between the bootstrap method and the ERQ will open up a new statistical testing method.

Finally, we hope that the research concerning the ERQ becomes increasingly important and anxiously wait for further research developments along the lines indicated in this study.

References

- [1] T. Amemiya: *Advanced Econometrics*, Harvard University Press, Cambridge, Massachusetts, 1985.
- [2] G. Appa and C. Smith: On L_1 and Chebychev estimation. *Mathematical Programming*, Vol. 5 (1973), 73-87.
- [3] A.C. Atkinson: Masking unmasked, *Biometrika*, Vol. 73 (1986), 533-541.
- [4] G. Bassett and R. Koenker: Asymptotic theory of least absolute error regression, *Journal of the American Statistical Association*, Vol. 73 (1978), 618-622.
- [5] G. Bassett and R. Koenker: An empirical quantile function for linear models with IID errors, *Journal of the American Statistical Association*, Vol. 77 (1982), 407-415.
- [6] A. Charnes and W.W. Cooper: *Management Models and Industrial Applications of Linear Programming*, John Wiley, New York, 1961.
- [7] A. Charnes and W.W. Cooper: Goal programming and multiple objective optimization: Part I, *European Journal of Operational Research*, Vol. 1 (1977), 39-54.

- [8] A. Charnes and W.W. Cooper: Goal programming and constrained regression: A comment, *Omega*, Vol. 3 (1975), 403-409.
- [9] A. Charnes, W.W. Cooper and R.O. Ferguson: Optimal estimation of executive compensation by linear programming, *Management Science*, Vol. 1 (1955), 138-151.
- [10] A. Charnes, W.W. Cooper and T. Sueyoshi: Least Squares/ridge regression and goal programming/constrained regression alternatives, *European Journal of Operational Research*, Vol. 27 (1986), 147-157.
- [11] A. Charnes, W.W. Cooper and T. Sueyoshi: A goal programming/constrained regression review of the Bell System breakup, *Management Science*, Vol. 34 (1988), 1-26.
- [12] T.E. Dielman and R.C. Pfaffenberger: Bootstrapping in least absolute value regression: An application to hypothesis testing, *Communications in Statistics - Simulation and Computation*, Vol. 17 (3) (1988), 843-856.
- [13] R.V. Hogg: Estimates of percentile regression lines using salary data, *Journal of the American Statistical Association*, Vol. 70 (1975), 56-59.
- [14] P.J. Huber: *Robust Statistics*, John Wiley & Sons, New York, 1981.
- [15] R. Koenker and G. Bassett: Regression quantiles, *Econometrica*, Vol. 46 (1978), 33-50.
- [16] T. Sueyoshi and Y.L. Chang: Goal programming approach for regression median, *Decision Sciences*, Vol. 20 (1989), 700-714.
- [17] T. Sueyoshi: Empirical regression quantile for measuring employment discrimination, Memo, 1989.
- [18] T. Sueyoshi: Estimation of stochastic frontier cost function using data envelopment analysis: an application to the AT&T divestiture, *Journal of the Operational Research Society*, Vol. 42 (1991), 463-477.
- [19] C. Toro-Vizcarrondo and T.D. Wallace: A test of the mean square error criterion for restrictions in linear regression, *Journal of the American Statistical Association*, Vol. 63 (1968), 558-572.

Toshiyuki Sueyoshi
The Ohio State University
College of Business
School of Public Policy and Management
1775 College Road
Columbus, Ohio 43210, U.S.A.