Journal of the Korean
Statistical Society
Vol. 24, No. 1, 1995

Asymptotic Properties of Nonlinear Least Absolute Deviation Estimators

Hae Kyung Kim and Seung Hoe Choi [1]

ABSTRACT

This paper is concerned with the asymptotic properties of the least absolute deviation estimators for nonlinear regression models. The simple and practical sufficient conditions for the strong consistency and the asymptotic normality of the least absolute deviation estimators are given. It is confirmed that the extension of these properties to wide class of regression functions can be established by imposing some condition on the input values. A confidence region based on the least absolute deviation estimators is proposed and some desirable asymptotic properties including the asymptotic relative efficiency are also discussed for various error distributions. Some examples are given to illustrate the application of main results.

Keywords: Nonlinear regression model, Least absolute deviation estimators, Consistency, Normality

[1]Department of Mathematics, Yonsei University, Seoul 120, Korea

1. INTRODUCTION

The least squares method plays an important role in the statistical inference of regression parameters, both for linear models and for nonlinear models. However, in the spite of the theoretical and practical merits, basic criticisms of procedures based on the least squares method in the past have pointed to the lack of the robustness even from a single outlier or from a slight departure from the normality assumption on the errors ; General discussions of robustness are given in Huber (1972) and Hampel (1973). To overcome this difficulty an alternative procedure based on the least absolute deviations using the L_1-norm rather than the least squares using the L_2-norm may be preferred.

In this paper we confine our attention to the nonlinear L_1-normed estimation by establishing the asymptotic properties of the least absolute deviation estimator under mild conditions, and confirm that the least absolute deviation estimator is a natural analog of the sample median for the nonlinear regression model.

We consider the following nonlinear regression model for a univariate response y,

$$y_t = f(x_t, \theta_o) + \epsilon_t, \qquad t = 1, ..., n, \tag{1.1}$$

where $x_t \in R^m$ denotes the tth fixed known input vector, $\theta_o \in R^p$ is the parameter vector from a parameter space $\Theta \subset R^p$, $f : R^{p+m} \to R^1$ is a function of x and θ, and ϵ_t are random errors.

The parameter θ_o is unknown and the regression problem is to make inference about θ in some optimal way on the basis of observations on y_t, and $x_t, t = 1, ..., n$. The least squares estimator (LSE) $\tilde{\theta}_n$ is any vector in Θ which minimizes the mean square deviation

$$S_n(y, \theta) = \frac{1}{n} \sum_{t=1}^{n} (y_t - f(x_t, \theta))^2,$$

where $y = (y_1, ..., y_n)$. Alternatively, the least absolute deviation estimator (LAD) $\hat{\theta}_n$ is defined by any vector minimizing the mean absolute deviation

$$D_n(y, \theta) = \frac{1}{n} \sum_{t=1}^{n} |y_t - f(x_t, \theta)|. \tag{1.2}$$

The LSE or LAD is a particular case with $\rho(x) = x^2$ or $\rho(x) = |x|$ of a general class of robust methods based on minimizing an expression $\frac{1}{n}\sum_{t=1}^{n}\rho(y_t - f(x_t, \theta))$ for a suitable choice of the ρ function.

Various authors have provided conditions which ensure the existence, consistency, and asymptotic normality of the nonlinear LSE and LAD. Jennrich (1969) and Malinvaud (1970) proved strong consistency and asymptotic normality of LSE when the parameter space is a compact subset of R^p and the errors, ϵ_t are independent and identically distributed (i.i.d.) random variables. Oberhofer (1982) gave weak consistency result of LAD when the parameter space is a compact subset of R^p and the errors are independent random variables. Richardson and Bhattacharyya (1987), in a recent paper, extends the strong consistency of LAD to any separable, completely regular topological parameter space. Asymptotic normality of LAD for the nonlinear model has received almost no attention in the statistical literature. Jennrich (1969) also showed that for the LSE, $\sqrt{n}(\hat{\theta}_n - \theta_o)$ is asymptotically normally distributed with mean zero and variance $\sigma^2 Q^{-1}$ where $Q = \lim_{n\to\infty} \frac{1}{n} J(\theta)' J(\theta)$ where $J(\theta)$ is the usual $n \times p$ Jacobian matrix of the $f(x_t, \theta)$.

Bassett and Koenker (1978) and Nyquist (1983) showed that under some regularity conditions for the linear LAD $\hat{\theta}_n$, asymptotically $\sqrt{n}(\hat{\theta}_n - \theta_o) \sim N(0, \omega^2 {Q_o}^{-1})$ where the moment ratio parameter $\omega = 1/2g(0)$ and $Q_o = \lim_{n\to\infty} \frac{1}{n} X'X$ which is positive definite with rank(Q_o)=p. Here, $g(0)$ is the ordinate of the error density function at median 0.

The results of Jennrich and Nyquist led Gonin and Money (1985) to conjecture that the nonlinear LAD including L_p-norm estimator, $\sqrt{n}(\hat{\theta}_n - \theta_o)$ is asymptotically normally distributed with mean zero and variance $\omega_p^2 Q^{-1}$ where ω_p depends on p. However, so far nobody has proved rigorously this result of the nonlinear LAD, which is the principal result of this paper.

The main purpose of this paper is to provide simple sufficient conditions for the asymptotic normality of the LAD in the nonlinear regression model (1.1), and to confirm that the asymptotic efficiency of the LAD are, in general, superior to these of the LSE. For these, under mild regularity conditions the asymptotic normality of the LAD is proved in Theorem 3.1 of Section 3. A confidence region based on the LAD is proposed and some desired asymptotic properties including the asymptotic relative efficiency are also discussed in Section 4. Some examples of the application of the main results are contained in sections.

2. STRONG CONSISTENCY

We start this section by introducing some conditions on regression function and input values which ensure existence and strong consistency of LAD $\hat{\theta}_n$.

Throughout the paper we make the following assumptions on the model (1.1) : Assumption A.

$A1$: θ_o is an interior point of a convex and compact set $\Theta \subset R^p$.

$A2$: The function $f(x,\theta)$ is continuous in (x,θ).

Under Assumption A, Its existence and measurability are then immediate result of Lemma 2 of Jennrich (1969).

In addition to Assumption A, we will assume Assumption B :

$B1$: $\{\epsilon_t\}$ are i.i.d. random variables with the continuous distributon function G for which $G(0) = 1/2$ uniquely.

$B2$: The sequence of inputs $\{x_t\}$ generates Cesàro summable sequences with respect to a probability measure μ defined on the Borel subsets of Ξ, where Ξ is a subset of $\{x \in R^m : |f(x,\theta) - f(x,\theta_o)| > c$ for some $c > 0, \theta \neq \theta_o$ in $\Theta\}$. i.e., for every real valued continuous function h with $\int_\Xi |h(x)|d\mu(x) < \infty$,

$$\lim_{n\to\infty} \frac{1}{n}\sum_{t=1}^{n} h(x_t) = \int_\Xi h(x)d\mu(x).$$

Remark 1. Assumption $B2$ is the regularity condition on the limiting behavior of inputs for the weak convergence of measures. One simple way of generating such sequence is to choose inputs as a random sample from some distribution function $H(x)$ defined on Ξ by the Strong Law of Large Numbers (SLLN).

Now, we shall state the strong consistency of LAD.

Theorem 2.1. For the model (1.1), suppose that Assumptions A and B are fulfilled. Then the LAD $\hat{\theta}_n$ defined on (1.2) is strongly consistent for θ_o.

Proof. The proof follows in the similar manner as the proof of Theorem of Oberhofer (1982), so that we shall be brief in here. Since Θ is compact, it suffices to show that for any $\delta > 0$, $\liminf_{n\to\infty} \inf_{|\theta-\theta_o|\geq\delta}(D_n(y,\theta) - D_n(y,\theta_o)) > 0$ almost surely (a.s.), due to $D_n(y,\hat{\theta}_n) - D_n(y,\theta) \leq 0$ for every $\theta \in \Theta$. Let $d_t(\theta) = f(x_t,\theta) - f(x_t,\theta_o)$. Note that $D_n(y,\theta) - D_n(y,\theta_o) = \frac{1}{n}\sum_{t=1}^{n}\{|\epsilon_t + d_t(\theta)| - |\epsilon_t|\}$. Let $X_t = |\epsilon_t + d_t(\theta)| - |\epsilon_t|$. Then $\{X_t\}$ is a sequence of independent random variables with uniformly bounded variances due to the boundedness of f, so that by the SLLN we have $\frac{1}{n}(S_n - ES_n) \to 0$ a.s., where $S_n = \sum_{t=1}^{n} X_t$. Therefore, it suffices to show that

$$\liminf_{n\to\infty} \inf_{|\theta-\theta_o|\geq\delta} \frac{ES_n}{n} \geq \eta \quad a.s.,$$

for some real $\eta > 0$. Let $T_1 = \{t | d_t(\theta) \leq 0\}$ and $T_2 = \{t | d_t(\theta) > 0\}$, with the number of elements n_1 and n_2 respectively. Then $\frac{1}{n}ES_n$ becomes

$$\frac{2}{n_1}\sum_{t\epsilon T_1}\int_{(0,-d_t(\theta)]}\{|d_t(\theta)| - x\}dG(x) + \frac{2}{n_2}\sum_{t\epsilon T_2}\int_{[-d_t(\theta),0)}\{|d_t(\theta)| + x\}dG(x). \quad (2.1)$$

From the fact that $x \leq -\frac{d_t(\theta)}{2}$ on $(0, -\frac{d_t(\theta)}{2}]$ and $x \geq -\frac{d_t(\theta)}{2}$ on $[-\frac{d_t(\theta)}{2}, 0)$, the (2.1) is greater than or equal to

$$\frac{1}{n_1}\sum_{t\epsilon T_1}\int_{(0,-\frac{d_t(\theta)}{2}]}|d_t(\theta)|dG(x) + \frac{1}{n_2}\sum_{t\epsilon T_2}\int_{[-\frac{d_t(\theta)}{2},0)}|d_t(\theta)|dG(x)$$

Thus, in virtue of $B2$, we obtain

$$\frac{1}{n}ES_n \geq \frac{h}{n}\sum_{t=1}^{n}|d_t(\theta)|$$

where $h = \min_{t\in T_1\cup T_2}\{|G(-\frac{d_t(\theta)}{2}) - G(0)|\}$. Then

$$\liminf_{n\to\infty} \inf_{|\theta-\theta_o|\geq\delta} \frac{ES_n}{n} \quad \geq \quad h\int_{\Xi}|f(x,\theta) - f(x,\theta_o)|d\mu > \eta.$$

Therefore the proof is complete.

Remark 2. Theorem 2.1 still holds when $\sum_{t=1}^{n}|f(x_t,\theta) - f(x_t,\theta_o)|$ diverges to infinity at rate faster than n, provided that for every $\theta \neq \theta_o$ in Θ, $|f(x,\theta) - f(x,\theta_o)| \geq c$ on Ξ for some $c > 0$. This result suggests that $\{x_t\}_{t=1}^{n}$ should be chosen such that $\sum_{t=1}^{n}|f(x_t,\theta) - f(x_t,\theta_o)|$ as large as possible. In this point of view, although the choice of $B2$ for input values $\{x_t\}$ may not be optimal, it does guarantee that minimum information is gathered to allow for the strong consistency of LAD.

Remark 3. The result of Theorem 2.1 may be extends to the more general parameter space which is noncompact but separable and completely reqular using same technical procedure used in Richardson and Bhattacharyya (1987).

For the applications of Theorem 2.1, we now consider several regression functions. Throughout the example we assume that the input values $\{x_t\}$ which give rise to observations $y_1, y_2, ..., y_n$ are chosen as a realization of a random sample from the distribution function $F(x)$ defined on $\Xi \subset R^m$.

Example 2.1. Consider the exponential model $y_t = f(x_t, \theta_o) + \epsilon_t$, with the regression function $f(x, \theta) = e^{-\theta x}$, where $\{\epsilon_t\}$ is a sequence of i.i.d. random variables with the distribution function G for which $G(0) = 1/2$. Suppose that Ξ is a subset of $[T, \infty)$, $T > 0$ and $\Theta = [0, \alpha] \subset R^1$, where $\alpha > 0$. Evidently, the regression function satisfies the assumptions of Theorem 2.1. We can guarantee, therefore, the strong consistency of the LAD $\hat{\theta}_n$ under the sampling scheme on Ξ.

Example 2.2. Consider the model $y_t = f(x_t, \theta_o) + \epsilon_t$, where $\theta_o \in \Theta = [0, \alpha_1] \times [0, \alpha_2], \alpha_1,\ \alpha_2 < \infty$ and $f(x, \theta) = \theta_1 e^{-\theta_2 x}$, $\theta = (\theta_1, \theta_2) \in \Theta$. Assume that $\{\epsilon_t\}$ are i.i.d. random variables each having median zero. The assumptions of Theorem 2.1 are easily satisfied. Note that $\theta_1 e^{-\theta_2 x} = \theta_1' e^{-\theta_2' x}$ if and only if $\theta_1 = \theta_1'$ and $\theta_2 = \theta_2'$. Let $\Xi \subset \{x : |f(x, \theta) - f(x, \theta')| \geq c, \theta \neq \theta' \in \Theta$ for some $c > 0\}$. Then under the sampling scheme on Ξ, the LAD $\hat{\theta}_n$ is strongly consistent.

The above result for the consistency is still applicable in the liner model.

Example 2.3. Consider the multiple regression

$$y_t = \sum_{u=1}^{m} \beta_u x_{tu} + \epsilon_t, \quad t = 1, ..., n \tag{2.3}$$

where the ϵ_t are random errors, β_u $(u = 1, ..., m)$ are unknown parameters, and y_t are observed response corresponding to the design vector $x_t = (x_{t1}, ..., x_{tm})'$. In this case, the assumptions of Theorem 2.1 are straightforward and if the input vector $\{x_t\}$ are chosen as a realization of a random sample from a uniform (say) $m-$variate distribution whose support is a subset of $\Xi = \{x \in R^m$: each component $\geq c > 0\}$ then the strong consistency of LAD follows.

3. ASYMPTOTIC NORMALITY

In this section we seek conditions on regression function and input values which ensure the asymptotic distribution of nonlinear LAD $\hat{\theta}_n$.

The following notation is used: $f_t(\theta) = f(x_t, \theta)$, $f_t'(\theta) = (\frac{\partial}{\partial \theta_j} f_t(\theta))_{j=1,\dots,p}$ and $f_t''(\theta) = [\frac{\partial^2}{\partial\theta_j \partial\theta_k} f_t(\theta)]_{j,k=1,\dots,p}$.

We will require the following additional Assumption C :

$C1$: $f_t'(\theta)$ and $f_t''(\theta)$ exist for all θ near θ_o.

$C2$: $\frac{1}{n}\sum_{t=1}^n f_t'(\theta_o) f_t'(\theta_o)^T$ converges to a positive definite matrix $V(\theta_o)$ as $n \to \infty$ where T denotes transpose.

The following theorem provides conditions for the asymptotic normality of LAD.

Theorem 3.1. Let $\hat{\theta}_n$ be a strongly consistent LAD of θ_o under the model (1.1). Then under Assumption A, B and C, $\sqrt{n}(\hat{\theta}_n - \theta_o)$ converges in distribution to a p-variate normal random vector with mean zero and variance-covariance matrix $V(\theta_o)^{-1}/[2g(0)]^2$.

Proof. Let be $D_n(\theta) = D_n(y, \theta)$ simply and let $\rho_n(x)$ be a smooth function such that as $n \to \infty$, $\rho_n(x) \to |x|$ and $|\tilde{D}_n(\theta) - D_n(\theta)|$ converges to zero a.s. unifomly in θ where $\tilde{D}_n(\theta) = \frac{1}{n}\sum_{t=1}^n \rho_n(y_t - f_t(\theta))$. As such function we use

$$\rho_n(x) = \frac{1 + (\beta_n x)^2}{2\beta_n} I_{[|x| \leq \frac{1}{\beta_n}]}(x) + |x| I_{[|x| > \frac{1}{\beta_n}]}(x),$$

where β_n is a function of n which is of the same order as n, and I_E is an indicator function on E, then clearly $\rho_n(x) \to |x|$ and $n|\tilde{D}_n(\theta) - D_n(\theta)| \leq \sum_{t=1}^n \frac{1}{\beta_n} I_{[|r_t(\theta)| \leq \frac{1}{\beta_n}]} \to 0$ a.s. uniformly in θ by the Kolmogorov's SLLN, where $r_t(\theta) = y_t - f_t(\theta)$.

Let $\tilde{\theta}_n$ be a minimizer of $\tilde{D}_n(\theta)$. Note that $\sqrt{n}(\hat{\theta}_n - \theta_o) = \sqrt{n}(\hat{\theta}_n - \tilde{\theta}_n) + \sqrt{n}(\tilde{\theta}_n - \theta_o)$. The theorem will be proved if we show that

(i). $\sqrt{n}(\hat{\theta}_n - \tilde{\theta}_n)$ converges to zero in probability.

(ii). $\sqrt{n}(\tilde{\theta}_n - \theta_o)$ converges to the p-variate normal distribution with zero vector mean, and variance-covariance matrix $V(\theta_o)^{-1}/[2g(0)]^2$.

For (i), note that from the Taylor formula, there exists an $\eta_n \in [0, 1]$ such that

$$\tilde{D}_n(\hat{\theta}_n) - \tilde{D}_n(\tilde{\theta}_n) = \tilde{D}_n'(\tilde{\theta}_n)(\hat{\theta}_n - \tilde{\theta}_n) + \frac{1}{2}(\hat{\theta}_n - \tilde{\theta}_n)^T \tilde{D}_n''(\bar{\theta}_n)(\hat{\theta}_n - \tilde{\theta}_n)$$

where $\bar{\theta}_n = (1-\eta_n)\tilde{\theta}_n + \eta_n\hat{\theta}_n$ which is measurable. Since $\tilde{\theta}_n$ is in the interior of Θ eventually, $\tilde{D}'_n(\tilde{\theta}_n) = 0$. Writing λ_n for the smallest eigenvalue of $\tilde{D}''_n(\bar{\theta}_n)$, we obtain

$$n(\hat{\theta}_n - \tilde{\theta}_n)^T(\hat{\theta}_n - \tilde{\theta}_n) \leq \frac{2n}{\lambda_n}|\tilde{D}_n(\hat{\theta}_n) - \tilde{D}_n(\tilde{\theta}_n)|$$

which converges to zero as $n \to \infty$. This result follows, because

$$n|\tilde{D}_n(\hat{\theta}_n) - \tilde{D}_n(\tilde{\theta}_n)| \leq n|\tilde{D}_n(\hat{\theta}_n) - D_n(\hat{\theta}_n)| + n|\tilde{D}_n(\tilde{\theta}_n) - D_n(\tilde{\theta}_n)|,$$

from the facts that $n|\tilde{D}_n(\theta) - D_n(\theta)| \to 0$ a.s. uniformly in θ, both $\hat{\theta}_n$ and $\tilde{\theta}_n$ are strongly consistent, and $\lambda_n \to \lambda_o$ which the smallest but postive eigenvalue of $2g(0)V(\theta_o)$ because of $\tilde{D}''_n(\hat{\theta}_n) \to 2g(0)V(\theta_o)$ which will be proved in the part (ii). In this case, the strong consistency of $\tilde{\theta}_n$ is due to $\tilde{D}_n(\theta) - D_n(\theta) \to 0$ a.s. uniformly in θ. It follows that $\sqrt{n}(\hat{\theta}_n - \tilde{\theta}_n)$ converges in probability to zero.

To prove (ii), note that the first two derivatives of $\tilde{D}_n(\theta)$ are

$$\tilde{D}'_n(\theta) = -\frac{\beta_n}{n}\sum_{t=1}^{n} f'_t(\theta)r_t(\theta)I_{E_{t,n}} - \frac{1}{n}\sum_{t=1}^{n} f'_t(\theta)Sign(r_t(\theta))I_{E^c_{t,n}},$$

and

$$\tilde{D}''_n(\theta) = -\frac{\beta_n}{n}\sum_{t=1}^{n} f''_t(\theta)r_t(\theta)I_{E_{t,n}} - \frac{1}{n}\sum_{t=1}^{n} f''_t(\theta)Sign(r_t(\theta))I_{E^c_{t,n}} + \frac{\beta_n}{n}\sum_{t=1}^{n} f'_t(\theta)f'_t(\theta)^T I_{E_{t,n}} \qquad (3.2)$$

where $E_{t,n} = [|r_t(\theta)| \leq \frac{1}{\beta_n}]$. From the mean value theorem and $\tilde{D}'_n(\tilde{\theta}_n) = 0$, we have

$$\tilde{D}'_n(\theta_o) = \tilde{D}''_n(\theta^*_n)(\theta_o - \tilde{\theta}_n)$$

for some θ^*_n on the line segment with end points θ_o and $\tilde{\theta}_n$. Thus, $\sqrt{n}(\tilde{\theta}_n - \theta_o)$ becomes

$$\left[\tilde{D}''_n(\theta^*_n)\right]^{-1}\left\{\frac{\beta_n}{\sqrt{n}}\sum_{t=1}^{n} f'_t(\theta_o)r_t(\theta_o)I_{\bar{E}_{t,n}} + \frac{1}{\sqrt{n}}\sum_{t=1}^{n} f'_t(\theta_o)Sign(r_t(\theta_o))I_{\bar{E}^c_{t,n}}\right\} \qquad (3.3)$$

where $\bar{E}_{t,n} = [|\epsilon_t| \leq \frac{1}{\beta_n}]$. First of all, it can be shown that $\tilde{D}''_n(\theta^*_n) \to 2g(0)V(\theta_o)$ because when $\theta = \theta_o$ the first and second terms of the right-hand side of (3.2) converge to zero matrices pointwisely and in probability respectively, while the third term converges to $2g(0)V(\theta_o)$ under the assumption C. The latter result follows from the fact that under $\theta = \hat{\theta}_n$, $\frac{\beta_n}{n}\sum_{t=1}^n f_t'(\theta)f_t'(\theta)^T I_{E_{t,n}} \to 2g(0)V(\theta_o)$ in probability due to $E[\beta_n I_{E_{t,n}}] \to 2g(0), Var[I_{E_{t,n}}] \leq PE_{t,n} \to 0$ for all t, and Assumption $C2$.

Since the first term in curly brackets in (3.3) converges to zero in law, it therefore suffices to show that the second term, denoted by W_n, converges to normal random vector with zero vector mean and variance-covariance matrix $V(\theta_o)$.

For this we first note that for any nonzero constant vector $\lambda = (\lambda_1, ..., \lambda_p)'$, $\lambda' W_n = \sum_{t=1}^n Z_t^{(o)}$ where

$$Z_t^{(o)} = \frac{1}{\sqrt{n}}\sum_{u=1}^p \lambda_u \frac{\partial f(x_t,\theta_o)}{\partial \theta_u} Sign(r_t(\theta_o)) I_{\bar{E}_{t,n}^c}.$$

Then, since $Z_t^{(o)}$ are independent, it remains to show that the Lindeberg condition of the central limit theorem holds with $EZ_t^{(o)} = 0$ and $\sigma_t^2 = Var Z_t^{(o)}$ equal to

$$\sigma_t^2 = \frac{1}{n}\Big[\sum_{u=1}^p \lambda_u \frac{\partial f(x_t,\theta_o)}{\partial \theta_u}\Big]^2 P[\bar{E}_{t,n}^c]. \tag{3.4}$$

Now, for arbitrary $\epsilon > 0$,

$$\sum_{t=1}^n \int_{[|z|>\epsilon s_n]} z^2 dF_t(z) = \frac{1}{n}\sum_{t=1}^n\Big[\sum_{u=1}^p \lambda_u \frac{\partial f(x_t,\theta_o)}{\partial \theta_u}\Big]^2 P[\bar{E}_{t,n}^c \cap D_{t,n}]$$

where $s_n^2 = \sum_{t=1}^n \sigma_t^2$, $F_t(z)$ is the distribution function of $Z_t^{(o)}$ and

$$D_{t,n} = \Big[w : \frac{1}{\sqrt{n}}\Big|\sum_{u=1}^p \lambda_u \frac{\partial f(x_t,\theta_o)}{\partial \theta_u} I_{\bar{E}_{t,n}^c}(w)\Big| > \epsilon s_n\Big].$$

Thus,

$$\frac{1}{s_n^2}\sum_{t=1}^n \int_{[|z|>\epsilon s_n]} z^2 dF_t(z) \leq \max_{1\leq t\leq n} P[D_{t,n}|\bar{E}_{t,n}^c]$$

which converges to zero. This result follows from the fact that

$$\Big[\sum_{u=1}^p \lambda_u \frac{\partial f(x_t,\theta_o)}{\partial \theta_u}\Big]^2 = \lambda^T\Big[f_t'(\theta_o)f_t'(\theta_o)^T\Big]\lambda, \tag{3.5}$$

and $D_{t,n} = [w | I_{\bar{E}^c_{t,n}}(w) > \epsilon\delta_n]$ with

$$\delta_n^2 = \frac{\lambda^T\left[\frac{1}{n}\sum_{t=1}^{n} f_t'(\theta_o)f_t'(\theta_o)^T\right]\lambda}{\lambda^T\left[\frac{1}{n}f_t'(\theta_o)f_t'(\theta_o)^T\right]\lambda} P[\bar{E}^c_{t,n}]$$

which converges to infinity as n go to infinity, from $C1$ and the boundedness of f. Thus the Lindeberg condition holds.

Furthermore, obviously $E[\lambda' W_n]$ converges to zero, and $Var[\lambda' W_n] = s_n^2 = \sum_{t=1}^{n}\sigma_t^2$ converges, by (3.4) and (3.5), to $\lambda' V(\theta_o)\lambda$ as $n \to \infty$, so that W_n converges to normal distribution with zero mean vector and variance-covariance vector $V(\theta_o)$. Hence, the desired limiting distribution of $\sqrt{n}(\tilde{\theta}_n - \theta_o)$ is obtained. This completes the proof.

The strongly consistenct estimators of Example 2.1 and 2.2 are also asymptocally normal since Assumption C is easily satisfied for each case.

In the linear model, Assumption A, B and $C1$ of C are straightforward. Also, Assumption $C2$ of C reduces to the condition that there exists a positive definite matrix Σ such that as $n \to \infty$

$$\frac{1}{n}X'X \to \Sigma, \tag{3.6}$$

where X is the design matrix of the model (2.3). Moreover, the condition (3.6) is implied by $B2$:

$$\frac{1}{n}X'X = \left[\frac{1}{n}\sum_{t=1}^{n} x_{tu}x_{tv}\right]_{u,v=1,\ldots,p} \to \quad \Sigma^* = \left[\int_{\Xi} x_u x_v d\mu(x)\right]_{u,v=1,\ldots,p}.$$

In this case, the matrix Σ^* is positive definite since for any nonzero $\lambda = (\lambda_1, ..., \lambda_p)'$ in R^p, $\lambda'\Sigma^*\lambda = \int_{\Xi}\lambda'[x^* x^{*\prime}]\lambda d\mu(x) > 0$ if X is of full rank, where $x^* = (x_1, ..., x_p)'$. Note [Drygas (1976)] that the (3.6) is a sufficient condition for the strong consistency of the LSE.

4. CONFIDENCE REGION AND ASYMPTOTIC RELATIVE EFFICIENCY

In this section we shall provide approximate confidence region for the parameter θ in the model (1.1), based on the large-sample normality of the LAD, and consider the asymptotic relative efficiency of the proposed region relative to the classical result derived from the limiting distribution of the LSE, using the ARE of two estimators.

The asymptotic normality of $\sqrt{n}(\hat{\theta}_n - \theta_o)$, derived in Theorem 3.1, under the regularity conditions, suggests the use of the pivotal quantity of the form

$$Q_n(y,\theta) = \frac{n}{[2g(0)]^2}(\hat{\theta}_n - \theta)V_n(\hat{\theta}_n)(\hat{\theta}_n - \theta)$$

where $V_n(\theta)$ is the $p \times p$ matrix with $(u,v)th$ element $\frac{1}{n}\sum_{t=1}^{n} f_t'(\theta)f_t'(\theta)^T$.

The following theorem gives the large sample distribution of $Q_n(\theta)$.

Theorem 4.1. Under the conditions of Theorem 3.1, $Q_n(\theta)$ has asymptotically a chi-square distribution with p degree of freedom.

Proof. Theorem 4.1 follows immediately from Theorem 3.1.

By reference to the limiting distribution of $Q_n(y,\theta_o)$, we define $C_{1-\alpha}(\theta)$ as the set of θ such that

$$(\hat{\theta}_n - \theta)V_n(\hat{\theta}_n)(\hat{\theta}_n - \theta) \leq \delta$$

where δ is $\frac{[2g(0)]^2}{n}\chi^2_{1-\alpha}(p)$ and $\chi^2_{1-\alpha}(p)$ is the $(1-\alpha)th$ the quantile of the chi-square distribution. Then, for n large $C_{1-\alpha}(\theta)$ provides a $100(1-\alpha)$ percent confidence region for θ.

Note that it is known [e.g., Jennrich (1976)] that under certain regularity conditions, the sequence of the LSE's $\check{\theta}_n$ has asymptotically a normal distribution in the sense that

$$\sqrt{n}(\hat{\theta}_n - \theta_o) \xrightarrow{L} N_p(0, \sigma^2 V(\theta_o))$$

where σ^2 is the common variance of errors in the model (1.1).

Thus, a $100(1-\alpha)$ percent confidence region based on the LSE, denoted by $C^*_{1-\alpha}(\theta)$, is the set of θ such that

$$(\check{\theta}_n - \theta)V_n(\check{\theta}_n)(\check{\theta}_n - \theta) \leq \delta^*$$

where δ is $\frac{\sigma^2}{n}\chi^2_{1-\alpha}(p)$ if σ^2 is known and $ps^2F_{1-\alpha}(p,\nu)$ if σ^2 is estimated independently using ν degrees of freedom.

If we define the asymptotic relative efficiency of $\{\tilde{\theta}_n\}$ with respect to the classical least squares estimator $\{\check{\theta}_n\}$ on the inverse ratio of their generalized limiting variances, and denote it by $e(A, S)$, which implies strictly smaller asymptotic confidence region, then we have

$$e(A, S) = \frac{\sigma^2}{[2g(0)]^2}$$

which coincides with the ratio of the variances of sample median and mean from the error distribution $G(x)$.

The above result implies that the LAD is relatively more efficient than LSE in the nonlinear model for any error distibution for which sample median is more efficient than the sample mean as an estimator of location, i.e., the heavy-tailed distributions and/or the distributions which have peaked density at the median, such as Cauchy, double-exponential, logistic distribution etc. This result also implies that the LAD has the strictly smaller asymptotic confidence regions than in LSE.

REFERENCES

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

(2) Drygas, H. (1976). Weak and strong consistency of the least square estimates in the regression models, *Z. Wahrscheinlickeitstheorie verw. Gebiete*, **34**, 119-127.

(3) Gonin, R. and Money, A.H. (1985). Nonlinear L_p-norm estimation: Part I - On the choice of the exponent, p, where the errors are additive, *Communications in Statistics: Theory and Method*, **14** , 827-840.

(4) Hampel, F.R. (1973). Robust Estimation: A Condensed Partial Survey, *Z. Wahrscheinlickeitstheorie verw. Gebiete*, **27**, 87-104.

(5) Huber, P.J. (1972). Robust Statistics: A Review, *Annals of Mathematical Statistics* , **43** , 1041-1067.

(6) Jennrich, R. I. (1969). Asymptotic properties of non-linear least squares estimators, *Annals of Mathematical Statistics*, **40**, No.2, 633-643.

(7) Malinvaud, E. (1970). The consistency of nonlinear regression, *Annals of Mathematical Statistics*, **41** , 956-969.

(8) Nyquist, H. (1983). The optimal L_p-norm estimator in linear regression models, *Communications in Statistics: Theory and Method*, **12**, 2511-2524.

(9) Oberhofer, W. (1982). The consistency of nonlinear regression minimizing the L_1−norm, *Annals of Statistics*, **10**, No.1. 316-319.

(10) Richardson, G. D. and Bhattacharyya, B. B. (1987). Consistent L_1-estimators in nonlinear regression for a noncompact parameter space, *Sankhy ā*, **49**, Series A, Pt.3, 377-387.