

LEAST ABSOLUTE DEVIATION ESTIMATES IN
AUTOREGRESSION WITH INFINITE VARIANCE

By

S. Gross

W. L. Steiger

Technical Report # 62

Dr. S. Gross
Dept. of Statistics
Baruch College
The City University of New York

Dr. W. L. Steiger
Dept. of Computer Science
Rutgers University

Department of Computer Science
Rutgers, The State University of New Jersey
New Brunswick, New Jersey 08903

Classification: Probability/Statistics

Title: Least Absolute Deviation Estimates in
Autoregression with Infinite Variance

Authors: S. Gross
Dept. of Statistics
Baruch College
The City University of New York

W.L. Steiger
Department of Computer Science
Rutgers University

Key Words: Least Absolute Deviation, Autoregression,
Infinite Variance

Running Title: Least Absolute Deviation Estimates

AMS (MDS) Classification: Primary 62J05, 60G10
Secondary 62P20, 93E10

Abstract

We consider an L_1 analogue of the least squares estimator for the parameters of stationary, finite order autoregressions. This estimator, the least absolute deviation (LAD), is shown to be strongly consistent via a result that may have independent interest. The striking feature is that the conditions are so mild as to include processes with infinite variance, notably the stationary, finite autoregressions driven by stable increments in L_α , $\alpha > 1$. Finally sampling properties of LAD are compared to those of least squares. Together with a known convergence rate result for least squares, the Monte-Carlo study provides evidence for a conjecture on the convergence rate of LAD.

1. Introduction and Summary

There has been a good deal of interest in stochastic processes based on random variables with infinite variance (e.g. [3] - [6], [8] - [16], [19]). No doubt this is due to the inherent challenge ([12], e.g.) and theoretical interest provided by the non-normal stable laws as well as the possibility that processes constructed from these laws may be appropriate models for many, diverse phenomena ([13] - [16]).

In this paper we consider the problem of estimating the regression parameter vector $\underline{a} = (a_0, \dots, a_k)$ in the k^{th} order autoregressive model

$$(1) \quad X_n = a_0 + a_1 X_{n-1} + \dots + a_k X_{n-k} + U_n$$

where $k > 0$ is a given integer and the error variables U_i form a sequence of nonlattice, integrable, independent and identically distributed random variables. We assume that \underline{a} satisfies the invertibility condition $a_1 z + \dots + a_k z^k \neq 1$ for any complex $z, |z| \leq 1$. Thus (1) is invertible and may be written as

$$(2) \quad X_n = \sum_{j=0}^{\infty} b_j U_{n-j}$$

for some constants b_j which are uniquely determined by \underline{a} .

It was recently shown by Yohai and Maronna [19] that integrability of the positive part of $\log |U_i|$ is a sufficient condition for the a.s. convergence of (2). Since $E |U_i| < \infty \Rightarrow E \log_+ |U_i| < \infty$, the process in (2) converges a.s. This in turn guarantees the existence of a stationary process $\{X_i\}$ that satisfies (1). We further assume that $\text{Median}(U_i) = 0$.

Least squares estimators are known to be consistent for \underline{a} . Kanter & Steiger [9,10] studied (1) in case the U_n are in the domain of attraction of a symmetric stable law of index $\alpha \in (0,2]$. They show that $\hat{\underline{C}}_N$, the least squares estimate of \underline{a} , based on N observations of $\{X_i\}$, satisfies

$$(3) \quad N^\delta |\hat{\underline{C}}_N - \underline{a}| \rightarrow 0$$

in probability for all $\delta < 2/\alpha - 1/\min(1,\alpha)$ while in [8], (3) was shown to hold a.s. for $\delta < 1/\alpha$. Yohai and Maronna [19] enlarged the class of processes for which $\hat{\underline{C}}_N$ is consistent by showing that $N^{1/2} |\hat{\underline{C}}_N - \underline{a}| \rightarrow 0$

in probability for U_i symmetric and satisfying $E(\log_+ |U_i|) < \infty$. The latter condition is obeyed by all symmetric stable variables U_i . The greater generality is obtained at the expense of rate of convergence.

In the present paper we consider the LAD, or least absolute deviation estimators of \underline{a} , defined as a random variable \underline{L}_N at which the function

$$(4) \quad f_N(\underline{c}) = \sum_{i=1}^N |X_{i+k} - (c_0 + c_1 X_{i+k-1} + \dots + c_k X_i)|$$

attains a minimum; as long as U_i is nonlattice, \underline{L}_N is a.s. unique. In the next section we establish the strong consistency of \underline{L}_N .

However consistency is not justification enough for considering LAD, especially in view of the favourable properties of least squares indicated

by (3) and (4). Any such justification must address the obvious complexity of computing L_N from real data by showing that it is first feasible to calculate L_N and then, that it is worthwhile to do so.

An attempt is made to do this in Section 3. We briefly discuss an algorithm [1] to compute L_N exactly and present some Monte-Carlo results that portray its efficiency. The conclusion is that while L_N is computationally several times more costly than least squares, it is affordable. The same Monte-Carlo experiments indicated that it should be worth the extra expense. In the stable autoregressions LAD appears to converge much faster than the least squares rate of N^δ as shown in (3).

It may be useful to close this introduction by recalling some basic facts about stable laws: further details may be found in Feller [7]. A random variable X is stable of index $\alpha \in (0,2]$ in case for each n there is a constant c_n that makes

$$L(X) = L((X_1 + \dots + X_n)/n^{1/\alpha} - c_n),$$

the X_i being i.i.d. variables with $L(X_i) = L(X)$ and $L(\cdot)$ referring to the law of the variable inside parentheses. A random variable Z is in the domain of attraction of a stable law of index $\alpha < 2$ if $t^\alpha P\{|Z| > t\} \rightarrow k \neq 0$, a constant, as $t \rightarrow \infty$, in which case a sequence $\{c_i\}$ may be found that makes $L((Z_1 + \dots + Z_n)/n^{1/\alpha} - c_n) \rightarrow L(X)$, the Z_i being i.i.d. copies of Z . Finally if $X \neq 0$ is symmetric and stable of index α , $E(e^{itX}) = e^{-c|t|^\alpha}$ for some $c > 0$ and if X, Y, Z are i.i.d. symmetric stable variables, then for any a, b ,

$$L(aY+bZ) = L((|a|^\alpha + |b|^\alpha)^{1/\alpha} X)$$

2. Strong Consistency of LAD.

The main result of this section is contained in Theorem 1, which is stated below after some preliminary definitions.

Let $\{\dots, U_{-1}, U_0, U_1, \dots\}$ denote a sequence of i.i.d. nonlattice integrable random variables defined on a probability space (Ω, A, p) . Fix an integer $k \geq 1$ and constants a_1, \dots, a_k and consider the k^{th} order autoregression

$$(5) \quad X_n = a_1 X_{n-1} + \dots + a_k X_{n-k} + U_n$$

It follows from Yohai and Maronna [19] that if $a_1 z + \dots + a_k z^k \neq 1$ for any complex z , $|z| \leq 1$, and if the $\{U_i\}$ are integrable then (5) may be written as

$$(6) \quad X_n = \sum_{j=0}^{\infty} b_j U_{n-j}$$

for appropriate constants b_j ; the sum on the right converges almost surely. In other words, there is a stationary process $\{X_i\}$ that satisfies (5). For this stationary, k^{th} order autoregression, we study the problem of estimating $\underline{a} = (a_1, \dots, a_k)$ based on N observations.

Consider the distance function

$$(7) \quad f_N(\underline{c}) = N^{-1} \sum_{j=1}^N |R_j(\underline{c})|$$

where $R_j(\underline{c}) = X_{j+k} - (c_1 X_{j+k-1} + \dots + c_k X_j)$ is the j^{th} residual determined by the vector \underline{c} . Clearly f_N is almost surely a continuous, convex function $R^k \rightarrow R$ because each term of (7) is. The least absolute deviation (LAD) estimator

for \underline{a} based on N residuals, \underline{L}_N is the location of the minimum of f_N , unambiguously defined almost surely by

$$(8) \quad f_N(\underline{L}_N) \leq f_N(\underline{u}), \text{ all } \underline{u} \in R^k.$$

Finally let $C = \{c \in R^k : c_1 z + \dots + c_k z^k = 1, |z| \leq 1\}$ be the set of all possible coefficients for k^{th} order stationary autoregressions that may be written as (6).

Theorem 1: Let $\{\dots, U_{-1}, U_0, U_1, \dots\}$ be an i.i.d. sequence of integrable random variables with a unique Median $(U_i) = 0$ and $\{X_i\}$ a stationary sequence that satisfies (5) for some $\underline{a} \in C \subset R^k$. The least absolute deviation estimators \underline{L}_N satisfy

$$(9) \quad \underline{L}_N \rightarrow \underline{a} \text{ a.s. as } N \rightarrow \infty.$$

In proving Theorem 1 we rely on a nonprobabilistic result (Lemma 1) that seems to be new and of independent interest due to its potential use in establishing consistency of general least distance estimators, provided the distance employed is convex. The proof of Lemma 1 together with a stronger result for $k=1$ is in an appendix.

Lemma 1: Let $\{h_n\}$ be an equicontinuous sequence of convex functions $R^k \rightarrow R$ that converge pointwise to a continuous convex function h on an open set $U \subset R^k$. Suppose h_n has a minimum at \underline{x}_n and h , a unique minimum at $\underline{x} \in U$; i.e.,

$$(10) \quad \begin{aligned} h_n(x_n) &\leq h_n(\underline{u}), \text{ all } \underline{u} \\ h(\underline{x}) &< h(\underline{u}), \quad \underline{u} \neq \underline{x}. \end{aligned}$$

Then

$$(11) \quad \underline{x}_n \rightarrow \underline{x} \text{ as } n \rightarrow \infty.$$

Proof of Theorem 1: We shall establish the following facts:

(i) For each fixed $\underline{c} \in C$ the sequence $f_N(\underline{c})$ converges almost surely as $N \rightarrow \infty$ to the limit function

$$(12) \quad f(\underline{c}) = E|X_{k-1} - (c_1 X_k + \dots + c_k X_1)|;$$

(ii) The limit function f is convex on C and has a unique minimum at \underline{a} .

Thus,

$$(13) \quad f(\underline{a}) < f(\underline{c}), \quad \underline{c} \neq \underline{a}.$$

Statement (i) follows from the ergodicity and stationarity of the process $\{X_n\}$, which in turn implies that the residual process $\{R_j(\underline{c})\}$ is stationary and ergodic (see Doob [2]). Hence $f_N(\underline{c}) \rightarrow E|R_0(\underline{c})| = f(\underline{c})$ by the ergodic theorem. Finally $|X_{k+1} - (c_1 X_k + \dots + c_k X_1)|$ is a convex function of \underline{c} on the set C , so $E(|X_{k+1} - (c_1 X_k + \dots + c_k X_1)|) = f(\underline{c})$ is convex.

To prove (13) write $\underline{c} = \underline{a} + \underline{d}$ and note that $f(\underline{c}) = f(\underline{a} + \underline{d}) = E(|X_{k+1} - (a_1 X_k + \dots + a_k X_1) - (d_1 X_k + \dots + d_k X_1)|) = E(|U_{k+1} - (d_1 X_k + \dots + d_k X_1)|)$, the last equality a consequence of (5).

Letting F_k denote the σ -field generated by X_1, \dots, X_k we have

$$(14) \quad f(\underline{c}) = E[E(|U_{k+1} - (d_1 X_k + \dots + d_k X_1)|) | F_k]$$

from which it is clear that f is minimized when

$$(15) \quad E(|U_{k+1} - (d_1 X_k + \dots + d_k X_1)| | F_k)$$

is. Since U_{k+1} is independent of F_k and has a unique median at 0, \underline{d} must be chosen so that

$$(16) \quad d_1 X_k + \dots + d_k X_1 = 0$$

on F_k , which because $\{U_i\}$ are nonlattice, implies that $\underline{d}=0$ at the unique minimum of f .

This completes the proof of statement (ii). As $f_N(\underline{c})$ is a sequence of continuous and convex functions on C for which $f_N(\underline{L}_N) \leq f_N(\underline{u})$ for all $\underline{u} \in R^k$, Theorem 1 will follow from Lemma 1 and statements (i), (ii) provided we verify that $\{f_N\}$ is almost surely equicontinuous on C , a fact that follows easily from

$$\begin{aligned} |f_N(\underline{c}) - f_N(\underline{c} + \underline{t})| &\leq N^{-1} \sum_{j=1}^N |t_1 X_{k+j-1} + \dots + t_k X_j| \\ &\leq \|\underline{t}\| \{N^{-1} \sum_{j=1}^N (|X_{k+j-1}| + \dots + |X_j|)\} \end{aligned}$$

and the ergodic theorem.

□

Remark. The theorem holds even if the $\{U_i\}$ are not centered at zero.

If $U_n = a_0 + V_n$, the $\{V_i\}$ i.i.d. with unique $\text{Median}(V_i) = 0$, the auto-regression (5) may be written as

$$(17) \quad X_n = a_0 + a_1 X_{n-1} + \dots + a_k X_{n-k} + V_n.$$

If we define $\underline{L}_N \in R^{k+1}$ as the location of the minimum of

$$(18) \quad g_N(\underline{c}) = N^{-1} \sum_{j=1}^N |X_{k+j} - (c_0 + c_1 X_{k+j-1} + \dots + c_k X_j)|$$

and

$$(19) \quad g(\underline{c}) = E |X_n - (c_0 + c_1 X_{n-1} + \dots + c_k X_{n-k})|$$

$g_N \rightarrow g$ a.s. on $R \times C$. Furthermore statement (ii) in Theorem 1 adapts to show that g has a unique minimum at $(a_0, a_1, \dots, a_k) = (a_0, \underline{a})$ from which, using Lemma 1, $\underline{L}_N \rightarrow (a_0, \underline{a})$.

3. Monte-Carlo Studies of LAD Estimates.

In this section we present Monte-Carlo results that compare sampling distributions of \underline{L}_N with those of the least squares estimators. Since rate of convergence results are known for the latter in the stable case, we chose to present simulation results for symmetric stable processes of index $0 < \alpha \leq 2$. These lend support to a conjecture about the rate of

convergence of L_N to \underline{a} and to a possible generalization of Theorem 1.

In the experiments we generated stable random numbers using the fact that for $\alpha < 1$ the random variable $S_\alpha = [f(U)/V]^{1/\alpha-1}$ is positive stable of index α , where U and V are independent with $P\{U \leq t\} = t$, $0 \leq t \leq 1$, and $P\{V \leq t\} = 1 - e^{-t}$, $0 \leq t < \infty$, and $f(x) = (\sin \pi \alpha x / \sin \pi x)^{1/(1-\alpha)} (\sin \pi x(1-\alpha) / \sin \pi \alpha x)$, for $0 \leq x \leq 1$ (see [11]). Moreover if X is symmetric stable of index $\alpha \in (0, 2]$ and Y is positive stable of index $\beta < 1$ and independent of X , then

$$R = X(Y)^{1/\alpha}$$

is symmetric stable of index $\alpha\beta$ (see [7;p172]) from which it follows that

$$(20) \quad N(S_{\alpha/2})^{1/2}$$

is symmetric stable of index α , where N is normal, $E(N)=0$, $E(N^2)=2$, and $S_{\alpha/2}$ positive stable of index $\alpha/2$ and independent of N ; in fact, the variable in (20) is "standard", with characteristic function $e^{-|t|^\alpha}$.

We simulated the stationary autoregression

$$(21) \quad X_n = a_1 X_{n-1} + a_2 X_{n-2} + \dots + a_k X_{n-k} + U_n$$

with U_i symmetric stable of index α , generated as in (20). This was done via the moving average representation of the process,

$$x_n = \sum_{j=0}^{\infty} b_j U_{n-j}$$

and approximating the solution to (21) by the (non stationary) process

$$(22) \quad x_n = \sum_{j=0}^{199+n} b_j U_{n-j}$$

For $N > 0$ a sample of size N, X_1, \dots, X_N is computed from U_{-199}, \dots, U_N according to (22).

Fifty samples of size $N+k$ were generated. For each, a value of the LAD estimator L_N , and of the least squares estimator C_N was computed. The estimate C_N was computed as the solution of the normal equation:

$$C_N^T = (W^T W)^{-1} W^T Y^T$$

where
$$W = \begin{pmatrix} X_k & \dots & X_1 \\ \vdots & & \\ X_{N+k-1} & \dots & X_N \end{pmatrix} \quad \text{and } Y = (X_{k+1}, \dots, X_{N+k})$$

The exact value of the LAD estimator was found using a recently developed algorithm [1]. It is based on the following observations.

- (i) Let X_1, \dots, X_{N+k} be a sample of size $N+k$ from (17) and let $\Pi = \{Z_\ell = (X_\ell, \dots, X_{\ell+k}), \ell=1, \dots, N\}$ be the set of N $k+1$ -tuples, or "points" formed by $k+1$ successive elements of $\{X_i\}$. It is well known ([18], e.g.) and easy to prove that at the minimum value of (18),

\underline{c} may be chosen so that at least $k+1$ terms in the sum are zero; i.e.,

$$X_{i_j} = c_0 + c_1 X_{i_j-1} + \dots + c_k X_{i_j-k} \quad \text{for certain distinct indices } i_j, j=1, \dots, k+1, \text{ so } \underline{c} \text{ may be regarded as a } k \text{ dimensional hyperplane in } R^{k+1}$$

containing the points $\underline{z}_{i_1}, \underline{z}_{i_2}, \dots, \underline{z}_{i_{k+1}} \in \Pi$. Since the minimum of (18) is unique a.s., \underline{L}_N is determined by one of the $\binom{N}{k+1}$ hyperplanes generated by the $k+1$ -tuples from Π .

(ii) Exhaustive search of these hyperplanes is avoided by starting with an initial hyperplane \underline{c}_0 and iterating through a sequence $\underline{c}_1, \dots, \underline{c}_n$ of hyperplanes until (18) is minimized. For the basic iteration suppose one is at \underline{c}_i , determined by points $\underline{z}_{j_1}, \dots, \underline{z}_{j_{k+1}}$ from Π . Since $\underline{z}_{j_2}, \dots, \underline{z}_{j_{k+1}}$ are k points in R^{k+1} , there are distinct hyperplanes $\underline{\theta} = (\theta_0, \dots, \theta_k) \neq 0$ and $\underline{V} = (V_0, \dots, V_k) \neq 0$ that contain them. Any hyperplane \underline{d} containing $\underline{z}_{j_2}, \dots, \underline{z}_{j_{k+1}}$ may be written as $\underline{d} = \underline{\theta} + t(\underline{V} - \underline{\theta}) = \underline{\theta} + t\underline{\delta}$ for some real t and $\underline{\delta} = \underline{V} - \underline{\theta}$. To minimize (18), one must minimize

$$\begin{aligned}
 f(\underline{d}) &= \sum_{j=1}^N |X_{j+k} - (d_0 + d_1 X_{j+k-1} + \dots + d_k X_j)| \\
 (23) \quad &= \sum_{j=1}^N |X_{j+k} - (\theta_0 + \theta_1 X_{j+k-1} + \dots + \theta_k X_j) - t(\delta_0 + \delta_1 X_{j+k-1} + \dots + \delta_k X_j)| \\
 &= \sum_{j=1}^N |w_j - t u_j|
 \end{aligned}$$

where we have written $w_j = X_{j+k} - (\theta_0 + \theta_1 X_{j+k-1} + \dots + \theta_k X_j)$ and $u_j = \delta_0 + \delta_1 X_{j+k-1} + \dots + \delta_k X_j$. Note that $w_{j_2} = \dots = w_{j_{k+1}} = 0$ because $\underline{z}_{j_2}, \dots, \underline{z}_{j_{k+1}}$ are in $\underline{\theta}$ and that $u_{j_2} = \dots = u_{j_{k+1}} = 0$ because $\underline{\delta} = \underline{V} - \underline{\theta}$ is orthogonal to $\underline{z}_{j_2}, \dots, \underline{z}_{j_{k+1}}$, so that k terms of (23) are already 0. The \underline{d} that minimizes (23) is determined by solving the one dimensional problem for t using the idea in Singleton [17] and thus forcing an additional, $(k+1)^{st}$ term of (23) to be zero. The next iterate, \underline{c}_{i+1} , is set to the minimizing \underline{d} and $f(\underline{c}_{i+1}) \leq f(\underline{c}_i)$.

(iii) If the model in (17) is constrained to have $a_0 = 0$, \underline{c}_0 in (18) is zero, as is θ_0 and V_0 , so the dimensionality, k , may be reduced by 1.

We point out that each iteration of the above procedure can be affected by solving $k+1$ linear equations and ordering N numbers whereas computation of C_N simply requires the solution of $k+1$ linear equations. Thus if n such iterations are required to compute L_N , $\text{cost}(L_N) = n(O_N + \text{cost}(C_N))$, where "cost" is a measure of computational complexity and O_N is the complexity of ordering N numbers.

The following tables depict the sampling distributions of L_N, C_N , and the number of iterations n that were required to compute L_N . The sampling distributions are based on 50 observations each, for $k=2, a_1=1.4, a_2=-.7$.

We are not really trying to estimate the sampling distributions of L_N and C_N . Instead, we want to compare the accuracy, speed of convergence, and computation times of the estimators. For this reason the tables show only the MIN, the MAX, and the three quartiles Q_1, Q_2, Q_3 of the relevant distributions. For each N and α , these are shown for the residuals $L_N - a$, for $C_N - a$, and for the number of iterations, n , required to compute L_N . Thus in Table 1 for $N=100$, the largest deviation of the least squares estimator of a_1 is 0.1816. The maximum number of iterations that were required to compute L_{100} among the 50 samples was 10 and the maximum deviation of the first component of L_{100} was 0.1101.

The four tables suggest the following observations.

Remark 1. The distribution of the number of iterations required to compute L_N stayed fairly stable as the sample size $N+2$ increased from 52 to 502 and α varied from 2.0 to 1.6 to 1.2 to .6. We estimate that each iteration to compute L_{100} (which is approximately equal to the time required to compute C_{100}) was approximately 0.37 sec of CPU time on the Rutgers PDP-10 computer. Thus the maximum CPU time to compute L_{100} was approximately 3.7 seconds when $\alpha=2$. \square

Remark 2. Comparison of \underline{C}_N and \underline{L}_N across tables reveals some unexpected results. For $\alpha=2.0$ where \underline{C}_N is expected to do better than \underline{L}_N , the distributions are quite comparable, the $N=500$ case possibly suggesting that $\underline{C}_N - \underline{a} \rightarrow 0$ a little faster than $\underline{L}_N - \underline{a}$ does. When $\alpha=1.6$ (Table 2), there is little to choose between them while for $\alpha=1.2$ (Table 3), \underline{L}_N is distributed much more tightly about \underline{a} than is \underline{C}_N , and to about the same extent for each N . Finally for $\alpha=.6$ (Table 4) \underline{L}_N is clearly superior to \underline{C}_N and becomes more so as N increases. \square

These remarks lead us to two conjectures that we cannot, as yet, prove. First, because of Table 4, we believe

Conjecture 1: For stable autoregressions, \underline{L}_N is consistent even if $E(|U_i|) = \infty$. \square

Next, it was proved in [8] that $N^\sigma ||\underline{C}_N - \underline{a}|| \rightarrow 0$ a.s. as long as $\sigma < 1/\alpha$. On the basis of the tables and Remark 2, we are led to the following statement.

Conjecture 2: For stable autoregressions, $N^\sigma | \underline{L}_N - \underline{a} | \rightarrow 0$ in probability, for $\sigma < 2/\alpha$.

Acknowledgement. Thanks are due to the referee for pointing out errors in the original simulations.

References

- [1] Bloomfield, P. and Steiger, W. L. (1978) Least absolute deviation curve-fitting. (submitted for publication).
- [2] Doob, J. L. (1953) Stochastic Processes, John Wiley, New York.
- [3] du Mouchel, W. H. (1973) Stable distributions in statistical inference: 1. Symmetric stable distributions compared to other symmetric long-tailed distributions. J. Amer. Statist. Assoc. 68, 469-477.
- [4] du Mouchel, W. H. (1973) On the asymptotic normality of the maximum likelihood estimates when sampling from a stable distribution. Ann. Statistics 1, 948-957.
- [5] Fama, E. and Roll, R. (1968) Some properties of symmetric stable distributions. J. Amer. Statist. Assoc. 63, 817-836.
- [6] Fama, E. and Roll, R. (1971) Parameter estimation for symmetric stable distributions. J. Amer. Statist. Assoc. 66, 331-338.
- [7] Feller, W. (1966) An Introduction to Probability Theory and its Applications, Vol. 2, John Wiley, New York.
- [8] Kanter, Marek and Hannan, E. J. (1977) Autoregressive Processes with Infinite Variance, J. Appl. Probability 14, 411-415.
- [9] Kanter, Marek and Steiger, W. L. (1974) Regression and autoregression with infinite variance, Adv. Appl. Probability 6, 768-783.
- [10] Kanter, Marek and Steiger, W. L. (1977) Estimating Linear Relationships for Models Based on Random Variables with Infinite Variance, Proceedings of the Fifth Brasov Conference on Probability Theory, 317-323.
- [11] Kanter, Marek (1975) Stable densities under change of scale and total variation inequalities, Ann. Probability 3, 697-707.
- [12] Kanter, Marek (1976) On the unimodality of stable densities, Ann. Probability 4, 1006-1008.
- [13] Mandelbrot, Benoit (1963) The variation of certain speculative prices, J. Business U. Chicago 26, 294-419.
- [14] Mandelbrot, Benoit (1967) The variation of some other speculative prices, J. Business U. Chicago 40, 393-413.
- [15] Mandelbrot, Benoit (1974) A population birth-and-mutation process I: explicit distributions for the number of mutants in an old culture of bacteria. J. Appl. Probability 11, 437-444.

- [16] Mandelbrot, Benoit and Taylor, H. (1967) On the distribution of stock price differences, Operations Research 15, 1057-1062.
- [17] Singleton, J. (1940) A Method for Minimizing the Sum of Absolute Deviations, Ann. Math. Statist. 11, 301-310.
- [18] Wagner, H. M. (1959) Linear Programming Techniques for Regression Analysis, J. Amer. Statist. Assoc., 206-212.
- [19] Yohai, Victor J. and Maronna, Ricardo A. (1977) Asymptotic behavior of Least Squares estimates for Autoregressive Processes with Infinite Variances, Ann. Statistics 5, 554-560.

4. Appendix.

Proof of Lemma 1. If $x_n \not\rightarrow x$ there is $\epsilon > 0$ and a subsequence n_j for which $\|x_{n_j} - x\| \geq \epsilon$. Choose $t < \epsilon$ so that $h_n(u) \rightarrow h(u)$ whenever $\|u - x\| \leq t$. Write $S(t)$ for the surface of the sphere $\{u: \|u - x\| = t\}$ and note that h_{n_j} attains a minimum (not necessarily uniquely) at, say, $u_{n_j} \in S(t)$.

There is a limit point, u_0 and a subsequence m_j of n_j such that $u_{m_j} \rightarrow u_0$. Equicontinuity implies $h_{m_j}(u_{m_j}) \rightarrow h(u_0)$. Pointwise convergence combines to imply that for large enough m_j , $h_{m_j}(u_{m_j}) > h(u_0) - \sigma/4$ and $h_{m_j}(x) < h(x) + \sigma/4$ where $\sigma = h(u_0) - h(x) > 0$; therefore $h_{m_j}(u_{m_j}) - h_{m_j}(x) > \sigma/2$. This shows that for m_j large enough, each h_{m_j} attains a minimum in $B(t) = \{u: \|u - x\| < t\}$ which, together with convexity, contradicts $x_n \not\rightarrow x$. \square

As a final remark, it is worth noting that if $k = 1$ in Lemma 1, the result holds without equicontinuity of $\{h_i\}$.

Lemma 2: $\{h_i\}$ and h are continuous convex functions and

$$h_n(x_n) \leq h_n(u) \quad \text{all } u \in R$$

$$h(x) < h(u) \quad u \neq x$$

If $h_n \rightarrow h$ pointwise on an open set containing x then $x_n \rightarrow x$.

Proof: If $x_n \not\rightarrow x$ there is $\epsilon > 0$ and a subsequence n_j such that $\|x_{n_j} - x\| \geq \epsilon$. Take $t < \epsilon$ such that $h_n \rightarrow h$ for all u in $S = \{u: \|x - u\| \leq t\}$. Let $\sigma > 0$ be

$\min(h(x+t)-h(x), h(x-t)-h(x))$ and using pointwise convergence, choose N so large that $h_{n_j}(x+t) > h(x+t) - \sigma/4$, $h_{n_j}(x-t) > h(x-t) - \sigma/4$, and $h_{n_j}(x) < h(x) + \sigma/4$ when $n_j > N$. This shows that $h_{n_j}(x+t) - h_{n_j}(x) > \sigma/2$ and $h_{n_j}(x-t) - h_{n_j}(x) > \sigma/2$ for $n_j > N$ which implies that each h_{n_j} has a minimum in $(x-t, x+t)$ by convexity, in contradiction to $|x_{n_j} - x| \geq \epsilon$. \square

Table 1 ($\alpha = 2.0$)

Sampling Distributions of \underline{L}_N and \underline{C}_N for various N ,
 based on 50 observations of each estimator and of
 n , the number of iterations required to compute \underline{L}_N .

<u>N = 50</u>					
	<u>n</u>	<u>\underline{C}_N</u>		<u>\underline{L}_N</u>	
MAX	9	.3479	.1782	.3728	.2664
Q3	5	.0804	.0526	.0945	.0966
Q2	5	.0357	.0082	.0178	.0045
Q1	4	-.0446	-.0817	-.0590	-.1099
MIN	2	-.1586	-.4752	-.2425	-.5366
<u>N = 100</u>					
	<u>n</u>	<u>\underline{C}_N</u>		<u>\underline{L}_N</u>	
MAX	10	0.1816	0.2015	0.1101	.1641
Q3	6	0.0403	0.0325	0.0221	0.0619
Q2	5	-0.0081	0.0027	-0.0029	0.0041
Q1	4	-0.0457	-0.0506	-0.0643	-0.0673
MIN	2	-0.1429	-0.1683	-0.2044	-0.2059
<u>N = 500</u>					
	<u>n</u>	<u>\underline{C}_N</u>		<u>\underline{L}_N</u>	
MAX	9	0.1056	0.0809	0.0744	0.0744
Q3	7	0.0207	0.0232	0.0359	0.0359
Q2	5	-0.0024	0.0036	-0.0053	-0.0069
Q1	4	-0.0223	-0.0168	-0.0322	-0.0239
MIN	2	-0.0722	-0.0904	-0.1028	-0.0893

Table 2 ($\alpha = 1.6$)

Sampling Distributions of \underline{L}_N and \underline{C}_N for various N ,
 based on 50 observations of each estimator and of
 n , the number of iterations required to compute \underline{L}_N .

$N = 50$

	<u>n</u>	<u>\underline{C}_N</u>		<u>\underline{L}_N</u>	
MAX	7	0.3431	0.1316	0.4556	0.2345
Q3	5	0.0848	0.0566	0.0828	0.0645
Q2	4	0.0165	0.0062	0.0069	0.0073
Q1	3	-0.0354	-0.0880	-0.0521	-0.0654
MIN	2	-0.1415	-0.4155	-0.2476	-0.4138

$N = 100$

	<u>n</u>	<u>\underline{C}_N</u>		<u>\underline{L}_N</u>	
MAX	8	0.1257	0.0692	0.1743	0.0776
Q3	6	0.0520	0.0050	0.0200	0.0299
Q2	5	0.0131	-0.0194	-0.0124	0.0088
Q1	3	-0.0287	-0.0454	-0.0413	-0.0349
MIN	2	-0.0817	-0.1916	-0.0873	-0.1117

$N = 500$

	<u>n</u>	<u>\underline{C}_N</u>		<u>\underline{L}_N</u>	
MAX	9	0.0527	0.0589	0.0509	0.0595
Q3	6	0.00082	0.0168	0.0097	0.0113
Q2	5	-0.0033	0.0021	-0.0001	0.0021
Q1	4	-0.0156	-0.0111	-0.0156	-0.0088
MIN	3	-0.0483	-0.0689	-0.0582	-0.0603

Table 3 ($\alpha = 1.2$) .

Sampling Distributions of \underline{L}_N and \underline{C}_N for various N , based on 50 observations of each estimator and of n , the number of iterations required to compute \underline{L}_N .

<u>N = 50</u>					
	<u>n</u>	<u>\underline{C}_N</u>		<u>\underline{L}_N</u>	
MAX	7	0.3852	0.1085	0.1537	0.0900
Q3	5	0.0778	0.0262	0.0251	0.0098
Q2	4	0.0042	0.0015	0.0015	-0.0040
Q1	3	-0.0295	-0.0452	-0.0090	-0.0259
MIN	2	-0.1204	-0.2313	-0.0567	-0.1481
<u>N = 100</u>					
	<u>n</u>	<u>\underline{C}_N</u>		<u>\underline{L}_N</u>	
MAX	8	0.1224	0.1140	0.0763	0.0552
Q3	6	0.0321	0.0265	0.0145	0.0115
Q2	5	0.0056	-0.0064	-0.0007	0.0023
Q1	4	-0.0187	-0.0346	-0.0136	-0.0150
MIN	2	-0.1248	-0.2027	-0.0709	-0.0849
<u>N = 500</u>					
	<u>n</u>	<u>\underline{C}_N</u>		<u>\underline{L}_N</u>	
MAX	10	0.0688	0.0374	0.0515	0.0202
Q3	6	0.0061	0.0063	0.0029	0.0028
Q2	6	0.0022	-0.0012	-0.0000	-0.0008
Q1	4	-0.0057	-0.0112	-0.0019	-0.0049
MIN	3	-0.0490	-0.0345	-0.0192	-0.0266

Table 4 ($\alpha = .6$)

Sampling Distributions of \underline{L}_N and \underline{C}_N for various N ,
 based on 50 observations of each estimator and of
 n , the number of iterations required to compute \underline{L}_N .

$N = 50$

	<u>n</u>	<u>\underline{C}_N</u>		<u>\underline{L}_N</u>	
MAX	7	0.2155	0.1176	0.0082	0.0167
Q3	5	0.0075	0.0200	0.0009	0.0007
Q2	4	-0.0000	0.0006	0.0000	0.0000
Q1	3	-0.0116	-0.0131	-0.0002	-0.0009
MIN	2	-0.1355	-0.3000	-0.0183	-0.0079

$N = 100$

	<u>n</u>	<u>\underline{C}_N</u>		<u>\underline{L}_N</u>	
MAX	8	0.2138	0.0667	0.0024	0.0023
Q3	5	0.0053	0.0072	0.0001	0.0001
Q2	4	0.0000	-0.0001	-0.0000	0.0000
Q1	4	-0.0049	-0.0094	-0.0001	-0.0002
MIN	2	-0.0675	-0.1764	-0.0012	-0.0031

$N = 500$

	<u>n</u>	<u>\underline{C}_N</u>		<u>\underline{L}_N</u>	
MAX	9	0.0834	0.0676	0.0008	0.0012
Q3	6	0.0029	0.0050	0.0000	0.0000
Q2	5	0.0000	-0.0000	0.0000	-0.0000
Q1	4	-0.0035	-0.0067	-0.0000	-0.0000
MIN	2	-0.0245	-0.5913	-0.0002	-0.0010