Journal of the Operations Research
Society of Japan
Vol. 38, No. 1, March 1995

EMPIRICAL REGRESSION QUANTILE AND LEVERAGE TREATMENT METHOD

Toshiyuki Sueyoshi
The Ohio State University

(Received December 28, 1992)

Abstract This article presents analytical and numerical approaches for statistically testing parameter estimates of a regression hyperplane, using Empirical Regression Quantile (ERQ) technique. The analytical approach uses its asymptotic property of ERQ to determine the standard error of parameter estimates. The asymptotical results are replaced and compared with a computer intensive technique referred to as "a bootstrap method." Furthermore, Leverage Treatment Method (LTM) is proposed for dealing with a leverage point problem that may seriously affect ERQ results. It is not trivial to detect the leverage point in a multivariate date set. An important feature of the LTM is that it fully utilizes dual variables derived from ERQ for identifying the leverage point. The proposed ERQ/LTM is applied to two illustrative examples in which the technique is compared with other conventional methods.

1. Introduction

The central research issue being proposed in this article is the development of analytical and numerical procedures for testing parameter estimates of Empirical Regression Quantile (ERQ), simultaneously dealing with a leverage point that could potentially affect ERQ results. This article proposes a Leverage Treatment Method (LTM) to handle the leverage point problem which often leads to ERQ misinterpretation.

The problem of this leverage point has been discussed in several articles such as Atkinson (1986), Belsley *et al.* (1980), Hawkins *et al.* (1984), Rousseeuw and Zomeren (1990), and Fung (1993). An important feature of these research efforts is that their approaches for dealing with a leverage point are all explored in the framework of conventional Least Squares (LS) regression. Meanwhile, this study will discuss the same issue from the perspective of ERQ, a special form of Goal Programming (GP) technique.

The ERQ, proposed first by Sueyoshi (1991b), is an analytical method to produce a linear regression hyperplane on the $100p^{th}$ percentile of an error distribution, where "p" indicates a predetermined percentile. In order to describe ERQ more clearly, this study starts with fitting a linear regression hyperplane, mathematically defined as $y_j = X_j\beta+\varepsilon_j (j = 1,\dots,n)$; where y_j is the j^{th} observed dependent variable, $\beta = (\beta_1, \beta_2,\dots,\beta_m)^T$ is a column vector representing parameter coefficients to be measured, $X_j = (x_{1j}, x_{2j},\dots,x_{mj})^T$ is the j^{th} row vector of an observed design matrix, and ε_j is an error related to the j^{th} observation. Formally, the ERQ can be considered as an algorithm that is designed to yield parameter estimates $\hat{\beta} = (\hat{\beta}_1, \hat{\beta}_2,\cdots,\hat{\beta}_m)$, satisfying the following required conditon:

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

where $N(\hat{\beta})$ indicates the number of observations such $y_j < X_j\hat{\beta}$. A methodological advantage of (1) is that this type of regression is usually very robust to the existence of an outlier(s) and/or non-normal error distributions (Hogg, 1977 and Huber, 1981). An initial effort to obtain (1), using GP, was first accomplished by Charnes *et al.* (1955). Their research work was further extended into the asymptotical theory of Bassett and Koenker (1978). As an extension of these studies, Sueyoshi (1991b) has opened up a new approach that can always maintain the required property of (1) to any data sets, including a data set with a small sample size. The method was referred to as "ERQ" by Sueyoshi (1991b). A research issue needed to be explored for ERQ is the development of various statistical tests for evaluating its parameter estimates. Moreover, ERQ needs an analytical approach to detect a leverage point(s) that may cause serious distortion of ERQ, leading to misinterpretation of ERQ results in many cases. This article attempts to overcome such methodological issues, simultaneously.

The structure of this article is organized as follows. The next section presents ERQ. Section 3 presents a statistical test for ERQ based upon its asymptotical property. The statistical test is reexamined by a bootstrap method that measures the standard error of a parameter estimate by resampling artificially generated data sets many times in the form of a computer intensive simulation. This task is also accomplished in Section 3. Section 4 presents LTM that is designed for detecting a leverage point in the framework of ERQ. The ERQ/LTM is applied to two illustrative data sets in Section 5. In this section, ERQ/LTM is compared with other traditional methods. Conclusion and future extensions are summarized in the last (6) section.

2. Empirical Regression Quantile

The ERQ can be broken down into three subprocesses. The initial stage of ERQ begins with a conventional regression quantile model that is formulated by the following GP model:

$$\begin{array}{ll} \text{minimize} & \sum_{j=1}^{n}[p\delta_j^+ + (1-p)\delta_j^-] \\ \text{subject to} & X_j\beta + \delta_j^+ - \delta_j^- = y_j, \quad j = 1,\cdots,n, \\ & \delta_j^+ \geq 0 \text{ and } \delta_j^- \geq 0, \quad j = 1,\cdots,n, \end{array} \qquad (2)$$

where δ_j^+ and δ_j^- are positive and negative deviations related to the jth observation, respectively. Model (2), originally proposed by Charnes *et al.* (1955) and further extended by Bassett and Koenker (1978), may yield a regression hyperplane that satisfies (1) on an approximate basis. When a data set has a small (e.g., less than 100) sample size, a resulting regression hyperplane measured by (2) does not exactly satisfy (1). Using the result of (2) as initial information, ERQ attempts to relocate the resulting regression hyperplane in a $X - y$ sample space, so that it satisfies exactly (1), simultaneously attempting to maintain the minimized sum of absolute errors, as formulated in the objective of (2).

In an effort to overcome this difficulty of (2) in terms of exactly maintaining (1), ERQ uses first dual variables measured by (2). The dual variables of ERQ can serve as a basis for classifying an observed data set into two subsets, as required in (1). Here, in order to describe the dichotomization, let w_j be the dual variable related to the j^{th} observation determined by (2). As found by Sueyoshi (1991b), the dual variable indicates the rate of change in the objective of (2) due to one unit of increase in y_j. Hence, the examination of the dual variable provides information on the locational relationship between y_j and its estimate. [See also Sueyoshi and Chang (1989a) for a detailed description on the implication of w_j when $p = 50\%$.]

Hereafter, this article uses a new symbol "k"$(k = 1, \ldots, n)$, representing the descending order of $w_j (j = 1, \ldots, n)$, so as to describe how w_j is fully utilized to dichotomize an observed data set. The descending order may be expressed by

$$w_1 \geq w_2 \geq \ldots \geq w_k \geq \ldots \geq w_n. \tag{3}$$

All the observations (J) are dichotomized into the following two subsets ($J = G_A U G_B$):

$G_A = \{j \mid \text{the} j^{th} \text{observation has } w_j \text{ belonging to the top} (1-p)^{th} \text{percentile in (3)}\}$and
$G_B = \{j \mid \text{the} j^{th} \text{observation has } w_j \text{ belonging to the bottom } p^{th} \text{percentile in (3)}\}$.

The sizes of G_A and G_B are $n(1-p)$ and np, respectively. It is important to note that in some case where ERQ cannot clearly separate an observed data set into its two subsets, ERQ needs to round off fractions to an integral value for this dichotomization. For instance, $n(1-p) = 4.8$ and $np = 3.2$ can be observed from the combination between $p = 0.4$ and $n = 8$. The round-off method determines that G_A and G_B have 5 and 3 observations, respectively.

The second stage of ERQ applies the following GP models to G_A and G_B, respectively:

$$\begin{aligned} \text{minimize} \quad & \sum_{j \in G_A} \delta_j^+ + L \sum_{j \in G_B} \delta_j^+ \\ \text{subject to} \quad & X_j\beta + \delta_j^+ = y_j, \ j \in G_A, \\ & X_j\beta + \delta_j^+ - \delta_j^- = y_j, \ j \in G_B, \\ & \delta_j^+ \geq 0 \text{ and } \delta_j^- \geq 0, \quad j = 1, \cdots, n, \end{aligned} \tag{4}$$

and

$$\begin{aligned} \text{minimize} \quad & L \sum_{j \in G_A} \delta_j^- + \sum_{j \in G_B} \delta_j^- \\ \text{subject to} \quad & X_j\beta + \delta_j^+ - \delta_j^- = y_j, j \in G_A, \\ & X_j\beta + \quad -\delta_j^- = y_j, j \in G_B, \\ & \delta_j^+ \geq 0 \text{ and } \delta_j^- \geq 0, \quad j = 1, \cdots, n, \end{aligned} \tag{5}$$

where all the symbols used in (4) and (5) are the same as used in (2), except L represecting a non-Archimedean large number. Two GP models (4) and (5) yield two distinct regression hyperplanes, each of which can be characterized as follows: First, as presented in (4), only positive deviations in G_A are minimized in its objective so that (4) yields a bottom hyperplane to G_A. Here, "bottom" means that all the sample observations in G_A locate above or on the resulting hyperplane. Meanwhile, (4) prevents any observation in G_B from being above the bottom hyperplane of G_A, because positive deviations in G_B are weighted by L in (4). Conversely, (5) produces an upper hyperplane to G_B, where "upper" means that all the sample observations in G_B are below or on the regression hypereplane. The function of (5) can be easily understood by changing the description regarding (4) in a converse manner.

The third and last stage of ERQ needs to determine the p^{th} hyperplane that can satisfy (1) between two regression hyperplanes derived from (4) and (5). In order to determine the p^{th} ERQ hyperplane, let $X\hat{\beta}_A$ and $X\hat{\beta}_B$ be the bottom hyperplane of G_A and the upper hyperplane of G_B, respectively. Then, the ERQ hyperplane $X\hat{\beta}$ can be determined by

$$\hat{\beta} = (1-\alpha)\hat{\beta}_A + \alpha\hat{\beta}_B \tag{6}$$

where α is a constant value on [0,1]. The upper and lower bounds of $\hat{\beta}$ may be obtained by selecting $\alpha = 0$ or 1, respectively. Thus, different α values produce different ERQ hyperplanes. An optimal α value is determined by the following model:

$$\begin{aligned} &\text{minimize} \quad \sum_{j \in G_A} p(y_j - X_j\beta)^2 + \sum_{j \in G_B} (1-p)(y_j - X_j\beta)^2 \\ &\text{subject to} \quad \hat{\beta} = (1-\alpha)\hat{\beta}_A + \alpha\hat{\beta}_B \text{ and } 0 \leq \alpha \leq 1. \end{aligned} \tag{7}$$

Here, the criterion of minimizing the sum of weighted square errors is incorporated in (7) so as to determine uniquely the optimal α^* value.

Following the proof of Sueyoshi (1991b), ERQ determines the optimal α^* by computing first:

$$\begin{aligned} &\alpha = [\sum_{j \in G_A} p\eta_j + \sum_{j \in G_B} (1-p)\eta_j]/[\sum_{j \in G_A} p\lambda_j + \sum_{j \in G_B} (1-p)\lambda_j] \\ &\text{where } \eta_j = (y_j - X_j\hat{\beta}_A)(X_j\hat{\beta}_B - X_j\hat{\beta}_A) \text{ and } \lambda_j = (X_j\hat{\beta}_A - X_j\hat{\beta}_B)^2. \end{aligned} \tag{8}$$

Then, using (8), the optimal α^* is determined as one of the following three cases: (a) if $0 \leq \alpha^* \leq 1$, then $\alpha^* = \alpha$, (b) if $\alpha > 1$, then $\alpha^* = 1$, and (c) if $\alpha < 0$, then $\alpha^* = 0$. The selection of α^* indicates the end of the ERQ algorithm proposed by Sueyoshi (1991b).

It is important to note that ERQ has other statistical properties, besides its robustness, as a percentile regression. For instance, ERQ can incorporate prior information representing various requirements on its resulting estimates. Moreover, using such additional side constraints, ERQ can deal with the problem of multicollinearity. [See Charnes ***et al.*** (1986, 1988) and Sueyoshi (1994a) which describe the theoretical relationship between the multicollinearity problem and the side constraints to be incorporated in ERQ. See also Charnes and Cooper (1961, 1975).]

3. Statistical Tests For ERQ

3.1 Asymptotic Property

The asymptotic behavior of regression quantile was first studied by Bassett and Koenker (1982). They studied first the asymptotical property of Least Absolute Value (LAV) estimation in their works (Bassett and Koenker, 1978 and Koenker and Bessett, 1978). Then, their study (1982) extended it to the asymptotical behavior of ERQ.

In an effort to describe the asymptotic theory from the perspective of our ERQ, this article starts with LAV estimation which is a special case of ERQ. The LAV estimation is obtained by setting $p = 50\%$ in (2). It is clear that the selection of $p = 50\%$ is the most important case of ERQ in term of its statistical applications. Their theory proved that (2) with $p = 50\%$ produces parameter estimates and these follow asymptotically a normal error distribution with $E(\hat{\beta}) = \beta$ and $Cov(\hat{\beta}) = \lambda^2(X^TX)^{-1}$. Here, $\lambda = 1/[2f(0)]$ and $f(0)$ is the height of the density of errors at zero (i.e., their median). Thus, it proves that LAV estimates have a smaller asymptotic covariance matrix than ordinal LS estimates when $\lambda^2 < \sigma^2$ is measured, where σ^2 is the variance of errors. This indicates that, for instance, if the errors follows the Laplace distribution, then $f(0) = 1/[\sqrt{2}\sigma]$ is observed, so that $\lambda^2 = \sigma^2/2$ is obtained. Hence, LAV has half the asymptotic variance of LS. This asymptotic property can be easily extended into regression quantile.

An empirical difficulty in using the asymptotical property of LAV is that $\lambda = 1/[2f(0)]$ is assumed to be known. This assumption is hardly achievable in real data sets. Many research works, including Dilman and Pfaffenberger (1982, 1988), have investigated an estimation method for λ. This article will use the result of Dilman and Pfaffenberger (1982, p.40) so as to estimate λ by the following equation:

$$\hat{\lambda} = 1/[2\hat{f}(0)] = [e(g) - e(-g)]/(4g/n), \tag{9}$$

where $e(g)$ and $e(-g)$ are the g^{th} and $-g^{th}$ errors on the symmetric index of the median sample.

A drawback related to (9) is that the median point cannot be uniquely determined by (2), because it is solved by linear programming (LP) and its solution is always determined on an extreme point(s). In the structure of (2), the extreme point is determined by the combination of sample observations. As a result of such LP property, (2) always needs several data points on its regression hyperplane. For instance, when a linear regression model with two parameters is fitted to a data set, at least two sample observations are usually required to be on the resulting hyperplane. Consequently, it is very difficult to determine the median point and its derived symmetric index. As discussed by Dielman and Pfaffenbergar (1988, p.846), the problem for identifying the median sample makes the determination of $\hat{\lambda}$ more difficult in many real applications. Meanwhile, the ERQ approach proposed in this study fully utilizes the information of dual variables, as presented in the previous section, so that it can easily determine the median point and thereby, the symmetric index required for (9). This clearly indicates a methodological advantage of ERQ in the estimation of $\hat{\lambda}$. [The problem of a leverage point, that will be discussed later in this article, is also due to the property of (2); the model always needs sample data points on its regression hyperplane. Such data points strongly pull down the location of a resulting regression hyperplane. For instance, if a data point on the regression hyperplane exists far away from the majority of its belonging data set in a $X - y$ space, then it often becomes the leverage point.]

The asymptotical theory on the LAV estimation is easily applicable to regression quantile, as well. That is, (2) yields parameter estimates $\hat{\beta}$ whose asymptotic distribution is multivariate normal with mean vector $E(\hat{\beta}) = \beta$ and covariance matrix is

$$\mathrm{Cov}(\hat{\beta}) = \frac{(X^T X)^{-1} p(1-p)}{f^2(\zeta_p)}, \tag{10}$$

where ζ_p is the p^{th} quantile of the error distribution.

As mentioned previously, Bassett and Koenker (1982, P.410) extended the asymptotic theory of regression quantile, proving that the asymptotical behavior of ERQ estimates converges to that of regression quantile estimates. Therefore, ERQ estimates behave like those measured by (2) in a very large sample size. [Of course, it can be very easily found that a large difference may occur between the properties of ERQ and regression quantile in a small sample.] As a result of their proof (1982), statistical tests for ERQ estimates can be conducted in the asymptotical properties of (9) and (10).

3.2 Bootstrap Method

As explored above, the asymptotical property of ERQ can serve as a theoretical basis for its statistical tests. However, this study needs to admit that in some case, where errors do not follow the normal distribution, we need to examine empirically how much they are different from the asymptotic distribution. For example, a data set including a student's GPA (Grade Point Average) maintains a skewed error distribution because GPA is measured between 0 and 4. Another example can be easily found in measuring parameter estimates of a cost function. The cost function needs to satisfy various economic conditions (e.g., linear homogeneity and negative own-price elasticity), so that it can be connected to a production function, as required in Shephard's duality theory (lemma) of production economics

(Shephard, 1970). [The duality theory is different from the duality of LP.See Charnes *et al.* (1988) and Sueyoshi (1991a) for a detailed description on regularity conditions required by the duality theory of economics.] As presented in Sueyoshi (1991a), side constraints representing the regulatory conditions often cause an assumed error distribution to be far from the normal. In such many cases, this article suggests that we need to use a computer intensive technique, referred to as "bootstrap method", to empirically measure the standard error of each parameter estimate.

The bootstrap procedure for LAV estimation was proposed first by Dielman and Pfaffenberger (1988, p.847). This study uses it for ERQ. The bootstrap approach for ERQ consists of the following four numerical steps:

(a) First, ERQ is applied to an observed data set. The estimated parameters $\hat{\beta}$ and the residuals $\hat{\epsilon}_j = y_j - \hat{y}_j$ are saved.
(b) Second, the errors are resampled with replacement to obtain new sample errors $\bar{\epsilon}_j$. The errors are used to create a pseudo-data set:

$$\bar{y}_j = X_j \hat{\beta} + \bar{\varepsilon}_j.$$

(c) Third, the parameters are reestimated to obtain new parameter estimates $\bar{\beta}$. Deviations in these parameter estimates can be directly observed by $\bar{\beta} - \hat{\beta}$.
(d) By replicating the above processes, the distribution of the true errors $(\hat{\beta} - \beta)$ can be approximated by the distribution of the pseudo-errors $(\bar{\beta} - \hat{\beta})$.

The variance of the bootstrap distribution can serve as an estimate of the variance of ERQ parameter estimates. In this study, 500 bootstrap replications are used. [Of course, this study is aware of the fact that more bootstrap replications may improve the quality of the variance estimation.]

4. Leverage Point

4.1 Leverage Treatment Method

When applying ERQ (also LAV), we need to pay attention to the existence of so-called "a leverage point(s)", or an outlying point(s) in X, that may cause serious distortion of an ERQ hyperplane, usually leading to misinterpretation of ERQ results.

Here, this study adapts a figure from Rousseeum and Zomeren (1990), as presented in Figure 1, to visually describe the leverage point problem. Using their terminology, all the data points in Figure 1 are classified into the following four types: (a) "regular observations", (b) "vertical outlier", (c) "good leverage point", or (d) "bad leverage point." Both data listed with (c) and (d) are leverage points, because their x values are outlying in its data space. Although (c) exists on the linearly extended line from the majority of regular observation, (d) deviates from the linear pattern. Therefore, (c) and (d) are referred to as a good leverage point and a bad leverage point, respectively. The vertical outlier (b) is not a leverage point. Since ERQ is robust to both the vertical outlier and the good leverage point, this study focuses upon identifying the bad leverage point only.

In an effort to identify the bad leverage point in our perspective, this article redefines it analytically in the framework of ERQ, returning to its capability to find a median point by fully utilizing dual variables of (2). The dual variables can be examined on a single dimension, even if it is obtained from multivariate ERQ regression. The examination makes it possible to identify the median point in m dimensions (i.e., the number of parameter estimates to be measured). This article uses the median point as initial information for identifying the bad leverage point.

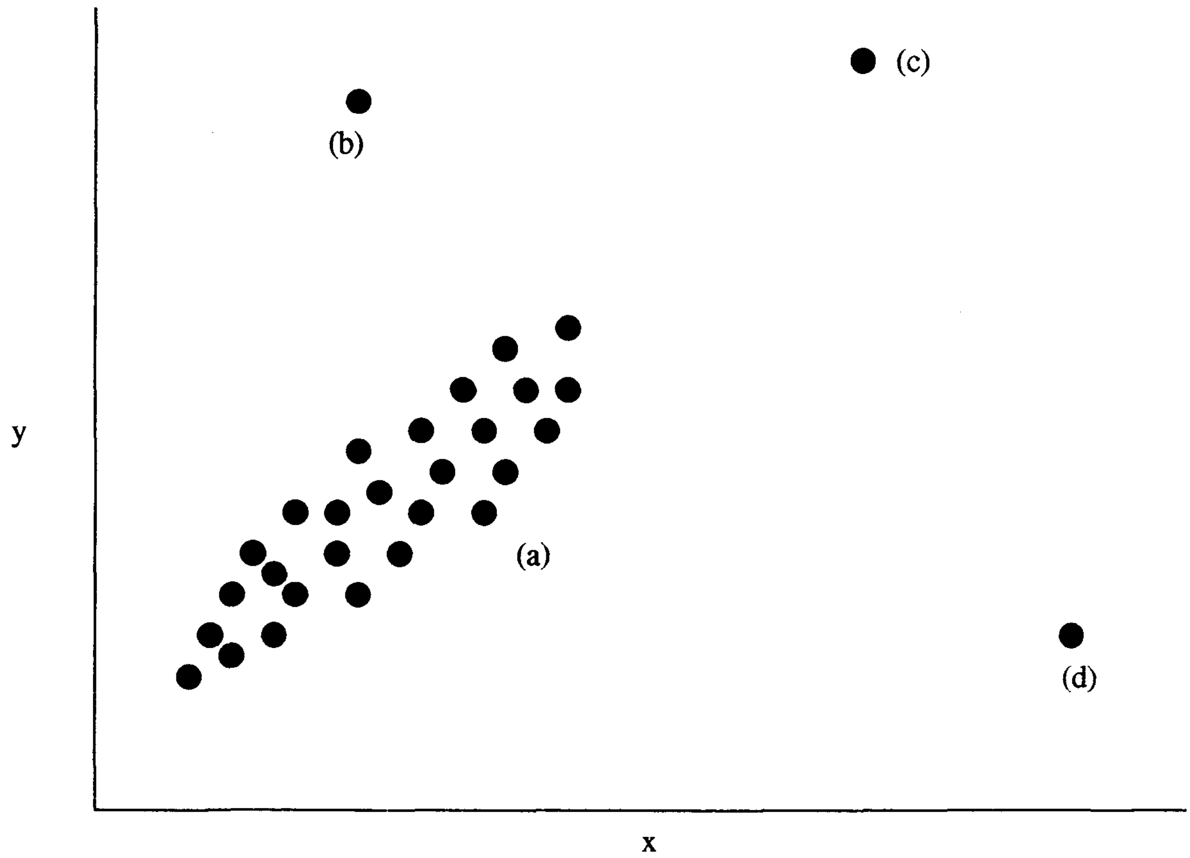

Figure 1 : Simple Example with (a) Regular Observations (b) Vertical Outlier, (c) Good Leverage Point and (d) Bad Leverage Point.

Source : Rousseeuw and Zomeren (1990, P.626).

Next, this study describes how the dual examination is incorporated in LTM for identifying the bad leverage point. The LTM is broken down into the following six subalgorithmic processes:

(a) First, LTM sets $p = 50\%$ in (2). Then, the median point (y_m, X_m) is determined on (3), using the following simple rule:

$$(a-1) \text{ if } n = 2r+1, \text{ then } (y_m, X_m) = (y_r, X_r) \text{ and}$$
$$(a-2) \text{ if } n = 2r, \text{ then } (y_m, X_m) = \{(y_r + y_{r+1})/2, (X_r + X_{r+1})/2\}. \tag{11}$$

This median point can strongly resist to the existence of a vertical outlier. However, it might become a bad leverage point.

(b) Second, in order to avoid such a case where the median point becomes the bad leverage point, LTM identifies all the data points on the regression hyperplane of (2), whose set

(H) can be defined as follows:

$$H = \{j \in J \text{ and } -1 < w_j < 1\}.$$

Then, LTM computes X_s that is the arithmetic sample median of the observed data set (X), i.e.,

$$X_s = \{x_{1s}, x_{2s}, \dots, x_{ms}\}, \tag{12}$$

where each component of X_S is simply a middle order statistic on the order $(x_{i1}, x_{i2}, \dots, x_{in})$ when n is odd. When n is even, we use the average of the order statistic with ranks $(n/2)$ and $(n/2)+1$. More formally, the i^{th} component of $X_s(i = 1, 2, \dots, m)$ is determined by

$$(b-1) \text{ if } n = 2r+1, \text{ then } (x_{is}) = (x_{ir}) \text{ and}$$
$$(b-2) \text{ if } n = 2r, \text{ then } (x_{is}) = \{(x_{ir} + x_{ir+1})/2\}. \tag{13}$$

(c) Third, LTM determines which data point in H is the closest point to X_m and X_s. For this purpose, the examination follows the order of the symmetric dual index from the median point (X_m). Hence, X_m is the initial data point to be examined. Let X_M be the selected data point in H that is the closest to both X_m and X_s. The X_M is a data point that is robust to not only an outlier but also a leverage point.

(d) Fourth, LTM computes a distance measure, named "LTM $- D_j$", that is formulated as follows:

$$\text{LTM} - D_j = \sqrt{(X_j - X_M)S(X)^{-1}(X_j - X_M)^T} \tag{14}$$

for each point $X_j (j = 1, \dots, n)$, where

$$S(X) = [\sum_{j=1}^{n}(X_j - X_M)(X_j - X_M)^T]/(n-1)$$

indicates the sample covariance matrix from X_M. The distance LTM $- D_j$ measures how far X_j is located from X_M.

(e) Fifth, LTM compares the score of LTM $- D_j$ with a cutoff value measured by $\chi^2_{m,1-\alpha/2}$.

(f) Finally, in order to confirm whether a data point with a high LTM-D score is a bad or good leverage point, LTM temporally omits it from an observed data set. If there is a bad leverage point in an observed data set, the data point must be on the ERQ regression hyperplane. [This feature of (2) indicates a methodological disadvantage in determining an ERQ regression hyperplane. However, it becomes an advantage in identifying a bad leverage point.] Then, using the remaining data set, LTM repeats ERQ. If a major change is found in the two resulting ERQ hyperplanes, the omitted data point can be identified as the bad leverage point. Otherwise, it is a good leverage point. This process is replicated until all the leverage points are discovered.

It is important to note that (14) is a modified from of Mahalanobis distance measure (Rousseeuw and Zomeren, 1990, P.633):

$$\text{MD}_j = \sqrt{(X_j - T(X))C(X)^{-1}(X_j - T(X))^T} \tag{15}$$

and

$$C(X) = [\sum_{j=1}^{n}(X_j - T(X))(X_j - T(X))^T]/(n-1)$$

where $T(X)$ is the arithmetic mean of the data set (X) and $C(X)$ is the usual sample covariance matrix. As discussed by Rousseeuw and Zomeren (1990), the Mahalanobis distance suffers from a masking effect, by which multiple outliers do not necessarily have a large MD_j score, because the mean value is easily influenced by the outlier(s). Equation (14) modifies the Mahalanobis distance in a way of replacing $T(X)$ and $C(X)$ by X_M and $S(X)$, respectively. Meanwhile, Rousseeuw and Zomeren (1990) used Minimum Volume Ellipsoid (MVE) estimates for $T(X)$ and $C(X)$ to increase the robustness of the Mahalanobis distance measure. [See Rousseeuw (1984) in which Least Median of Squares (LMS) regression is used to measure MVE estimates. A mathematical description on LMS is presented, with an illustrative example, in the next section. See also Atkinson (1986) for a description on identifying outliers by another LS approach.] A problem of their approach is that the MVE estimation is computationally inefficient. Meanwhile, the examination of dual variables produced by ERQ easily identifies the data point (X_M) whose finding plays an important role in reducing computational effort to determine LTM $- D_j$ scores.

5. Illustrative Examples

5.1 A Small Data Set

5.1.1 LTM Results

An illustrative data set, presented in Table 1, is artificially generated so as to describe how LTM is used for identifying a leverage point. Furthermore, ERQ/LTM results applied to the data set are compared with other conventional regression techniques.

Table 1 : Resulting Dual Variables and LTM - D Scores

Observation (j)	x	y	Dual Variable	LTM - D Score
1	100	16	-0.9315	3.275 **
2	15	15	-1.0000	0.538
3	20	18	-1.0000	0.314
4	22	20	-1.0000	0.224
5	25	19	-1.0000	0.090
6	30	22	-1.0000	0.135
7	27	23	-0.0685 (m)	0
8	29	24	1.0000	0.090
9	33	25	1.0000	0.269
10	34	26	1.0000	0.314
11	35	27	1.0000	0.359
12	36	38	1.0000	0.404
13	40	31	1.0000	0.583

Note : (m) indicates a median point and " ** " indicates that its LTM - D score is larger than the chi-square cutoff point (2.24). S (X) = 496.92.

The first stage of ERQ sets $p = 50\%$ in (2). The ERQ application yields resulting dual variables as listed at the third column of Table 1. The dual examination determines the seventh observation (27,23) as the median point (X_m). This point is also the closest point to the arithmetic sample median (X_s) measured by (12). Therefore, (27,23) becomes (X_M, y_M), as well. Using this data point (27,23) and the illustrative data set of Table 1, LTM computes $S(X) = 496.92$ and all LTM $- D_j(j = 1, 2, .., 13)$ values, using (14). All of these LTM $- D_j$ results are listed at the fourth column of Table 1. In this table, the LTM $- D_1(= 3.275)$ exceeds the cutoff value $\sqrt{\chi_1^2, 0.975} = 2.24$. [This study used a chi-square table listed in Hines and Montgomery (1980,p.594).] Based upon this result, the first observation is determined as an ***outlier.***

Second, the first data point can be considered to be on the resulting regression line, because of its dual value ($i.e., -1 < w_1 = 0.9315 < 1$). This result indicates that the first observation is a ***leverage point and outlier.***

Finally, LTM omits the first data point and applies (2) to the remaining data set ($j = 2, .., 13$). The two ERQ applications yield the following two distinct regression lines, depending upon the inclusion of the first data point:

$$\text{(a) } y = 25.5890 - 0.0959x \text{(with the first observation) and}$$

$$\text{(b) } y = 6.000 + 0.600x \quad \text{(without the first observation).}$$

Thus, we can easily find that the two regression hyperplanes have a major difference in these signs and magnitudes. This clearly indicates that the first observation is a ***bad leverage point and outlier.***

5.1.2 Comparison among Four Regression Methods

As a result of omitting the leverage point (the first data) from the observed data set, ERQ determines its initial regression line as $y = 6 + 0.6x$ and produces dual variables as follows: $w_2 = -0.6, w_3 = 0.6, w_4 = 1.0, w_5 = -1.0, w_6 = -1.0, w_7 = 1.0, w_8 = 1.0, w_9 = 1.0, w_{10} = -1.0, w_{11} = -1.0, w_{12} = 1.0$ and $w_{13} = 1.0$. The dual examination classifies the data set into the following two group:

$$G_A = \{j \mid 3, 4, 7, 8, 12 \text{ and } 13\} \text{ and } G_B = \{j \mid 2, 5, 6, 9, 10 \text{ and } 11\}.$$

The second stage of ERQ, using (4) and (5), yields $y = 6 + 0.6x$ both as the upper regression line for G_B and as the bottom regression line for G_A. [It is just a coincidence that G_A and G_B produce a same regression line.]

Finally, since the two lines are same, this study omits the determination of optimal α. Thus, the estimated ERQ/LTM regression becomes $y = 6 + 0.6x$.

Hereafter, this research compares the resulting ERQ/LTM regression line with other regression lines measured by traditional regression methods so as to visually examine the level of estimation improvement of ERQ/LTM by comparing its result with other methods. This research selects LS, LAV, and LMS (Least Median of Squares) methods as estimation alternatives of ERQ/LTM. This comparison among the four regression methods can be visually summarized in Figure 2. This study does not need to describe LS, because the technique is very well known and presented in any statistical text book. Furthermore, the LAV estimation indicates ERQ with $p = 50\%$ in (2), which has been long considered as an estimation alternative to the LS method since the 18^{th} century. This article omits its description on LAV, as well. However, LMS (Least Median of Squares) regression, first proposed by Rousseeuw (1984), needs a description on its formulation and statistical application. The LMS technique is currently proposed as an alternative estimation method for identifying outliers and leverage points (Rousseeuw and Zomeren, 1990).

Mathematically, the LMS method is expressed by

$$\underset{\beta}{\text{minimize}} \quad \underset{j}{\text{median}}\ \epsilon_j^2 \tag{16}$$

where $\epsilon_j = y_j - \beta_1 x_{1_j} - \beta_2 x_{2_j} -, \cdots, -\beta_m x_{mj}$ indicates the j^{th} residual. The sum of squared errors in LS is replaced by the median of the squared residuals in (16). This replacement adds LMS a capability to produce more robust estimators to an outlier than LS.

This study summarizes important findings in Figure 2 as follows:

(a) First, Figure 2 clearly depicts a fact that both LAV and LS are seriously influenced by the bad leverage point (i.e., the first data).

(b) Second, as discussed previously, the leverage point exists on the LAV regression line. Consequently, the point completely distorts the LAV regression line from the direction of the data set.

(c) Third, the level of LAV distortion is much worse than LS, indicating that LAV is more seriously influenced by a leverage point than LS.

(d) Finally, LMS regression ($y = 3.6+0.7x$) and ERQ/LTM ($y = 6+0.6x$) fit to the direction of the data set. This result obviously indicates the robustness of the two approaches to the bad leverage point.

5.2 A Data Set Regarding OECD Public Telecommunication Systems

5.2.1 Data Set and Regression Model

The next illustrative data set used in this ERQ/LTM application represents the performance of public telecommunications in 24 OECD countries (1987). The data set, originally presented in an OECD policy study (1990,pp.139-157), was adapted from Sueyoshi (1994b). In his study, Sueyoshi (1994b) first developed a new type of stochastic frontier probuction by combining Data Envelopment Analysis (DEA) with LAV. As visually described in Figure 2, LAV estimation is seriously affected by the existence of a bad leverage point. In an effort to extend his original work (DEA/LAV) here, this article presents a new combination among DEA, ERQ, and LTM that can measure parameter estimates of a stochastic frontier production function. This study expects that the combination DEA/ERQ/LTM may provide a new empirical result that cannot be obtained from DEA/LAV.

Since the data set used in this study is presented in Sueyoshi (1994b), this article does not provide its detailed description concerning the data set, here, except noting that each data point indicates OECD country's Public Telecommunications Operation (PTO). The PTO is measured in the form of a production activity that uses the following three independent variables (inputs): (a) the number of telephone main lines installed in PTO [unit = 1000 lines], (b) the total amount of capital investment [unit = 1 million US dollars], and (c) the number of employees working for PTO [unit = 1000 employees], as well as the following dependent variable (output): (d) the total amount of telecommunications service revenues [unit = 1 million US dollars].

A regression model applied to the data set is the Cobb-Douglas production function:

$$\ln y = A + a \ln x_1 + b \ln x_2 + c \ln x_3 \tag{17}$$

where y and $x_i (i = 1,2,$ and $3)$ indicate the dependent and three independent variables, respectively. Parameter estimates to be measured by DEA/ERQ/LTM are listed with "A", "a", "b", and "c". As presented in (17), all the data points are transformed in a natural logarithm.

A stochastic frontier production function, applied to (17), has the following error structure:

$$\ln y = \ln(f(X) + u) + v \text{ or } y = \text{EXP}\{[\ln f(X) + u] + v\} \tag{18}$$

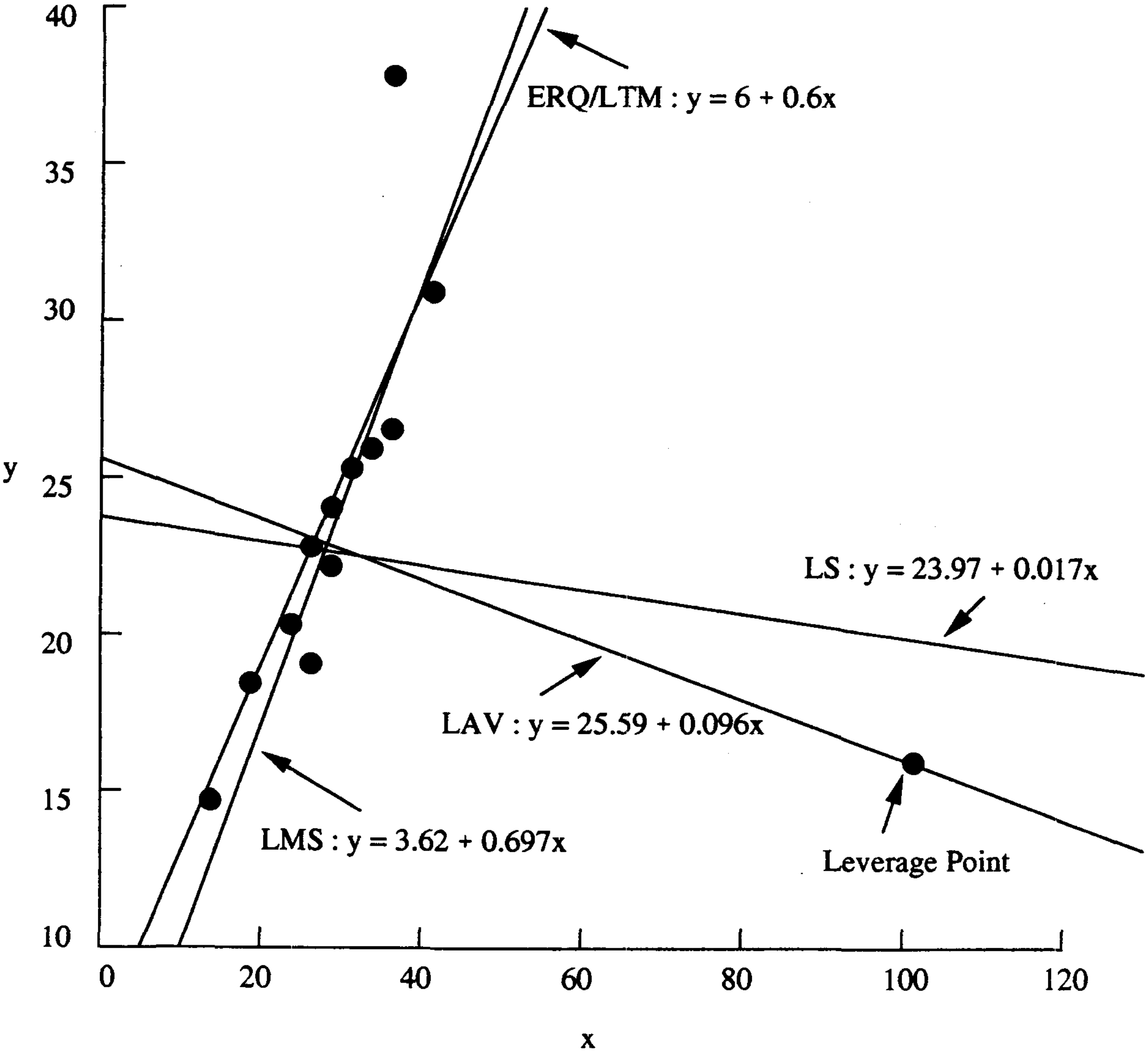

Figure 2 : Illustrative Example with LS, LAV, LMS and ERQ/LTM

where $f(X) = A + a \ln x_1 + b \ln x_2 + c \ln x_3$. The managerial erro ($u$) is due to managerial/decisional inefficiency such as inappropriate decision making, which is visually defined by a deviation from a frontier function. The DEA capability is fully utilized for measuring the level of the managerial error. It is very well known that a frontier production function is easily influenced by the existence of even a single outlier. It is also true that a bad leverage point affects seriously the resulting frontier production function. This application to the PTO data set, thus, illustrates practicality and empirical usefulness of ERQ/LTM in productivity analysis, both of which are not yet fully explored in Sueyoshi (1994b).

It is important to note that this research is aware of the existence of conventional studies on the stochastic frontier production analysis, including Aigner and Chu (1968), Aigner *et al.*

(1977), Greene (1990), Kopp and Smith (1980), Meusen and Van den Broeck (1977), Schmidt (1986), and Van den Broeck *et al.* (1980). The approach proposed in this article may serve as an alternative method when the conventional approaches have a difficulty in measuring a stochastic frontier production function, particularly in the cases where bad leverage points and outliers exist in an observed data set, and/or where underlying error distributions are difficult to determine for the managerial (u) and observational (v) errors.

5.2.2 LTM Application

When applying DEA/ERQ/LTM to the stochastic frontier production function of PTO in 24 OECD nations, this study needs to provide a comment on LTM's use and its application. That is, the purpose of LTM is to identify a bad leverage point that affects seriously the result of ERQ. However, it is not always true that its objective is to delete the bad leverage point. In some case where the leverage point is essential to make a production function, the data point needs to be still kept in its estimation process. For example, the PTO of the United States of America (U.S.A.) is much larger than the other OECD nations. Thus, the data point might have a high likelihood to become a bad leverage point and/or an outlier. If the American PTO were deleted from this empirical study, the resulting performance analysis would have only minor implication for the PTO evaluation in OECD nations. A user(s) needs to decide whether the bad leverage point is to be kept, deleted or corrected, depending upon the type of research and application.

The LTM results, applied to the PTO in the 24 OECD nations, are summarized in Table 2. The table exhibits dual variables derived from (2) and LTM-D scores measured by (14). Here, the dual examination determined a set of data points (H) whose dual variables belonged to $-1 < w < 1$.

$$H = \{\text{Denmark, Ireland, Luxembourg and Switzerland}\}.$$

Furthermore, the median point X_m was found to locate between the data points regarding Denmark and Luxembourg. Using (12), the arithmetic sample median X_S was identified as

$$X_S = \{8.237, 6.839, 3.388\}.$$

By comparing X_S and X values of the four nations in H, X_M was determined as the X of Switzerland, i.e.,

$$X_M = \{8.160, 7.231, 2.929\}.$$

Along with the X_M value, the LTM-D_j scores regarding 24 OECD countries were measured by (14). These LTM results are listed at the third column of Table 2. The cutoff value $\sqrt{\alpha_3^2, 0.975} = 3.058$ indicates there is no serious outlier in this logarithmically transformed data set. Exceptions may be found in Iceland and the United States. The two nations have a large magnitude (2.108 and 2.005, respectively) in these LTM-D scores. Furthermore, all the data points in H do not have high LTM-D scores, indicating that there is no bad leverage point affecting seriously the regression hyperplane: $\ln y = 1.849 + 0.127 \ln x_1 + 0.663 \ln x_2 + 0.190 \ln x_3$. Consequently, this study may now proceed for measuring a stochastic frontier production function to the PTO data set of 24 OECD nations.

Table 2 :Resulting Dual Variables and LTM-D Scores

Country	Dual Variable	LTM -D Score
Australia	-1.0000	0.5783
Austria	-1.0000	0.1736
Belgium	-1.0000	0.3612
Canada	1.0000	0.7343
Denmark	0.3108 (m)	0.3790
Finland	-1.0000	0.3742
France	1.0000	1.0520
Germany	-1.0000	1.2196
Greece	-1.0000	0.6861
Iceland	1.0000	2.1077
Ireland	0.7361	0.8194
Italy	-1.0000	0.9043
Japan	1.0000	1.4214
Luxembourg	0.3192 (m)	1.9681
Netherlands	1.0000	0.3253
New Zealand	1.0000	0.6682
Norway	1.0000	0.3663
Portugal	-1.0000	0.6189
Spain	-1.0000	0.5383
Sweden	-1.0000	0.3294
Switzerland	0.6338 (M)	0
Turkey	-1.0000	0.4304
United Kingdom	1.0000	1.0578
United States	1.0000	2.0048

Note : (m) indicates two data points between which a median point exists, and (M) indicates the data point that is the closest to the median point and the arithmetic sample median : Xs = (8.264, 6.839, 3.388).

5.2.3 DEA Application

This study has applied Data Envelopment Analysis (DEA), proposed first by Charnes *et al.* (1978), to the PTO data set so as to empirically determine the magnitude of $u_j (j = 1, .., n)$ incorporated in (18). Seiford (1993) provided a bibliography regarding DEA. Since a detailed description on DEA can be now found in Tone (1993) and Cooper *et al.* (1994) in

Japan, this article omits such a detailed comment on the DEA use and its implications. [See Chan and Sueyoshi (1991), Sueyoshi (1990a, 1991c, 1992b, 1992c) and Shang and Sueyoshi (1994) for various DEA applications. See also Chang and Sueyoshi (1991), Sueyoshi (1990b, 1992a, 1992d), and Sueyoshi and Chang (1989b) for algorithmic development related to several DEA models.] The original DEA, referred to as "Ratio-Form," can be mathematically formulated in the following GP model:

$$\begin{aligned} &\text{maximize } h \\ &s.t. \quad \sum_{j=1}^{n} X_j^T \lambda_j \leq X_o^T, \\ &\qquad -\sum_{j=1}^{n} y_j \lambda_j + y_o h \leq 0, \\ &\qquad \lambda_j \geq 0 \text{ and } h \geq 0, j = 1,..,n. \end{aligned} \tag{19}$$

An efficiency score (θ^*) of the specific o^{th} DMU is measured by $\theta^* = 1/h^*$, where h^* is derived from (19). The DEA efficiency score represents the degree of the specific o^{th} DMU on [0,1].

Utilizing the result of h^*, this study determines the estimated managerial error (u_o) as

$$u_o = y_o h^* - y_o. \tag{20}$$

In our proposed approach, DEA repeats n times the two computational processes: (19) and (20). [The subscript "o" is used to denote a specific DMU to be measured by (20). Meanwhile, the subscript "j" indicates the observed order of DMUs.]

After identifying $u_j (j = 1, \dots, n)$ by (20), each y_j is replaced by y_j^*, i.e.,

$$y_j^* = y_j + u_j = y_j h^*, \quad j = 1, \dots, n. \tag{21}$$

Then, (21) can be replaced by

$$\ln y_j^* = A + a \ln x_1 + b \ln x_2 + c \ln x_3 + v_j, \ j = 1, \dots, n. \tag{22}$$

5.2.4 ERQ Application

After completing DEA, this study applied ERQ to the adjusted data set ($\ln y^*, \ln X$) so that the resulting ERQ hyperplane became a stochastic frontier production function to the observed data set ($\ln y$, $\ln X$). First, the bottom hyperplane of G_A was estimated by (4) as follows:

$$\ln y = 1.861 + 0.261 \ln x_1 + 0.472 \ln x_2 + 0.284 \ln x_3.$$

Second, the upper hyperplane of G_B was estimated by (5) as follows:

$$\ln y = 1.731 + 0.265 \ln x_1 + 0.499 \ln x_2 + 0.249 \ln x_3.$$

Finally, the optimal α^* was estimated as 0.6003 by (8). Threfore, the estimated ERQ hyperplane of the stochastic frontier production function became

$$\ln y = 1.783 + 0.263 \ln x_1 + 0.488 \ln x_2 + 0.263 \ln x_3.$$

5.2.5 Statistical Test

As presented previously, this article proposes analytical and numerical approaches for conductiong various statistical tests. This study presents first the asymptotical approach and then compares it with the bootstrap method. The t-test is used as an example of many statistical tests, such as confidence intervals and hypothesis tests, in this illustration. Computational processes related to the t-test may be easily extended to other statistical tests.

In conducting the t-test based upon the ERQ asymptotical result, this study first measured the λ estimate, using (10). The error deviations (v_j) were rearranged by the symmetric order from the median sample point. Resulting $e(g), e(-g)$ and λ are listed in Table 3, along with the order of $g(g = 1, \ldots, 12)$. The λ estimates were used with a covariance matrix $(X^T X)^{-1}$ of the PTO data set, so as to estimate the standard deviation of the four parameters. All of them are listed in the right side of Table 3.

Besides the asymptotical approach, the bootstrap method presented in Section 3.2 was applied to estimate the standard deviation of the four parameters, as well. The resulting bootstrap estimates are summarized at the bottom of Table 3.

Table 3 : Estimate and Standard Deviation of Parameter Estimate

g	e (g)	e (-g)	$\hat{\lambda}$	Standard Deviation			
				I	ML	TI	PTO-E
1	0.000002 [21]	-0.00541 [2]	0.0324	0.0899	0.0246	0.0168	0.0173
2	0.022675 [19]	-0.01273 [15]	0.1062	0.2942	0.0805	0.0551	0.0567
3	0.034183 [18]	-0.01525 [7]	0.0989	0.2738	0.0750	0.0513	0.0527
4	0.034394 [23]	-0.01599 [14]	0.0756	0.2093	0.0573	0.0392	0.0403
5	0.050299 [5]	-0.01628 [8]	0.0799	0.2213	0.0606	0.0414	0.0426
6	0.051944 [4]	-0.01985 [1]	0.0718	0.1988	0.0544	0.0372	0.0383
7	0.056089 [17]	-0.02496 [16]	0.0695	0.1924	0.0527	0.0360	0.0371
8	0.057208 [20]	-0.03938 [22]	0.0724	0.2006	0.0549	0.0376	0.0386
9	0.073244 [3]	-0.04093 [13]	0.0761	0.2108	0.0577	0.0395	0.0406
10	0.074177 [24]	-0.04860 [12]	0.0737	0.2040	0.0559	0.0382	0.0393
11	0.092394 [6]	-0.08544 [11]	0.0970	0.2687	0.0735	0.0503	0.0517
12	0.096745 [10]	-0.20499 [9]	0.1509	0.4178	0.1144	0.0782	0.0805
Average			0.0837	0.2318	0.0635	0.0434	0.0447
Bootstrap Result				0.1724	0.0483	0.0329	0.0351

Note : The number in [ ] indicates the observed order of 24 data. The symbols "I," "ML," "TL," and "PTO-E" indicate intercept, main lines, telecommunications investment, and PTO employment , respectively.

By comparing the two results at the bottom of Table 3, this study finds that the bootstrap method has produced slightly smaller deviations in the four parameters than the asymptotic approach. Furthermore, it is found in Table 3 that the λ estimate depends upon the selection of "g". There is no way for uniquely determining the best choice of the g value. This feature indicates that different choices of g yield different λ estimates. Therefore, this study used its average to compute the standard error of ERQ parameter estimates. This operational difficulty can be considered as a methodological shortcoming of the asymptotic approach. Consequently, this study believes that it is better to measure the asymptotic standard errors in combination with another estimation technique such as the bootstrap method. Comparing the two approaches, we can confirm whether there is no major difference in the two results.

Table 4 : Test for H_o by Asymptotical Approach and Bootstrap Method

Variable	$\hat{\beta}_i$	$(\hat{\lambda}^2 c_{ii})^{1/2}$	$\hat{\beta}_i / (\hat{\lambda}^2 c_{ii})^{1/2}$
Intercept	1.783	[0.2318](0.1724)	[7.692](10.342)**
Main Lines	0.263	[0.0635](0.0483)	[4.142](5.445)**
Telecommunications Investment	0.488	[0.0434](0.0329)	[11.244](14.833)**
PTO Employment	0.263	[0.0447](0.0351)	[5.884]((7.493)**

Note : The symbol " ** " indicates the level at 1% significance of the t - test.
The numbers with [ ] and ( ) in the third column indicate the standard deviations measured by the asymptotical approach and the bootstrap method, respectively. The corresponding results are listed in the last column.

Table 4 exhibits a result on the t-test for the null hypotheses $H_o : \beta_i = 0$, using the standard errors measured by both the asymptotical approach and the bootstrap method, respectively. An important finding is identified in Figure 4; the bootstrap method rejected the null hypotheses related to all the four parameter estimates at the level of 1% significance. This result indicates that all the parameter estimates are important in making the stochastic frontier production function for PTO of the 24 OECD nations. The asymptotical approach also rejected the null hypotheses at the same level of significance. However, the magnitudes of $\hat{\beta}_i/(\hat{\lambda}^2 c_{ii})^{1/2} (i = 1,..,m)$ measured by the asymptotical approach are all smaller than those of the bootstrap method. This result is due to a fact that the standard error measured by the bootstrap technique is smaller than that of the asymptotical approach. [A different data may produce a distinct result between the two methods.]

5.2.6 Productivity Analysis

As mentioned previously, the resulting stochastic frontier production function can serve as an empirical basis for not only predicting future PTO performance, but also evaluating current PTO performance in OECD nations. As a research extension of Sueyoshi (1994b), this section compares the productivity measures of PTO performance in 24 OECD nations determined by DEA/ERQ/LTM with those of the previous DEA/LAV. Such empirical comparison is summarized in Table 5.

Table 5 : Efficiency Comparison between DEA/LAV and DEA/ERQ/LTM

Method / Country	DEA/LAV	DEA/ERQ/LTM
Australia	0.5726	0.5783
Austria	0.7443	0.7402
Belgium	0.7677	0.7891
Canada	0.8216	0.8330
Denmark	0.8340	0.8516
Finland	0.6912	0.7015
France	0.8458	0.8493
Germany	0.6961	0.6847
Greece	0.5290	0.5695
Iceland	1.0000	1.0000
Ireland	0.8678	0.8845
Italy	0.7065	0.7024
Japan	0.9611	0.9595
Luxembourg	0.9594	0.9841
Netherlands	0.9585	0.9873
New Zealand	0.8731	0.9016
Norway	1.0000	1.0000
Portugal	0.5718	0.5916
Spain	0.6143	0.6282
Sweden	0.6896	0.6991
Switzerland	1.0000	1.0000
Turkey	0.2534	0.2574
United Kingdom	0.8652	0.8861
United States	1.0000	1.0000

Note: DEA/LAV results are adapted from Sueyoshi (1994).

Table 5 indicates that there is no major difference between the two approaches, both reporting that the four nations, (Iceland, Norway, Switzerland, and the U.S.A.) are evaluated as 100% production efficiency. [This research believes that such similarity in productivity analysis is just a coincidence between DEA/ERQ/LTM and DEA/LAV. A different data set may produce a large discrepancy between the two methods.] The empirical result indicates that the U.S. can be considered to achieve the high productivity by utillzing its scale efficiency. This perception regarding the U.S. may be applicable to Japan, as well. Small nations such as Iceland, Norway and Switzerland also presented the high productivity. The other OECD nations represented some amount of inefficiency by comparing relatively these PTO performances with those of the four efficient nations. Thus, the resulting DEA/ERQ/LTM can serve as empirical basis for not only prediction but also productivity analysis.

6. Conclusion and Future Extensions

This article has achieved three research objectives that can be summarized as follows: First, this study has presented an asymptotic framework for ERQ/LTM that can be served as an analytical basis for conducting various statistical tests and measuring confidence intervals. The asymptotical approach was compared with the bootstrap method, so that this study confirmed whether the two methods yielded any significant difference in the t-test. Second, LTM was incorporated into ERQ. The LTM can identify a bad leverage point that may affect seriously ERQ results. As a consequence of incorporating LTM into ERQ, ERQ/LTM estimates become robust to not only an outlier but also a bad leverage point. Finally, as an important application, this study used ERQ/LTM in combination with DEA so that the three combination produced a stochastic frontier production function. This study has applied the DEA/ERQ/LTM production analysis to a data set regarding PTO performance in 24 OECD counties. The empirical result of DEA/ERQ/LTM was compared with DEA/LAV proposed by Suehoshi (1994b).

As extensions of this study, the following research issues need to be explored in the near future: First, we need to confirm the LTM performance for identifying bad leverage points by comparing it with other statistical methods in a simulation study. Second, DEA/ERQ/LTM needs to be compared with other econometric methods in terms of producing a stochastic frontier production function. Finally, it is hoped that this study can make a small contribution to the use and development of LTM, ERQ and DEA/ERQ/LTM. This reseach waits anxiously further extensions along the lines specified in this article.

References

[1] Aigner, D & Chu S. F., "On Estimation the Industry Production Function," *The American Economic Review,* 58 (1968), 826-839.

[2] Aigner, D., Lovell, K. & Schimidt P., "Formulation and Estimation of Stochastic Frontier Production Models," *Journal of Econometrics*, 6 (1977), 21-37.

[3] Atkinson, A. C., "Masking Unmasked," *Biometrika*, 73 (1986), 533-541.

[4] Bassett, G. & Koenker, R., "Asymptotic Theory of Least Absolute Error Regression," *Journal of the American Statistical Association*, 73 (1978), 618-622.

[5] Bassett, G. & Koenker, R., "An Empirical Quantile Function for Linear Models with IID Errors," *Journal of the American Statistical Association,* 77 (1982), 407-415.

[6] Belsley, D.A., Kuh, E. & Welsch, R.E., *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity,* John Wiley, New York, 1980.

[7] Chan, P. S. & Sueyoshi, T., "Environmental Change, Competition, Strategy, Structure and Firm Performance: an Application of Data Envelopment Analysis in the Airline Industry," *International Journal of System Science*, 22 (1991), 1625-1636.

[8] Chang, Y. & Sueyoshi, T., "An Interactive Application of Data Envelopment Analysis in Microcomputers," *Computer Science in Economics and Management,* 4 (1991), 51-64.

[9] Charnes, A. & Cooper, W. W., *Management Models and Industrical Applications of Linear Programming,* Wiley, New York, 1961.

[10] Charnes, A. & Cooper, W. W. "Goal Programming and Constrained Regression -a Comment," *Omega*, 3 (1975), 403-409.
[11] Charnes, A., Cooper, W. W. & Fergurson, R. O., "Optimal Estimation of Executive Compensation by Linear Programming," *Management Science*, 1 (1955), 138-150.
[12] Charnes, A., Cooper, W. W. & Rhodes, E., "Measuringthe Efficiency of Decision Making Units," *European Journal of Operational Research*, 2 (1978), 429-444.
[13] Charnes, A., Cooper, W. W. & Sueyoshi, T., "Least Square / Ridge Regression and Goal Programming / Constrained Regression Alternatives," *European Journal of Operational Research*, 27 (1986), 147-157.
[14] Charnes, A., Cooper, W. W. & Sueyoshi, T., "A Goal Programming / Constrained Regression Review of Bell System Breakup," *Management Science*, 34 (1988), 1-26.
[15] Cooper, W. W., Tone, K., Takamori, H., & Sueyoshi, T., "Data Envelopment Analysis: Survey and Interpretation (in Japanese)," *Operations Research: Communication of the Operations Research Society of Japan* (Forthcoming in 1994).
[16] Dielman, T. E. & Pfaffenberger, R. C., "LAV (Least Absolute Value) Estimation in Linear Regression: A Review," In S. H. Zanakis & J. S. Rustagi (Eds.), *TIMS Studies in the Management Sciences: Optimization in Statistics*, 19, North Holland, Amsterdam, 1982.
[17] Dielman, T. E. & Pfaffenberger, R. C., "Bootstrapping in Least Absolute Value Regression: an Application to Hypothesis Testing," *Communications in Statistics - Simulation and Computation*, 17 (1988),843-856.
[18] Fung, W. K., "Unmasking Outliers and Leverage Points: A Confirmation," *Journal of the American Statistical Association*, 88 (1933), 515-519.
[19] Greene, W. H., "A Gamma-Distributed Stochastic Frontier Model," *Journal of Econometrics*, 46 (1990), 141-163.
[20] Hawkins, D. M., Bradu, D., & Kass, G. V.. "Location of Several Outliers in Multiple Regression Data Using Elemental Sets," *Technometrics*, 26 (1984) 197-208.
[21] Hines, W. W. & Montgomery, D. C., *Probability and Statistics in Engineering and Management Science*, John Wiley & Sons, New York, 1980.
[22] Hogg, R. V., "An Introduction to Robust Procedures," *Communications in Statistics - Simulation and Computation*, A6-9 (1977), 789-794.
[23] Huber, P. J., *Robust Statistics*, Wiley, New York, 1981.
[24] Koenker, R. & Bassett, G., "Regression Quantiles," *Econometrica*, 46 (1978), 33-50.
[25] Kopp, R. & Smith, K., "Frontier Production Function Estimates for Steam Electric Generation: a Comparative Analysis," *Southern Economic Journal*, 46 (1980), 1049-1059.
[26] Meusen, W. & Van den Broeck, J., "Technical Efficiency and Dimension of the Firm: Some Results on the Use of Frontier Production Function," *Empirical Economics*, 2 (1977), 109-122.
[27] OECD, *Performance Indicators for Public Telecommunications Operations*, Paris, France, 1990.
[28] Rousseeuw, P. J., "Least Median of Squares Regression," *Journal of the American Statistical Association*, 79 (1984), 871-880.
[29] Rousseeuw, P. J. & Zomeren, B. C. V., "Unmasking Multivariate Outliers and Leverage Points," *Journal of the American Statistical Association*. 85 (1990), 633-639.
[30] Schmidt, P., "Frontier Production Function," *Econometric Review*, 4 (1986), 289-328.
[31] Seiford, L. H., "A Bibliography of Data Envelopment Analysis (1978-1993), "DEA Bibliography, Working Paper, Department of Industrial Engineering and Operations Research, University of Massachusetts, 1993.

[32] Shephard, R. W., *Theory of Cost and Production Function,* Princeton University Press, Princeton, 1970.

[33] Shang, J. & Sueyoshi, T., "A Unified Framework for the Selection of a Flexible Manufacturing System," *European Journal of Operational Research,* (Forthcoming in 1994).

[34] Sueyoshi, T., "A study of Efficiency Analysis Using DEA (in Japanese)," *Operations Research: Communications of the Operations Research Society of Japan,* 35 (1990a), 167-173.

[35] Sueyoshi, T., "Special Algorithm for an Additive Model in Data Envelopment Analysis," *Journal of the Operational Research Society,* 41 (1990b), 249-257.

[36] Sueyoshi, T., "Estimation of Stochastic Frontier Cost Function Using Data Envelopment Analysis: an Application to the AT&T Divestiture," *Journal of the Operational Research Society,* 42 (1991a), 463-477.

[37] Sueyoshi, T., "Empirical Regression Quantile," *Journal of the Operations Research Society of Japan,* 34 (1991b), 250-262.

[38] Sueyoshi, T., "Knowledge Utilization in Productivity Analsis: Use of Organizational Intelligence in Data Envelopment Analysis," *Journal of Management Science & Policy Analysis,* 8 (1991c), 319-330.

[39] Sueyoshi., T., "Measuring Technical, Allocative, Overall Efficiencies Using DEA," *Journal of the Operational Research Society,* 43 (1992a), 141-155.

[40] Sueyoshi, T., "Measuring Industrial Performance of Chinese Cities by Data Envelopment Analysis," *Socio-Economic Planning Science,* 26 (1992b),75-88.

[41] Sueyoshi, T., "Scale Efficiency of Nippon Telegram & Telephone: an Application of DEA (in Japanese)," *Operations Reseaerch: Communications of the Operations Research Society of Japan,* 37 (1992c), 210-219.

[42] Sueyoshi, T., "Algorithm Strategy for Assurance Regin Analysis in DEA," *Journal of the Operations Research Society of Japan,* 35, (1992d), 62-76.

[43] Sueyoshi, T., "Constrained Regression Median for Measuring Possible Salary Discrimination," *European Journal of Operational Research,* (Forthcoming in 1994a).

[44] Sueyoshi, T., "Stochastic Frontier Production Analysis: Measuring Performance of Public Telecommunications in 24 OECD Countries," *European Journal of Operational Research,* (Forthcoming in 1994b).

[45] Sueyoshi, T. & Anselmo P.C., "The Evans and Hechman Subadditivity Test," *The American Economic Review,* 76 (1986) 854-855.

[46] Sueyoshi, T. & Chang, Y., "Goal Programming Approach for Regression Median," *Decision Sciences,* 20 (1989a), 700-714.

[47] Sueyoshi, T. & Chang, Y., "Efficient Algorithm for the Additive and Multiplicative Models in Data Envelopment Analysis," *Operations Research Letters,* 8 (1989b), 205-213.

[48] Tone, K., *Data Envelopment Analysis*, Nippon Kagaku Gijutsu Renmei, Tokyo, 1993.

[49] Van den Broeck, J., Forsund, F., Hjalmarsson, L. & Meusen, W., "On the Estimation of Deterministic and Stochastic Frontier Production," *Journal of Econometrics,* 13 (1980), 117-138.

Toshiyuki Sueyoshi
The Ohio State University
College of Business
1775 College Road
Columbus, Ohio 43210, USA