Asymptotic Theory of Least Absolute Error Regression

GILBERT BASSETT, JR. and ROGER KOENKER*

In the general linear model with independent and identically distributed errors and distribution function F, the estimator which minimizes the sum of absolute residuals is demonstrated to be consistent and asymptotically Gaussian with covariance matrix $\omega^2 Q^{-1}$, where $Q = \lim T^{-1}X'X$ and ω^2 is the asymptotic variance of the ordinary sample median from samples with distribution F. Thus the least absolute error estimator has strictly smaller asymptotic confidence ellipsoids than the least squares estimator for linear models from any F for which the sample median is a more efficient estimator of location than the sample mean.

KEY WORDS: Least absolute error estimators; Linear models; Asymptotic distribution theory.

1. INTRODUCTION

The methods of minimizing the sum of absolute and squared deviations from hypothesized linear models have vied for statistical favor for more than 250 years. While least squares enjoys certain well-known optimality properties within strictly Gaussian parametric models, the least absolute error (LAE) estimator[1] is a widely recognized superior *robust* method especially well-suited to longer-tailed error distributions. Increasingly, the LAE estimator is recommended as a preliminary (consistent) estimator for one-step and iteratively reweighted least squares procedures; cf. Andrews (1974), Bickel (1975), Harvey (1977), and Hill and Holland (1977).

This article resolves a long-standing, open question concerning the LAE estimator by establishing its asymptotic normality under general conditions, thereby extending a result of Laplace (1818) (see Stigler 1973) to the general linear model. The result confirms that the LAE estimator is a natural analog of the sample median for the general linear model. Recently we have extended this analogy to the estimation of arbitrary quantile hyperplanes for linear models; see Koenker and Bassett (1978).

We consider the familiar problem of estimating a K-dimensional vector of unknown parameters $\boldsymbol{\beta}$ from a sample of independent observations on random variables $Y_1, Y_2, \ldots, Y_T$ distributed according to

$$\Pr[Y_t < y] = F(y - \sum_{k=1}^{K} x_{kt}\beta_k) , \quad t = 1, \ldots, T , \quad (1.1)$$

where x_{kt} denotes an element of a known $T \times K$ design matrix $\mathbf{X}_T$. We assume that $\boldsymbol{\beta}$ is located so that the distribution function F has median zero; if the column space of $\mathbf{X}_T$ contains $\mathbf{1}_T = (1, 1, \ldots, 1)$, then this involves no loss in generality. The LAE estimator $\boldsymbol{\beta}_T^*$ is a solution to the problem

$$\min_{\mathbf{b} \in \mathbb{R}^K} [\rho(\mathbf{b}) = \sum_{t=1}^{I} |y_t - \sum_{k=1}^{K} x_{kt}b_k|] . \quad (1.2)$$

We prove the following theorem in Section 3.

Theorem: Let $\{\boldsymbol{\beta}_T^*\}$ denote a sequence of unique solutions to the problem (1.2) for model (1.1), and assume

(i) F is continuous and has continuous and positive density f at the median, and
(ii) $\lim T^{-1}\mathbf{X}_T'\mathbf{X}_T = \mathbf{Q}$, a positive definite matrix.

Then $\sqrt{T}(\boldsymbol{\beta}_T^* - \boldsymbol{\beta})$ converges in distribution to a K-dimensional Gaussian random vector with mean $\mathbf{0}$ and covariance matrix $\omega^2\mathbf{Q}^{-1}$, where ω^2 is the asymptotic variance of the sample median from random samples from distribution F; i.e., $\omega = [2f(0)]^{-1}$.

The result implies that for any error distribution for which the median is superior to the mean as an estimator of location, the LAE estimator is preferable to the least squares estimator in the general linear model, in the sense of having strictly smaller asymptotic confidence ellipsoids.[2] This condition holds for an enormous class of distributions which have peaked density at the median and/or long tails.

Laplace (1818) proved the preceding result for the special case of bivariate regression through the origin,

* Gilbert Bassett, Jr. is Assistant Professor of Economics, University of Illinois at Chicago Circle, Chicago, IL 60680. Roger Koenker is a member of the technical staff, Bell Laboratories, Murray Hill, NJ 07974. The authors wish to express their appreciation to Lester D. Taylor for stimulating their interest in this topic.

[1] This venerable method, which Laplace called the "method of situations," has had a bewildering variety of names. Recently it has accumulated a large array of acronyms: LAE, LAR, LAD, MAD, MSAE, and others. It is also frequently referred to as L_1-regression, and less frequently as median regression.

[2] A number of recent studies have sought to investigate the sampling distribution of the LAE estimator by Monte Carlo methods. Most notable among these is the unpublished work of Rosenberg and Carlson (1971), who conclude that the Gaussian distribution given above provides an acceptable approximation to their sampling distributions for modest sample sizes and well-conditioned designs. Their finding that the approximation was significantly worse for these sample sizes and ill-conditioned designs should come as no surprise.

© Journal of the American Statistical Association
September 1978, Volume 73, Number 363
Theory and Methods Section

which in turn includes the median as a special case. Since $\boldsymbol{\beta}^*$ may be viewed as a limiting form of the well-known Huber (1972, 1973) M-estimator defined by

$$\min_{\mathbf{b}\in\mathbb{R}^K} \sum_{t=1}^{T} \rho(y_t - \sum_{k=1}^{K} x_{tk}b_k) \ , \tag{1.3}$$

where

$$\begin{aligned}\rho(u) &= \tfrac{1}{2}u^2 \qquad |u| < c \ , \\ &= c|u| - \tfrac{1}{2}c^2 \quad |u| \geq c \ ,\end{aligned} \tag{1.4}$$

our result may be interpreted as an extension of well-known results on the large-sample theory of M-estimators to an important nonregular special case. General methods of proof of asymptotic results on M-estimators for linear models break down when $\rho(t) = |t|$ because differentiability conditions are violated. However, note that the familiar asymptotic variance formula for M-estimators (see, e.g., Huber 1973, Relles 1968, Yohai 1974),

$$V(\psi, F) = \int_{-\infty}^{\infty} \psi^2(t) f(t) dt \Big/ \left[\int_{-\infty}^{\infty} \psi'(t) f(t) dt \right]^2 ,$$

equals $[2f(0)]^{-2}$ for $\psi(t) = \text{sgn}(t)$ with $\psi'(t)$ interpreted as twice the Dirac delta function.

2. NOTATION AND PRELIMINARY RESULTS

Let $\mathcal{T} = \{1, 2, \ldots, T\}$ and let $\mathcal{H}$ denote the set of K-element subsets of $\mathcal{T}$. Elements $h \in \mathcal{H}$ have relative complement $\bar{h} = \mathcal{T} - h$, and serve to partition $\mathbf{y}$ and $\mathbf{X}$. Thus $\mathbf{y}(h)$ denotes the K-vector with elements $\{y_t : t \in h\}$, while $\mathbf{X}(\bar{h})$ denotes a $(T - K) \times K$ matrix with rows $\{\mathbf{x}_t : t \in \bar{h}\}$. Finally, let

$$H = \{h \in \mathcal{H} | \text{rank } \mathbf{X}(h) = K\} \ .$$

We may now state some fundamental properties of elements $\boldsymbol{\beta}^*$ of the solution set $\mathbf{B}^*$ of the minimization problem (1.2).

Lemma 1: If $\mathbf{X}$ has rank K, then the solution set $\mathbf{B}^*$ to (1.2) has at least one element of the form

$$\boldsymbol{\beta}^* = \mathbf{X}(h)^{-1}\mathbf{y}(h)$$

for some $h \in H$. Moreover, $\mathbf{B}^*$ is the convex hull of all solutions having this form.

Proof: This result follows immediately from the linear programming formulation of problem (1.1). See, for example, the recent work by Abdelmalek (1974), Bassett (1973), and Taylor (1974) for details.

Remark: One introductory econometrics text asserts that the LAE estimator "ignores sample information" because it "passes through" K sample points. However, the entire sample serves to determine *which* K points the $\boldsymbol{\beta}^*$ hyperplane passes through. Also, the fact that $\boldsymbol{\beta}^*$ is a linear function of *some* of the sample observations has apparently led some authors to state that $\boldsymbol{\beta}^*$ must be inferior to least squares under Gauss–Markov conditions. However, the selection of h such that $\boldsymbol{\beta}^* = \mathbf{X}(h)^{-1}\mathbf{y}(h)$ makes $\boldsymbol{\beta}^*$ a manifestly nonlinear estimator; thus no Gauss–Markov claims apply to it. Indeed, there is abundant evidence that $\boldsymbol{\beta}^*$ outperforms least squares for a large class of distributions satisfying Gauss–Markov conditions. For examples we need go no further than the comparison of the sample median and sample mean as estimates of location.

We now introduce some additional notation and state two more lemmas. For $\mathbf{v} \in \mathbb{R}^K$ let $\|\mathbf{v}\| = \max_{k=1,\ldots,K} |v_k|$ and consider the K-dimensional closed hypercubes centered at $\boldsymbol{\delta}$ of "radius" ϵ,

$$C[\boldsymbol{\delta}, \epsilon] = \{\mathbf{c} \in \mathbb{R}^K : \|\mathbf{c} - \boldsymbol{\delta}\| \leq \epsilon\} \ . \tag{2.1}$$

The corresponding open cubes will be denoted by $C(\boldsymbol{\delta}, \epsilon)$. Our characterization of the estimator $\boldsymbol{\beta}^*$ relies extensively on the vector-valued function

$$\begin{aligned}\boldsymbol{\zeta}(h, \mathbf{v}) = \sum_{t\in\bar{h}} \boldsymbol{\zeta}_t(h, \mathbf{v}) = \sum_{t\in\bar{h}} \text{sgn}^*(y_t - \mathbf{x}_t\mathbf{b}(h); \\ - \mathbf{x}_t\mathbf{X}(h)^{-1}\mathbf{v})\mathbf{x}_t\mathbf{X}(h)^{-1} \ ,\end{aligned} \tag{2.2}$$

where

$$\begin{aligned}\text{sgn}^*(u; w) &= \text{sgn}u \quad \text{if} \quad u \neq 0 \ ; \\ &= \text{sgn}w \quad \text{otherwise} \ .\end{aligned}$$

Lemma 2: For any $h \in H$, $\mathbf{b}(h) \equiv \mathbf{X}(h)^{-1}\mathbf{y}(h) \in \mathbf{B}^*$ if and only if $\boldsymbol{\zeta}(h, \mathbf{v}) \in C[\mathbf{0}, 1]$ for all $\mathbf{v} \neq \mathbf{0}$, and $\mathbf{b}(h) = \mathbf{B}^*$ (is unique) if and only if $\boldsymbol{\zeta}(h, \mathbf{v}) \in C(\mathbf{0}, 1)$ for all $\mathbf{v} \neq \mathbf{0}$.

Proof: Since $\rho(\cdot)$ is convex, it suffices to show that our existence and uniqueness conditions are equivalent to nonnegativity and strict positivity of the directional derivative function,

$$\psi(\mathbf{b}(h); \mathbf{w}) = -\sum_{t=1}^{T} \text{sgn}^*(y_t - \mathbf{x}_t\mathbf{b}(h); -\mathbf{x}_t\mathbf{w})\mathbf{x}_t\mathbf{w} \ ,$$

in all directions $\mathbf{w} \neq \mathbf{0}$. Now,

$$\begin{aligned}\psi(\mathbf{b}(h); \mathbf{w}) = \sum_{t\in h} \text{sgn}(\mathbf{x}_t\mathbf{w})\mathbf{x}_t\mathbf{w} \\ - \sum_{t\in\bar{h}} \text{sgn}^*(y_t - \mathbf{x}_t\mathbf{b}(h); \mathbf{x}_t\mathbf{w})\mathbf{x}_t\mathbf{w} \geq 0\end{aligned}$$

for all $\mathbf{w} \neq \mathbf{0}$ is equivalent to, setting $\mathbf{v} = \mathbf{X}(h)\mathbf{w}$,

$$\sum_{k=1}^{K} |v_k| - \boldsymbol{\zeta}(h, \mathbf{v})\mathbf{v} \geq 0 \tag{2.3}$$

for all $\mathbf{v} \neq \mathbf{0}$. But (2.3) is equivalent to $\boldsymbol{\zeta}(h, \mathbf{v}) \in C[\mathbf{0}, 1]$. Uniqueness is argued in the same way, with strict inequalities replacing weak.

Remarks: Note that if $y_t - \mathbf{x}_t\mathbf{b}(h) \neq 0$ for all $t \in \bar{h}$, then $\boldsymbol{\zeta}(h, \mathbf{v})$ is independent of $\mathbf{v}$. This is a nondegeneracy or "no ties" condition for the general linear model.

It is instructive to consider the sample median in the light of Lemmas 1 and 2. Suppose that $x_t = 1$ for $t = 1, \ldots, T$, so $H = \mathcal{T}$. Then $\beta^* = y(h)$ is a sample median if and only if

$$-1 \leq \sum_{t\in\bar{h}} \text{sgn}^*(y_t - y(h); w) \leq 1$$

for all $w \neq 0$. If F is continuous, so that $y_t - y(h) = 0$ for $t \in \bar{h}$ with probability zero, then a unique median obtains if and only if T is odd. With continuous F, and

T even, there is, with probability one, an "interval of medians" between adjacent order statistics. Note that in the absence of degeneracy, the condition for uniqueness is purely a design condition, reducing in the location model to the requirement that T be odd. This suggests that for any sequence $\{\mathbf{X}_T\}$ of designs, one should be able to extract a subsequence, or at worst a "perturbed" subsequence, whose elements have unique solutions. An alternative approach which is frequently employed in the location model is to adopt some arbitrary rule to choose a single element from sets of solutions to problem (1.2) when they occur. Either approach suffices to obtain the sequence of unique solutions considered in the next section.

We now state a number of equivariance properties of the LAE estimator.

Lemma 3: If $\boldsymbol{\beta}^*(\mathbf{y}, \mathbf{X}) \in \mathbf{B}^*(\mathbf{y}, \mathbf{X})$, then the following are elements of the solution of the specified transformed problem:

(i) $\boldsymbol{\beta}^*(\lambda\mathbf{y}, \mathbf{X}) = \lambda\boldsymbol{\beta}^*(\mathbf{y}, \mathbf{X}), \quad \lambda \in \mathbb{R}$;

(ii) $\boldsymbol{\beta}^*(\mathbf{y} + \mathbf{X}\boldsymbol{\gamma}, \mathbf{X}) = \boldsymbol{\beta}^*(\mathbf{y}, \mathbf{X}) + \boldsymbol{\gamma}, \quad \boldsymbol{\gamma} \in \mathbb{R}^K$;

(iii) $\boldsymbol{\beta}^*(\mathbf{y}, \mathbf{XA}) = \mathbf{A}^{-1}\boldsymbol{\beta}^*(\mathbf{y}, \mathbf{X}), \quad \mathbf{A}_{K\times K}$ nonsingular;

(iv) $\boldsymbol{\beta}^*(\mathbf{X}\boldsymbol{\beta}^* + \mathbf{D}\mathbf{u}^*, \mathbf{X}) = \boldsymbol{\beta}^*(\mathbf{y}, \mathbf{X})$, $\mathbf{D}_{T\times T}$ diagonal with nonnegative elements; $\mathbf{u}^* \equiv \mathbf{y} - \mathbf{X}\boldsymbol{\beta}^*$.

Proof: Let

$$\psi(\mathbf{b}; \mathbf{y}, \mathbf{X}) = \sum_{t=1}^{T} |y_t - \mathbf{x}_t\mathbf{b}|$$

and note that

(i) $|\lambda|\psi(\mathbf{b}; \mathbf{y}, \mathbf{X}) = \psi(\lambda\mathbf{b}; \lambda\mathbf{y}, \mathbf{X})$;

(ii) $\psi(\mathbf{b}; \mathbf{y}, \mathbf{X}) = \psi(\mathbf{b} + \boldsymbol{\gamma}; \mathbf{y} + \mathbf{X}\boldsymbol{\gamma}, \mathbf{X})$;

(iii) $\psi(\mathbf{b}; \mathbf{y}, \mathbf{X}) = \psi(\mathbf{A}^{-1}\mathbf{b}; \mathbf{y}, \mathbf{XA})$.

For (iv), $\boldsymbol{\beta}^* \in \mathbf{B}^*(\mathbf{y}, \mathbf{X})$ implies (by Lemma 2);

$$-\sum_{t=1}^{T} \operatorname{sgn}^*(y_t - \mathbf{x}_t\boldsymbol{\beta}^*; -\mathbf{x}_t\mathbf{w})\mathbf{x}_t\mathbf{w} \geq 0, \quad \mathbf{w} \in \mathbb{R}^K .$$

Note that

$$\operatorname{sgn}^*(\mathbf{x}_t\boldsymbol{\beta}^* + d_t(y_t - \mathbf{x}_t\boldsymbol{\beta}^*) - \mathbf{x}_t\boldsymbol{\beta}^*; -\mathbf{x}_t\mathbf{w})\mathbf{x}_t\mathbf{w} \leq \operatorname{sgn}^*(y_t - \mathbf{x}_t\boldsymbol{\beta}^*; -\mathbf{x}_t\mathbf{w})\mathbf{x}_t\mathbf{w}$$

for $d_t \geq 0$, and (iv) follows.

Remark: Estimators with properties (i) and (ii) are termed affine equivariant, and scale and shift equivariant, respectively (see Bickel 1975). Estimators with property (iii) may be termed equivariant to reparameterization of design. Properties (i)–(iii) are shared by the least squares estimator, but typically robust alternatives to least squares are not equivariant in one or more of the above senses; see, e.g., Bickel (1973, 1975).

The fourth property generalizes an invariance property of the median to the linear model. It has the following geometric interpretation. Imagine a scatter of sample observations in $\mathbb{R}^2$ with the LAE solution line slicing through the scatter. Now consider the effect (on the position of the LAE line) of moving observations up or down in the scatter. The result states that as long as these movements leave observations on the same side of the original line, the solution is unaffected. This property is not shared by least squares, and although obvious in the case of the median, it seems to capture part of the intuitive flavor of $\boldsymbol{\beta}^*$'s median-type robustness and insensitivity to outlying observations.

3. PROOF OF THE THEOREM

We begin by establishing that the finite sample density $\phi_T(\boldsymbol{\delta})$ of the random K-vector $\sqrt{T}(\boldsymbol{\beta}_T^* - \boldsymbol{\beta})$ is given by

$$\phi_T(\boldsymbol{\delta}) = T^{-K/2} \sum_{h\in H} |\mathbf{X}(h)| \prod_{t\in h} f(T^{-\frac{1}{2}}\mathbf{x}_t\boldsymbol{\delta}) \cdot \Pr[\mathbf{Z}_T(\boldsymbol{\delta}, h) \in C(\mathbf{0}, 1)] , \quad (3.1)$$

where

$$\mathbf{Z}_T(\boldsymbol{\delta}, h) = \sum_{t\in\bar{h}} \mathbf{z}_t(\boldsymbol{\delta}, h) = \sum_{t\in\bar{h}} \operatorname{sgn}(u_t - T^{-\frac{1}{2}}\mathbf{x}_t\boldsymbol{\delta})\mathbf{x}_t\mathbf{X}(h)^{-1} .$$

We argue as follows. Let $\boldsymbol{\delta}^* = \boldsymbol{\beta}^* - \boldsymbol{\beta}$ and $\mathbf{d}(h) = \mathbf{b}(h) - \boldsymbol{\beta} = \mathbf{X}(h)^{-1}\mathbf{u}(h)$, and note that $y_t - \mathbf{x}_t\mathbf{b}(h) = \mathbf{u}_t - \mathbf{x}_t\mathbf{d}(h)$. Consider the events

$$E_1(h, \boldsymbol{\delta}, \epsilon) = \{\mathbf{u} \in \mathbb{R}^T | \mathbf{d}(h) \in C(\boldsymbol{\delta}, \epsilon)\} ,$$

$$E_2(h) = \{\mathbf{u} \in \mathbb{R}^T | \boldsymbol{\zeta}(h, \mathbf{v}) \in C(\mathbf{0}, 1), \forall\, \mathbf{v} \neq \mathbf{0}\} .$$

By Lemmas 1 and 2,

$$\Pr[\boldsymbol{\delta}^* \in C(\boldsymbol{\delta}, \epsilon)] = \bigcup_{h\in H} [E_1(h, \boldsymbol{\delta}, \epsilon) \cap E_2(h)] . \quad (3.2)$$

Let

$$\|\mathbf{X}_T\| = \max_{\substack{k=1,\ldots,K \\ t=1,\ldots,T}} |x_{kt}| ,$$

set $M_T = K\|\mathbf{X}_T\|$, and define the event

$$E_3(h, \boldsymbol{\delta}, \epsilon) = \{\mathbf{u} \in \mathbb{R}^T: |u_t - \mathbf{x}_t\boldsymbol{\delta}| > \epsilon M \text{ for all } t \in \bar{h}\} .$$

Clearly,

$$\Pr(E_1 \cap E_2) = \Pr(E_1 \cap E_2 \cap E_3) + \Pr(E_1 \cap E_2 \cap \sim E_3) . \quad (3.3)$$

Since $E_1(h, \boldsymbol{\delta}, \epsilon)$ implies $|u_t - \mathbf{x}_t\boldsymbol{\delta}| < \epsilon M$ for all $t \in h$,

$$\Pr[\bigcup_{h\in H} (E_1 \cap E_2 \cap E_3)] = \sum_{h\in H} \Pr(E_1 \cap E_2 \cap E_3) = \sum_{h\in H} \Pr(E_1) \Pr(E_2 | E_1 \cap E_3) \Pr(E_3 | E_1) . \quad (3.4)$$

Set $E_2'(h, \delta) = \{\mathbf{u} \in \mathbb{R}^T | \mathbf{Z}(\sqrt{T}\boldsymbol{\delta}, h) \in C(\mathbf{0}, 1)\}$. It is readily shown that $\Pr(E_2') = \Pr(E_2 | E_1 \cap E_3)$. If we denote the Lebesgue measure of C by $\lambda\{C\}$, then since $\lim_{\epsilon\to 0} \Pr(E_3(h, \boldsymbol{\delta}, \epsilon)) = 1$, we have:

$$\lim_{\epsilon\to 0} \frac{\Pr[\boldsymbol{\delta}^* \in C(\boldsymbol{\delta}, \epsilon)]}{\lambda\{C(\boldsymbol{\delta}, \epsilon)\}} = \lim_{\epsilon\to 0} \frac{\Pr[E_1(h, \boldsymbol{\delta}, \epsilon)]}{\lambda\{C(\boldsymbol{\delta}, \epsilon)\}} \Pr(E_2') = |\mathbf{X}(h)| \prod_{t\in h} f(\mathbf{x}_t\boldsymbol{\delta}) \Pr(E_2') . \quad (3.5)$$

Normalizing by $T^{-\frac{1}{2}}$ yields (3.1).

We now demonstrate that $g_T(\boldsymbol{\delta})$ converges to a specified Gaussian density, and Scheffé's Theorem (1947) on convergence of densities completes the proof.

Note that

$$\begin{aligned} \mathbf{z}_t(\boldsymbol{\delta}, h) &= \mathbf{x}_t\mathbf{X}(h)^{-1} \text{ with probability } 1 - F(T^{-\frac{1}{2}}\mathbf{x}_t\boldsymbol{\delta}) \ , \\ &= -\mathbf{x}_t\mathbf{X}(h)^{-1} \text{ with probability } F(T^{-\frac{1}{2}}\mathbf{x}_t\boldsymbol{\delta}) \ . \end{aligned} \quad (3.6)$$

Expanding F around zero, noting that $T^{-1}\sum_{t\epsilon\bar{h}} \mathbf{x}_t'\mathbf{x}_t \rightarrow Q$, and noting that $\|\mathbf{X}_T\| = o(\sqrt{T})$ (see Malinvaud 1970, pp. 226–227), it is readily shown that the stabilized sum

$$T^{-\frac{1}{2}}\mathbf{Z}_T(\boldsymbol{\delta}, h) = T^{-\frac{1}{2}} \sum_{t\in\bar{h}} \mathbf{z}_t(\boldsymbol{\delta}, h) \quad (3.7)$$

converges in law to a K-variate Gaussian random vector with mean $-2f(0)\boldsymbol{\delta}'\mathbf{Q}\mathbf{X}(h)^{-1}$ and covariance matrix $\mathbf{X}'(h)^{-1}\mathbf{Q}\mathbf{X}(h)^{-1}$. Note that the condition $\|\mathbf{X}_T\| = o(\sqrt{T})$ implies the standard multivariate Lindeberg condition (e.g., Cramér 1970, p. 114) immediately.

Suppressing $\boldsymbol{\delta}$ and h, let G_T denote the probability measure induced on $\mathfrak{R}^K$ by $\tilde{\mathbf{Z}}_T = \frac{1}{2}T^{-\frac{1}{2}}\mathbf{Z}_T$, and let $G_T \rightarrow G$. Define the K-dimensional hypercubes centered at the origin,

$$C_T = C[\mathbf{0}, 1/(2\sqrt{T})] = \{\mathbf{c} \in \mathfrak{R}^K : \|\mathbf{c}\| < 1/(2\sqrt{T})\} \ , \quad (3.8)$$

with Lebesgue measure $\lambda\{C_T\} = T^{-K/2}$. If the design sequence $\{X_T\}$ makes G_T nonlattice after some T_0, then arguing as in Shepp (1964) and Stone (1965), the Radon–Nikodyn derivative,

$$\lim_{T\rightarrow\infty} \frac{G_T\{C_T\}}{\lambda\{C_T\}} = \lim_{T\rightarrow\infty} T^{-K/2}\Pr[\tilde{\mathbf{Z}}_T \in C_T] = g(0) \ , \quad (3.9)$$

where $g(0)$ is the density of G evaluated at the origin. That is,

$$\begin{aligned} &T^{K/2}\Pr[\mathbf{Z}_T(\boldsymbol{\delta}, h) \in C_T] \\ &\quad = (2\pi)^{-K/2}|\tfrac{1}{4}\mathbf{X}'(h)^{-1}\mathbf{Q}\mathbf{X}(h)^{-1}|^{-\frac{1}{2}} \\ &\qquad \cdot \exp\{-\tfrac{1}{2}f^2(0)\boldsymbol{\delta}'\mathbf{Q}\mathbf{X}'(h)^{-1} \\ &\qquad \cdot [\tfrac{1}{4}\mathbf{X}'(h)^{-1}\mathbf{Q}\mathbf{X}(h)^{-1}]^{-1}\mathbf{X}(h)^{-1}\mathbf{Q}\boldsymbol{\delta}\} + o(1) \\ &\quad = (2\pi)^{-K/2}2^K|\mathbf{X}(h)||\mathbf{Q}|^{-\frac{1}{2}}\{\exp - \tfrac{1}{2}\omega^2\boldsymbol{\delta}'\mathbf{Q}\boldsymbol{\delta}\} + o(1) \ . \end{aligned}$$

The continuity of the density at the median implies

$$\prod_{t\in h} f(T^{-\frac{1}{2}}\mathbf{x}_t\boldsymbol{\delta}) = f^K(0) + o(1) \ . \quad (3.10)$$

So substituting back into (3.1), we have

$$\begin{aligned} \phi_T(\boldsymbol{\delta}) = \sum_{h\in H} T^{-K}|\mathbf{X}(h)|^2|\mathbf{Q}|^{-\frac{1}{2}}(2\pi)^{-K/2}\omega^{-K} \\ \cdot \exp\{-\tfrac{1}{2}\omega^2\boldsymbol{\delta}'\mathbf{Q}\boldsymbol{\delta}\} + o(1)\sum_{h\in H} T^{-K}|\mathbf{X}(h)|^2 \ . \end{aligned}$$

But $\sum_{h\in H} |\mathbf{X}(h)|^2 = |\mathbf{X}'\mathbf{X}|$ (see Rao 1973, p. 32). Thus, simplifying, we have

$$\phi_T(\boldsymbol{\delta}) \rightarrow (2\pi)^{-K/2}|\omega^{-2}\mathbf{Q}|^{\frac{1}{2}}\exp\{\tfrac{1}{2}\omega^2\boldsymbol{\delta}'\mathbf{Q}\boldsymbol{\delta}\} \ . \quad (3.11)$$

Finally, Scheffé's Theorem (1947) on the convergence of densities yields the desired conclusion.

If G_T is lattice in one or more directions, the expression (3.9) holds only up to some bounded constant of proportionality involving counting measure on the lattice points in C_T. Summing over H yields a uniformly bounded density proportional to $\exp\{-\frac{1}{2}\omega^2\boldsymbol{\delta}'\mathbf{Q}\boldsymbol{\delta}\}$; thus, by the Lebesgue dominated convergence theorem, the sequence of integrals of $g_T(\boldsymbol{\delta})$ converges, and our result follows as in the nonlattice case (cf. Sheffé 1947, p. 437).

4. CONCLUSION

The least absolute error estimator has been demonstrated to be more efficient than the least squares estimator in the linear model for any error distribution for which the median is more efficient than the mean. This result considerably strengthens the existing rationale for the use of the LAE estimator, making it particularly attractive relative to least squares when the regression process is thought to be potentially long-tailed or to possess peaked density at the median.

Our main theorem also provides, for the first time, the foundation for a large-sample hypothesis testing apparatus for the LAE estimator. Any of the conventional schemes for calculating confidence intervals for the median, applied to the LAE residuals, will provide a consistent estimator of the density at the median, and therefore, normal theory tests with asymptotic justification may be readily constructed.

[*Received November 1976. Revised January 1978.*]

REFERENCES

Abdelmalek, N.N. (1974), "On the Discrete Linear L_1 Approximation and L_1 Solutions of Overdetermined Linear Equations," *Journal of Approximation Theory*, 11, 38–53.

Andrews, David F. (1974), "A Robust Method for Multiple Linear Regression," *Technometrics*, 16, 523–531.

Bassett, Gilbert W., Jr. (1973), *Some Properties of the Least Absolute Error Estimator*, unpublished Ph.D. thesis, Department of Economics, University of Michigan.

Bickel, Peter J. (1973), "On Some Analogues to Linear Combinations of Order Statistics in the Linear Model," *Annals of Statistics*, 1, 597–616.

——— (1975), "One-Step Huber Estimates in the Linear Model," *Journal of the American Statistical Association*, 70, 428–434.

Cramér, Harald (1970), *Random Variables and Probability Distributions*, Cambridge: Cambridge University Press.

Harvey, Andrew C. (1977), "A Comparison of Preliminary Estimators for Robust Regression," *Journal of the American Statistical Association*, 72, 910–913.

Hill, Richard W., and Holland, Paul W. (1977), "Two Robust Alternatives to Least-Squares Regression," *Journal of the American Statistical Association*, 72, 828–833.

Huber, Peter J. (1972), "Robust Statistics: A Review," *Annals of Mathematical Statistics*, 43, 1041–1067.

——— (1973), "Robust Regression: Asymptotics, Conjectures and Monte Carlo," *Annals of Statistics*, 1, 799–821.

Koenker, Roger W., and Bassett, Gilbert W., Jr. (1978), "Regression Quantiles," *Econometrica*, 46, 33–50.

de Laplace, Pierre Simon (1818), *Deuxième Supplement à la Théorie Analytique des Probabilités*, Paris: Courcier.

Malinvaud, E. (1970), *Statistical Methods of Econometrics*, 2nd rev. ed., Amsterdam: North–Holland Publishing Co., New York: American Elsevier Publishing Co.

Rao, C. Radhakrishna (1973), *Linear Statistical Inference and Its Applications*, 2nd ed., New York: John Wiley & Sons.

Relles, Daniel A. (1968), "Robust Regression by Modified Least Squares," unpublished Ph.D. dissertation, Yale University.

Rosenberg, Barr, and Carlson, D. (1971), "The Sampling Distribution of Least Absolute Residuals Regression Estimates," Working Paper No. IP-164, Institute of Business and Economic Research, University of California, Berkeley.

Scheffé, Henry (1974), "A Useful Convergence Theorem for Probability Distributions," *Annals of Mathematical Statistics*, 18, 434–438.

Shepp, L.A., (1964), "A Local Limit Theorem," *Annals of Mathematical Statistics*, 35, 419–423.

Stigler, Stephen M. (1973), "Laplace, Fisher, and the Discovery of the Concept of Sufficiency," *Biometrika*, 60, 439–445.

Stone, Charles (1965), "A Local Limit Theorem for Nonlattice Multi-dimensional Distribution Functions," *Annals of Mathematical Statistics*, 36, 546–551.

Taylor, Lester D. (1974), "Estimation by Minimizing the Sum of Absolute Errors," in *Frontiers in Econometrics*, ed. Paul Zarembka, New York: Academic Press.

Yohai, Victor J. (1974), "Robust Estimation in the Linear Model," *Annals of Statistics*, 2, 562–567.