Weighted Least Absolute Error Estimators of Regression Parameters

Moon Sup Song*

ABSTRACT

In the multiple linear regression model a class of weighted least absolute error estimators, which minimize the sum of weighted absolute residuals, is proposed. It is shown that the weighted least absolute error estimators with Wilcoxon scores are equivalent to the Koul's Wilcoxon type estimator. Therefore, the asymptotic efficiency of the proposed estimator with Wilcoxon scores relative to the least squares estimator is the same as the Pitman efficiency of the Wilcoxon test relative to the Student's t-test. To find the estimates the iterative weighted least squares method suggested by Schlossmacher is applicable.

1. Introduction

Consider the linear regression model

$$Y_i=\sum_{j=1}^{p} x_{ij}\beta_j+E_i,\quad i=1,\cdots,n,$$

where the x_{ij}'s are known constants, the β_j's are regression parameters to be estimated, and the E_i's are identically and independently distributed (iid) random errors.

Classically, the regression parameters are estimated by minimizing the sum of squared deviations. That is, the least squares estimator is a solution to the minimization problem

* Department of Computer Science and Statistics, Seoul National University. The author is grateful to Professor Sung H. Park for his careful reading of the manuscript and for several helpful suggestions.

$$\sum_{i=1}^{n}(Y_i-\sum_{j=1}^{p}x_{ij}\beta_j)^2=\min! \qquad (1.1)$$

or equivalently, is a solution of the system of equations

$$\sum_{i=1}^{n}(Y_i-\sum_{j=1}^{p}x_{ij}\beta_j)x_{ij}=0,\ \ j=1,\cdots,p.$$

The least squares estimators enjoy some optimality properties, if the errors are iid normal. But, it is well known that the least squares estimators are very poor when the errors have a long-tailed distribution.

Various mathods have been proposed for obtaining robust estimators of the regression parameters which are insensitive to the departures from normality and to the effects of outliers in the data.

A class of M-estimators in location problems was introduced by Huber (19-64). In the linear regression case (e.g. Huber (1973)), the class of M-estimators is defined by the solutions to the minimization problem

$$\sum_{i=1}^{n}\rho(Y_i-\sum_{j=1}^{p}x_{ij}\beta_j)=\min! \qquad (1.2)$$

where ρ is some convex function. Note that (1. 1) is a special case of (1. 2). If we differentiate (1. 2) with respect to β_j's, we obtain the system of equations

$$\sum_{i=1}^{n}\psi(Y_i-\sum_{j=1}^{p}x_{ij}\beta_j)x_{ij}=0,\ \ j=1,\cdots,\ p,$$

where $\psi(x)=(d/dx)\rho(x)$ is assumed to be continuous and bounded.

A second class of L-estimators of regression parameters based on linear combination of ordered statistics was investigated by Bickel (1973). A third class of R-estimators based on rank tests, in location problems, was proposed by Hodges and Lehmann (1963). The class of R-estimators was generalized to regression problems by Adichie (1967), Sen (1968), Jaeckel (1972), Jurečkova (1971), Koul (1969), Scholz (1978), and Sievers (1978), among others.

Jaeckel (1972) proposed a method which minimizes

$$\sum_{i=1}^{n}a_n(R_i)(Y_i-\sum_{j=2}^{p}x_{ij}\beta_j), \qquad (1.3)$$

where R_i is the rank of $Y_i - \sum x_{ij}\beta_j$ among $\{Y_i - \sum x_{ij}\beta_j, 1 \leq i \leq n\}$ and $a_n(1) \leq \cdots \leq a_n(n)$ are some nondecreasing scores satisfying $\sum a_n(i) = 0$. Note that the intercept β_1 of the linear regression model can not be obtained by the Jaeckel's method. If we differentiate (1. 3), the Jaeckel's method can be approximated by the method of Jurečková (1971) which minimizes

$$\sum_{j=1}^{p} \left| \sum_{i=1}^{n} a_n(R_i) x_{ij} \right|. \tag{1. 4}$$

Jaeckel (1972) showed that the two minimization problems given by (1. 3) and (1. 4) are asymptotically equivalent.

The least absolute error (LAE) estimator is a maximum likelihood estimator when the errors have a double exponential distribution. The asymptotic theory of the LAE estimators was established by Basset and Koenker (1978). According to their results (Basset and Koenker, p.618), for any error distribution for which the median is superior to the sample mean as an estimator of location, the LAE estimator is preferable to the least squares estimator, in the sense of having strictly smaller asymptotic confidence ellipsoids.

In this paper we propose an weighted LAE estimator which minimizes

$$\sum_{i=1}^{n} a_n(R_i) \left| Y_i - \sum_{j=1}^{p} x_{ij}\beta_j \right|, \tag{1. 5}$$

where R_i is the rank of $|Y_i - \sum x_{ij}\beta_j|$ among $\{|Y_i - \sum x_{ij}\beta_j|, 1 \leq i \leq n\}$. Note that when $a_n(R_i) = 1$, $i = 1, \cdots, n$, (1. 5) reduces to the LAE problem.

Schlossmacher (1973) suggested a procedure to compute the LAE estimates which uses the iterative weighted least squares method. His algorithm may also be applied to the minimization problem (1. 5) to obtain the weighted LAE estimates.

In Section 2 we introduce a class of dispersion measures and discuss some of their properties. The definition of our estimators is given and the invariance properties are studied. In Section 3 we show the asymptotic equivalence of the weighted LAE estimator with Wilcoxon scores to the Wilcoxon type estimator proposed by Koul (1969). But, it seems easier to compute our estimates than Koul's estimates. From the asymptotic equivalence, it follows that the asymptotic relative efficiency of the weighted LAE estimator with Wilcoxon scores relative to the least squares estimator is the same as that of the Wilcoxon test relative to the Student's t-test.

2. Notations and Preliminaries

Let $Y_1, \cdots, Y_n$ be independent random variables with continuous cumulative distribution function (cdf)

$$P(Y_i \leq y) = F(y - x_i\beta_0),\ \ i=1, \cdots, n,$$

where the $\beta_0' = (\beta_{01}, \cdots, \beta_{0p})$ is the vector of regression parameters to be estimated and the $x_i = (x_{i1}, \cdots, x_{ip})$ are the vectors of known constants which form the ith row of the $n \times p$ matrix

$$X_n = ((x_{ij}^{(n)})),\ \ j=1, \cdots, p;\ \ i=1, \cdots, n. \tag{2.1}$$

In the sequel we shall suppress the index n whenever feasible.

For fixed $Y_1, \cdots, Y_n$ and for any β, let R_i be the rank of $|Y_i - x_i\beta|$ among $\{|Y_i - x_i\beta|,\ 1 \leq i \leq n\}$. Let $a_n(i)$, $i=1, \cdots, n$, be a set of positive scores generated by

$$a_n(i) = \varphi(i/(n+1)),\ \ i=1, \cdots, n \tag{2.2}$$

where $\varphi(u)$, $0<u<1$, is the score generating function which is nondecreasing and square integrable on $(0, 1)$. Define

$$T(Y - X\beta) = \sum_{i=1}^{n} a_n(R_i)\,|Y_i - x_i\beta|. \tag{2.3}$$

We now derive some properties of $T(Y-X\beta)$ in the following theorems.

Theorem 2.1. For any fixed Y, $T(Y-X\beta)$ is a continuous and convex function of β.

〈**Proof**〉 Let $p = (p_1, p_2, \cdots, p_n)$ be any permutation of the indices $1, 2, \cdots, n$. Let P be the set of all such permutations. According to the same argument as in Theorem 1 of Jaeckel (1972), for any fixed residuals $Y_i - x_i\beta$, $i=1, \cdots, n$,

$$T_p = \sum_{i=1}^{n} a_n(p_i)\,|Y_i - x_i\beta|$$

is maximized over P when $p_i = R_i$, $i = 1, \cdots, n$. Thus, we have

$$T(Y-X\beta) = \max_{p \in P} \sum_{i=1}^{n} a_n(p_i) \, | Y_i - x_i \beta |. \tag{2.4}$$

Note that each term in the summation of (2.4) is a continuous and convex function in β. Therefore $T(Y-X\beta)$, which is the maximum of a finite number of such functions, is also continuous and convex (concave upward). This completes the proof.

For the case of a single parameter, we have

$$T(Y-X\beta) = \sum_{i=1}^{n} a_n(R_i) \, | Y_i - \beta x_i | \tag{2.5}$$

where β and each x_i are real numbers. From the convexity of T we may state the following lemma without proof.

Lemma 2.1. When $p=1$,

$$\frac{dT}{d\beta} = -\sum_{i=1}^{n} a_n(R_i) x_i \ \mathrm{sgn}(Y_i - \beta x_i)$$

is a nondecreasing step function in β. Furthermore, the minimum of $dT/d\beta$ is $-\sum a_n(R_i)|x_i|$ and the maximum is $\sum a_n(R_i)|x_i|$.

Thus, for the case of a single parameter, we can see that the graph of $T(Y-X\beta)$ will be an open convex polygon in the (T, β) plane. At each vertex the slope of T is increased by the amount of one of the following:

i) When the order of absolute residuals changes in one sign,

$$(a_n(k+1) - a_n(k)) \, | \, |x_{i(k+1)}| - |x_{i(k)}| \, |.$$

ii) When the order of absolute residuals changes in different signs,

$$(a_n(k+1) - a_n(k)) \, | \, |x_{i(k+1)}| + |x_{i(k)}| \, |.$$

iii) When the sign of $Y_{i(k)} - \beta x_{i(k)}$ changes,

$$2 \, | a_n(1) x_{i(1)} |.$$

Here $i(k)$ is the index of observation corresponding to the kth ordered residual.

Now for the general case, we are ready to prove the boundedness of the set of β for which $T(Y-X\beta) \leq T_0$ for any T_0. Using the notations in Jaeckel (1972), let E be the $n \times n$ matrix all of whose entries are $1/n$, and let $\bar{X} = EX$. Combining Theorem 2.1 and Lemma 2.1, we may state the following theorem whose proof is exactly the same as that of Theorem 2 of Jaeckel (1972).

Theorem 2. 2. If $\overline{X}-X$ has rank p, then for any T_0, the set $\{\beta : T(Y-X\beta)\leq T_0\}$ is bounded.

By Theorem 2. 2, If $\overline{X}-X$ has rank p, then the set of β for which $T(Y-X\beta)$ attains its minimum is bounded. Since $T(Y-X\beta)$ is actually a measure of dispersions of residuals, it is possible to estimate the vector of regression parameter β_0 by minimizing $T(Y-X\beta)$.

Definition. The weighted LAE estimator β_T is a solution to the minimization problem

$$T(Y-X\beta)=\sum_{i=1}^{n} a_n(R_i)\,|\,Y_i-x_i\beta\,|=\min!. \qquad (2.6)$$

Note that the solutions to (2. 6) may not be unique. However, according to Theorem 3. 2 below, they are asymptotically equivalent in the sense that they all have the same asymptotic distribution.

We now state some invariance properties of the weighted LAE estimator β_T.

Theorem2. 3. If $\beta_T(Y,X)$ is a weighted LAE estimate defined by (2. 6), then

i) $\beta_T(Y+X\beta_1,X)=\beta_T(Y,X)+\beta_1$ for any p-vector β_1,

ii) $\beta_T(\lambda Y,X)=\lambda\beta_T(Y,X)$ for any real λ.

⟨**Proof**⟩ i) Note that

$$\sum_{i=1}^{n} a_n(R_i)\,|\,(Y_i+x_i\beta_1)-x_i(\beta+\beta_1)\,|$$

$$=\sum_{i=1}^{n} a_n(R_i)\,|\,Y_i-x_i\beta\,|.$$

Thus, if β_T minimizes $\sum a_n(R_i)\,|\,Y_i-x_i\beta\,|$, then $\beta_T+\beta_1$ minimizes $\sum a_n(R_i)\,|\,(Y_i+x_i\beta_1)-x_j\beta\,|$.

ii) From the equality

$$\sum_{i=1}^{n} a_n(R_i)\,|\,\lambda Y_i-x_i(\lambda\beta)\,|=\lambda\sum_{i=1}^{n} a_n(R_i)\,|\,Y_i-x_i\beta\,|,$$

the second part is also obvious.

(2. 6) defines a class of weighted LAE estimators, one corresponding to

each score generating function φ in (2. 2). When $a_n(i)$ are Wilcoxon scores, that is when $\varphi(u)=u$, $0<u<1$, the asymptotic equivalence between the weighted LAE estimator and the Koul's Wilcoxon type estimator is shown in the next section.

3. Asymptotic Equivalence to Koul's Estimator and Asymptotic Properties

To show the asymptotic equivalence of the weighted LAE estimators defined by (2. 6) to the Koul's estimator, we shall briefly make assumptions in Koul (1969) and introduce the Koul's estimator.

We assume that the cdf F is absolutely continuous and has finite Fisher's information, that is

$$\int [f'(x)/f(x)]^2 f(x)dx < \infty$$

with f being the density of the distribution. We also assume that f is absolutely continuous, symmetric and bounded. Let

$$\Sigma_n = \frac{1}{n} X_n' X_n \qquad (3.1)$$

where X_n is defined by (2. 1). For the design matrix we assume that

$$\lim_{n\to\infty}(\max_{1\le i\le n} x_{ij}^2)/\left(\sum_{i=1}^{n} x_{ij}^{1/2}\right)=0$$

for all $j=1,\cdots,p$, and also assume that

$$\lim_{n\to\infty}\Sigma_n=\Sigma \qquad (3.2)$$

exists and is a positive definite matrix.

Define

$$S_{nj}(Y)=\frac{1}{n}\sum_{i=1}^{n} x_{ij}a_n(R_i)\ \mathrm{sgn}(Y_i),\ j=1,\cdots,p \qquad (3.3)$$

where R_i is the rank of $|Y_i|$ among $\{|Y_i|,\ 1\le i\le n\}$. Let

$$S_n'(Y)=(S_{n1}(Y),\cdots,S_{np}(Y)),$$

$$M_n(Y)=nS_n'(Y)\ \hat{\Sigma}_n S_n(Y)$$

where $\hat{\Sigma}_n$ is defined by

$$\hat{\sum}_n = \left(\int_0^1 \varphi^2(u)du\right)\sum_n. \tag{3.4}$$

Then by Lemma 1.1 of Koul (1969), $M_n(Y-X\beta)$ has a chi-square distribution with p degrees of freedom. Thus a confidence region with confidence coefficient $1-\alpha$ may be defined by

$$R_n(Y-X\beta)=\{\beta : M_n(Y-X\beta)\leq K\alpha\}. \tag{3.5}$$

The Koul's estimator $\hat{\beta}$ of the regression parameters is defined by the center of gravity of the confidence region (3.5), that is

$$\hat{\beta}=[\lambda(R_n(Y-X\beta))]^{-1}\int_{R_n(Y-X\beta)} t\lambda(dt) \tag{3.6}$$

where λ denotes p-dimensional Lebesque measure on p-dimensional Euclidean space. When $\varphi(u)=u$, the asymptotic normality of the $\hat{\beta}$ in (3.6) is proved by approximating $\hat{\beta}$ by $\tilde{\beta}$ defined below. Therefore, in this section we shall assume that $\varphi(u)=u$, $0<u<1$. Note that when $\varphi(u)=u$, the scores $a_n(i)$ in (2.2) are Wilcoxon scores.

However, as Koul mentioned (1969, p. 1953), it is believed that asymptotic results will remain valid for a class of score generating function φ which is monotone, square integrable and have first two integrable derivatives.

We introduce some further notations. Define

$$\hat{S}_{nj}(Y)=\frac{1}{n}\sum_{i=1}^{n} x_{ij}\{2F(|Y_i|-1)\}\ \mathrm{sgn}(Y_i)$$

for $j=1,\cdots,p$, and let

$$\hat{S}_n'(Y)=(\hat{S}_{n1}(Y),\cdots,\hat{S}_{np}(Y)).$$

Then by Equation (3.45) of Koul, for every $\varepsilon>0$

$$\lim_{n\to\infty} P[\sqrt{n}|\hat{S}_{nj}(Y)-S_{nj}(Y)|\geq\varepsilon]=0 \tag{3.7}$$

for $j=1,\cdots,p$, where $S_{nj}(Y)$ is defined by (3.3).

We now define $\tilde{\beta}$, which is asymptotically equivalent to $\hat{\beta}$, by

$$\tilde{\beta}=(6\gamma)^{-1}\hat{\sum}_n\hat{S}_n(Y) \tag{3.8}$$

where $\hat{\sum}_n$ is defined by (3.4) and

$$\gamma=\int f^2(x)dx \tag{3.9}$$

Notice that $\hat{\sum}_n=(1/3)\sum_n$ when $\varphi(u)=u$, which will be used later.

Definition. Two sequences of random vectors $\{\bar{\beta}_n\}$ and $\{\beta_n\}$ are said to be asymptotically equivalent if

$$\lim_{n\to\infty} P[\sqrt{n}\|\beta_n-\bar{\beta}_n\|\geq\varepsilon]=0$$

holds for every $\varepsilon>0$, where $\|\cdot\|$ is defined for a p-vector t by

$$\|t\|=\sum_{i=1}^{p}|t_i|.$$

By Lemma 4.1 and Theorem 4.1 of Koul (1969), we have the following lemma.

Lemma 3.1. Under the assumptions of Koul (1969), $\{\hat{\beta}\}$ in (3.6) and $\{\tilde{\beta}\}$ in (3.8) are asymptotically equivalent. It follows that $\sqrt{n}\ (\tilde{\beta}-\beta_0)$ is asymptotically normal

$$(0,\ (6\gamma)^{-2}\hat{\Sigma}^{-1})$$

where $\hat{\Sigma}^{-1}=\lim\ \hat{\Sigma}_n{}^{-1}$ and γ is defined by (3.9).

Asymptotic normality of the weighted LAE estimator β_T will be proved by showing the asymptotic equivalence of β_T to $\tilde{\beta}$. For any $\varepsilon>0$, define

$$V_n(a)=\{\beta:\ \sqrt{n}\|\beta\|\leq a\}.$$

By Corollary 3.1 of Koul (1969) we have the following lemma which combining with (3.7), represents the asymptotic linearity of $S_{nj}(Y-X\beta)$ in β.

Lemma 3.2. Under the assumptions of Koul (1969), for every $\varepsilon>0$

$$\lim_{n\to\infty} P\Big[\sup_{\beta\in V_n(a)}\sqrt{n}\,|S_{nj}(Y-X\beta)-\hat{S}_{nj}(Y)+2r\sigma_j\beta|\geq\varepsilon\Big]=0$$

for all $i=1,\cdots,p$ and any $0<a<\infty$, where σ_j is the jth row of $\sum_n$ in(3.1).

Let

$$D(Y-X\beta)=\frac{1}{\sqrt{n}}T(Y-X\beta)=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}a_n(R_i)\,|\,Y_i-x_i\beta|.$$

Then the partial derivatives of $D(Y-X\beta)$ are

$$\frac{\partial D(Y-X\beta)}{\partial\beta_j}=-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}x_{ij}a_n(R_i)\ \mathrm{sgn}(Y_i-x_i\beta), \qquad (3.10)$$

for $j=1,\cdots,p$. To show the asymptotic equivalence we will firstly approximate $D(Y-X\beta)$ by a quadratic form, and then we will show that the points where they are minimized approach each other.

We now define the quadraitc function

$$Q(\beta)=\sqrt{n}(r\beta'\sum_n\beta-\hat{S}_n'(Y)\beta)+D(Y). \qquad (3.11)$$

The partial derivatives of $Q(\beta)$ are

$$\frac{\partial Q(\beta)}{\partial\beta_j}=\sqrt{n}(2r\sigma_j\beta-\hat{S}_{nj}(Y)),\ j=1,\cdots,p,$$

and $Q(\beta)$ is minimized at the unique solution of the system

$$2\gamma\sum_n\beta=\hat{S}_n(Y).$$

Note that this solution is the same as the $\tilde{\beta}$ defined by (3.8).

The following Lemma 3.3 and Theorem 3.1 are similar to Lemma 1 and Theorem 3 of Jaeckel (1972), respectively.

Lemma 3.3. Under the assumptions of Koul (1969),

$$\lim_{n\to\infty} P\Big[\max_{\beta\in V_n(\mathrm{a})}|Q(\beta)-D(Y-X\beta)|\geq\varepsilon\Big]=0$$

holds for every $\varepsilon>0$ and $0<a<\infty$.

⟨**Proof**⟩ By (3.10), (3.11) and Lemma 3.2, it can be easily shown that

$$\lim_{n\to\infty} P\Big[\sup_{\beta\in V_n(a)}\Big|\frac{\partial Q(\beta)}{\partial\beta_j}-\frac{\partial D(Y-X\beta)}{\partial\beta_j}\Big|\geq\frac{a}{\sqrt{n}}\varepsilon\Big]=0 \qquad (3.12)$$

holds for every $\varepsilon>0$ and $a>0$. Given $\varepsilon>0$ and $a>0$, we choose $\beta_1\in V_n(a)$. Then for $0\leq t\leq 1$,

$$\frac{d}{dt}[Q(t\beta_1)-D(Y-tX\beta_1)]$$

$$=\sum_{j=1}^{p}\beta_{1j}\Big[\frac{\partial}{\partial\beta_{1j}}Q(t\beta_1)-\frac{\partial}{\partial\beta_{1j}}D(Y-tX\beta_1)\Big]. \qquad (3.13)$$

Combining (3.12) and (3.13), we have for $0\leq t\leq 1$

$$\lim_{n\to\infty} P\Big[\Big|\frac{d}{dt}\{Q(t\beta_1)-D(Y-tX\beta_1)\}\Big|\geq\varepsilon\Big]=0 \qquad (3.14)$$

for every $\beta_1\in V_n(a)$.

Since $Q-D$ is continuous on $V_n(a)$ which is closed, it takes the maximum on $V_n(a)$. Note also that $Q=D$ at $t=0$. Therefore by (3.14) with $t=1$, we have

$$\lim_{n\to\infty} P\Big[\max_{\beta\in V_n(a)}|Q(\beta)-D(Y-X\beta)|\geq\varepsilon\Big]=0,$$

completing the proof.

Let

$$B_N = \{\beta : T(Y-X\beta) \text{ is minimized}\}. \tag{3.15}$$

Then the weighted LAE estimators β_T are points of B_N, and they are asymptotically equivalent by the following theorem.

Theorem 3. 1. Under the assumptions of Koul (1969),

$$\lim_{n\to\infty} P\Big[\sup_{\beta_T\in\beta_N} \sqrt{n}\,\|\beta_T-\tilde{\beta}\|\geq\varepsilon\Big]=0$$

holds for every $\varepsilon>0$.

〈**Proof**〉 Because of Theorem 2. 3 we may assume that $\beta_0=0$ without any loss of generality. To prove the theorem it is enough to show that for all $\varepsilon>0$ and $\delta>0$,

$$P\Big[\sup_{\beta\in B_N} \sqrt{n}\,\|\beta-\tilde{\beta}\|\geq\varepsilon\Big]\leq\delta$$

for sufficiently large n.

Choose $\varepsilon>0$ and $\delta>0$, and let

$$r=\min\{Q(\beta) : \sqrt{n}\,\|\beta-\tilde{\beta}\|=\varepsilon\}-Q(\tilde{\beta}). \tag{3.16}$$

Then, since $Q(\beta)$ has a unique minimum at $\tilde{\beta}$, $r>0$. therefore by Lemma 3. 3, there exists $n(\delta_1)$ such that

$$P\Big[\max_{\beta\in V_n(a)} |Q(\beta)-D(Y-X\beta)|\geq\frac{r}{2}\Big]\leq\frac{\delta}{2} \tag{3.17}$$

for all $n\geq n(\delta_1)$.

However, by Lemma 3. 1, $\sqrt{n}\,\tilde{\beta}$ has a limiting normal distribution. It follows that for a given $\delta>0$ there exist a_0 and $n(\delta_2)$ such that $n\geq n(\delta_2)$ implies

$$P\Big[\sqrt{n}\,\|\tilde{\beta}\|\leq a_0\Big]\geq 1-\frac{\delta}{2}. \tag{3.18}$$

By (3. 17) and (3. 18), we have for all $n\geq n(\delta)=\max\{n(\delta_1),n(\delta_2)\}$,

$$P\Big[\sqrt{n}\,\|\tilde{\beta}\|\leq a_0 \text{ and } D(Y-X\tilde{\beta})<Q(\tilde{\beta})+\frac{r}{2}\Big]\geq 1-\delta. \tag{3.19}$$

Also, for any β such that $\sqrt{n}\,\|\beta-\tilde{\beta}\|=\varepsilon$ and $n>n(\delta)$,

$$P\Big[\sqrt{n}\,\|\beta\|\leq a_0+\varepsilon \text{ and } D(Y-X\beta)>Q(\beta)-\frac{r}{2}\Big]\geq 1-\delta. \tag{3.20}$$

But, by the definition of r in (3. 16), for every β such that $\sqrt{n}\,\|\beta-\tilde{\beta}\|=\varepsilon$

we have

$$Q(\beta)\geq Q(\tilde{\beta})+r. \tag{3.21}$$

Combining (3.19), (3.20) and (3.21), we have that $n\geq n(\delta)$ implies

$$P[D(Y-X\beta)>D(Y-X\tilde{\beta})]\geq 1-\delta \tag{3.22}$$

for every β such that $\sqrt{n}\,\|\beta-\tilde{\beta}\|=\varepsilon$. However, by the convexity of D, the strict inequality in (3.22) holds for all β such that $\sqrt{n}\,\|\beta-\tilde{\beta}\|\geq\varepsilon$. Therefore, $\beta\in B_N$ implies $\sqrt{n}\,\|\beta-\tilde{\beta}\|\leq\varepsilon$ with probability at least $1-\delta$ for all $n\geq n(\delta)$. Thus we have for every $n\geq n(\delta)$,

$$P\Big[\sup_{\beta\in B_N}\sqrt{n}\,\|\beta-\tilde{\beta}\|\leq\varepsilon\Big]\geq 1-\delta.$$

This concludes the proof.

By Lemma 3.1 and Theorem 3.1, we have the following main theorem of this section.

Theorem 3.2. Under the assumptions of Koul (1969), the weighted LAE estimators $\beta_T\in B_N$ are asymptotically equivalent. Moreover, $\sqrt{n}\,(\beta_T-\beta_0)$ has asymptotically multivariate normal distribution with mean zero and covariance matrix

$$\left[12\left\{\int f^2(x)dx\right\}^2\right]^{-1}\Sigma^{-1},$$

where $\Sigma=\lim_{n\to\infty}\left(\frac{1}{n}X_n'X_n\right)$.

〈**Proof**〉 If we notice that $\hat{\Sigma}_n=(1/3)\Sigma_n$ when $\varphi(u)=u$, the theorem follows immediately from Lemma 3.1 and Theorem 3.1.

It is well known that the asymptotic distribution of the least squares estimator $\sqrt{n}\,\beta^*$ is multivariate normal with mean $\sqrt{n}\,\beta_0$ and covariance matrix $\sigma^2\Sigma^{-1}$, where σ^2 is the common variance of Y_i's. Therefore, if we define the asymptotic relative efficiency of β_T relative to β^* as the inverse ratio of their generalized limiting variances, and denote it by $e(\beta_T,\beta^*)$, then we have

$$e(\beta_T,\beta^*)=12\sigma^2\left[\int_{-\infty}^{\infty}f^2(x)dx\right]^2. \tag{3.23}$$

Note that this is nothing but the Pitman efficiency of the Wilcoxon test relative to the Student's t-test. Thus, when $F(x)$ is normal, the asymptotic efficiency is $\pi/3=0.955$. When $F(x)$ is double exponential, the Pitman efficiency of the Wilcoxon test to t-test is 1.5. If $F(x)$ has longer tails (such as Cauchy), then the asymptotic efficiency in (3.23) may be infinitely large, and for any continuous $F(x)$ with finite variance it cannot be less than 0.864.

COMMENTS

As mentioned in Section 1 the Schlossmacher's algorithm, which uses the iterative weighted least squares method, may be used for the minimization problem (2.6). But, since the weights are changing in each iteration according to the order of residuals, it is possible that the convergence fails very often. We note also that the second deritives of the dispersion measure function T in (2.6) are identically zero wherever they exist. Therefore we cannot use any algorithm which is based on second derivatives, like Newton's method. One possibility is the use of steepest descent method, which is unfortunately very slow in convergence and oscillates very often around the minimum point. Thus, developing more efficient algorithms for the minimization problem (2.6) is subject to further study. Because of its nice asymptotic properties, a comparative study for small samples is also worthy to be done.

REFERENCES

[1] Adichie, J.N., "Estimates of regression parameters based on rank tests," *The Annals of Mathematical Statistics*, 38, 1967, 894-904.

[2] Basset, G., Jr. and Koenker, R., "Asymptotic theory of least absolute error regression," *Journal of the American Statistical Association*, 73, 1978, 618-622.

[3] Bickel, P.J., "On some analogues to linear combinations of order statistics and linear model," *The Annals of Statistics* 1, 1973, 597-616.

[4] Hodges, J.L. and Lehmann, E.L., "Estimates of location based on rank tests," *The Annals of Mathematical Statistics*, 34, 1963, 598-611.

[5] Huber, P.J., "Robust estimation of a location parameter," *The Annals of Mathematical Statistics*, 35, 1964, 73-101.

[6] Huber, P. J., "Robust regression: Asymptotics, conjectures and Monte Carlo," *The Annals of Statistics* 1, 1973, 799-821.

[7] Jaeckel, L.A., "Estimating regression coefficients by minimizing the dispersion of the residuals," *The Annals of Mathematical Statistics*, 43, 1972, 1449-1458.

[8] Jurečkové, J., "Nonparametric estimate of the regression coefficients," *The Annals of Mathematical Statistics*, 42, 1971, 1328-1338.

[9] Koul, H.L., "Asymptotic behavior of Wilcoxon type confidence regions in multiple linear rgression," *The Annals of Mathematical Statistics*, 40, 1969, 1950-1979.

[10] Schlossmacher, E. J., "An iterative technique for absolute deviations curve fitting," *Journal of the American Statistical Association*, 68, 1973, 857-859.

[11] Scholz, F.W., "Weighted median regression estimates," *The Annals of Statistics*, 6, 1978, 603-609.

[12] Sen, P.K., "Estimates of the regression coefficients based on Kendall's tau," *Journal of the American Statistical Association*, 63, 1968, 1379-1389.

[13] Sievers, G. L. "Weighted rank statistics for simple linear regression," *Journal of the American Statisttical Association*," 73, 1978, 628-631.