COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
AT YALE UNIVERSITY

Box 2125, Yale University
New Haven, Connecticut 06520

Cowles Foundation Discussion Paper No. 949

A SHORTCUT TO LAD ESTIMATOR ASYMPTOTICS

by

Peter C.B. Phillips

July 1990

A SHORTCUT TO LAD ESTIMATOR ASYMPTOTICS

by

P. C. B. Phillips[1]

Cowles Foundation for Research in Economics
Yale University

Preliminary draft: November 1988

Completed version: June 1990

0. ABSTRACT

Using generalized functions of random variables and generalized Taylor series expansions, we provide almost trivial demonstrations of the asymptotic theory for the LAD estimator in a regression model setting. The approach is justified by the smoothing that is delivered in the limit by the asymptotics, whereby the generalized functions are forced to appear as linear functionals wherein they become real valued. Models with fixed and random regressors, autoregressions and autoregressions with infinite variance errors are studied. Some new analytic results are obtained including an asymptotic expansion of the distribution of the LAD estimator and the results of some earlier simulation studies are explained.

Running Head: LAD Asymptotics.

"Shall I refuse my dinner because I do not fully understand the process of digestion." O. Heaviside

1. INTRODUCTION

Classical asymptotic methods such as those given in Cramér (1946, Ch. 33) are usually thought to apply only in cases of regular estimation. Here smoothness conditions facilitate the use of Taylor expansions of the objective function and the associated first order conditions. When supplemented with preliminary consistency arguments, these expansions then yield the required asymptotic distribution theory in a few simple steps. With more tiresome algebra and higher order smoothness conditions, they also supply the formulae for Edgeworth expansions to second and higher orders.

The absence of smoothness in the criterion function that occurs in what are called non–regular cases is generally thought to prevent the use of this classical approach. Early studies that dealt with non–regularity complications by direct methods were Daniels (1961) and Huber (1967). More recently there has been a move towards the use of stochastic equicontinuity arguments and empirical process techniques to address the complications that are presented by the absence of smoothness and continuity in the criterion function. These techniques have been successful in accommodating a wide range of non–regular cases, including simulation based optimization estimators such as those that are employed by McFadden (1989) and Pakes and Pollard (1989). They have also found attractive applications in the development of asymptotics for semiparametric models (Andrews, 1989). The work of Pollard (1984, 1985, 1989a, b) has been especially influential in advancing the use of these techniques in econometric applications. However, as pointed out by Pollard (1989a), stochastic equicontinuity arguments are less accessible to many potential users even though they often capture the key technical difficulty in the asymptotics. Partly in response to this objection, Pollard presents an alternative approach for studying the asymptotic theory of the least absolute deviation (LAD) estimator in a simple regression

1

context. Pollard's alternative approach builds on the convexity of the LAD criterion function to construct a quadratic approximation whose minimand is close enough to the LAD estimator for the latter to share the same asymptotic normal distribution.

The present note is related closely to Pollard (1989a). However, instead of putting forward an alternative approach, our objective is to show the serviceability of the classical approach in non—regular problems like that of the LAD estimator. The idea we put forward is very simple. If the criterion function has non—regularities like discontinuities in its derivatives, these may be accommodated directly by the use of generalized functions, provided the discontinuities are smoothed out asymptotically. First order conditions and Taylor representations can be written down in the usual way but they take the form of generalized Taylor series. They may be formally interpreted as linear functionals in terms of the empirical distribution function. As the sample size $n \rightarrow \infty$, these linear functionals become well behaved provided some basic smoothness conditions are imposed on the underlying probability law of the data. In effect, with this generalization of the classical approach the asymptotics provide the smoothness that is required to justify the Taylor development and thereby the resulting asymptotics.

Our approach is heuristic and we do not claim to deal rigorously with all of the mathematical issues that arise. However, it is hoped that our extended treatment of LAD asymptotics will serve to illustrate the utility of these ideas and to stimulate the interest of others in the use of these methods. To continue the theme put forward by Heaviside in the line that heads this article, we believe there is still good food to enjoy in classical dinners.

2. LAD ASYMPTOTICS: THE HEURISTICS

Suppose y_t is generated by the linear regression

$$(1) \qquad y_t = x_t'\beta^0 + u_t \quad (t = 1, \ldots, n),$$

where the parameter vector $\beta^0 \in \mathbb{R}^k$, the errors u_t satisfy (A_1) below and (x_t) constitutes a bounded, deterministic sequence for which

$$(2) \qquad Q_n = n^{-1}\Sigma_1^n x_t x_t' \to Q,$$

a positive definite limit, as $n \to \infty$.

(A_1) *The sequence* (u_t) *is iid with zero median and probability density* $f(\cdot)$ *that is positive and analytic at zero.*

The LAD estimator $\hat{\beta}_n$ is chosen as a solution of the extremum problem

$$(3) \qquad \hat{\beta}_n = \text{argmin}[n^{-1}\Sigma_1^n |y_t - x_t'\beta|].$$

This is the standard framework for deriving the consistency and asymptotic normality of the LAD estimator. The original work is due to Bassett and Koenker (1978). An extensive study with an analysis of autoregressions as well as the regression model (1) is provided by the monograph by Bloomfield and Steiger (1983). A recent and novel treatment of the subject that includes an historical overview of research and some additional references is Pollard (1989a). Our condition (A_1) is stronger than Pollard's "error assumption" on u_t in that we require the density $f(\cdot)$ to be analytic rather than simply continuous at the origin. But there will be gains to making this stronger assumption. Not only does it help in developing generalized Taylor series but in so doing it facilitates the subsequent development of higher order asymptotics.

In most cases of interest and certainly under the standard assumptions of the regression model given above, the consistency of $\hat{\beta}_n$ is easily established by conventional arguments that involve the limit of the objective function in (3). Amemiya (1985, pp.

152–153) is a convenient source for the details of this approach. In what follows, we shall assume this argument has already been made and that $\hat{\beta}_n \to_p \beta^0$ as $n \to \infty$.

Our concern is with the asymptotic distribution of $\hat{\beta}_n$. Our approach is to proceed as if the problem were regular and the objective function were differentiable in β. Although the derivatives do not exist in the usual sense, they do have a meaning as generalized functions. Moreover, they are real valued and unique when they appear in an appropriate linear functional form. Since this is precisely how they do arise in the limit as $n \to \infty$, it turns out that we may proceed with the usual Taylor series expansion of the first order conditions to extract the asymptotic theory.

We start with the first order conditions for $\hat{\beta}_n$ from (3). These are

$$(4) \qquad n^{-1}\Sigma_1^n \, \text{sgn}(y_t - x_t'\hat{\beta}_n)x_t = 0 \, ,$$

where $\text{sgn}(X) = -1$ for $X < 0$ and $= 1$ for $X \geq 0$. We now expand (4) in a Taylor series about its value at β^0. Note that $d/dX(\text{sgn}(X)) = 2\delta(X)$ where $\delta(X)$ is the delta (generalized) function (Gelfand and Shilov (1964), hereafter simply GS, p. 4). We denote successive derivatives of the delta function by $\delta^{(k)}(X)$ for $k = 1, 2, \ldots$. Proceeding in a purely formal way by treating $\text{sgn}(\cdot)$ as analytic (it is already piecewise analytic) and its derivatives as ordinary functions, we would have the expansion

$$(5) \qquad 0 = n^{-1}\Sigma_1^n \, \text{sgn}(u_t)x_t - 2n^{-1}\Sigma_1^n \delta(u_t)x_t x_t'(\hat{\beta}_n - \beta^0)$$
$$+ 2\Sigma_{k=2}^{\infty}(-1)^k(1/k!)n^{-1}\Sigma_1^n \delta^{(k-1)}(u_t)x_t\left[x_t'(\hat{\beta}_n - \beta^0)\right]^k \, .$$

Let us now suppose that we can ignore all but the first two terms of (5) as $n \to \infty$. Scaling by $n^{1/2}$ and taking the error to be $o_p(1)$, we have

$$(6) \qquad n^{-1/2}\Sigma_1^n \, \text{sgn}(u_t)x_t = 2\left[n^{-1}\Sigma_1^n \delta(u_t)x_t x_t'\right]n^{1/2}(\hat{\beta}_n - \beta^0) + o_p(1) \, .$$

Since $(\text{sgn}(u_t))$ is iid(0,1), the left side of (6) satisfies a multivariate extension of the Lindeberg–Lévy theorem, leading to

$$(7) \qquad n^{-1/2}\Sigma_1^n \ sgn(u_t)x_t \rightarrow_d N(0,Q) \ .$$

On the right hand side of (6) the matrix factor in square brackets satisfies a weak law of large numbers, viz.

$$(8) \qquad n^{-1}\Sigma_1^n \delta(u_t)x_t x_t' \rightarrow_p \lim_{n\rightarrow\infty} n^{-1}\Sigma_1^n E(\delta(u_t))x_t x_t' = f(0)Q \ .$$

Putting (6), (7) and (8) together we deduce directly the limit theory for the LAD estimator, i.e.

$$(9) \qquad n^{1/2}(\hat{\beta}_n - \beta^0) \rightarrow_d N\left[0, (1/2 \ f(0))^2 Q^{-1}\right] \ .$$

3. AN ATTEMPT AT RIGOR

To attach some rigor to this heuristic skeleton, we need to justify three of the steps just taken. Specifically, (i) the Taylor expansion (5); (ii) the $o_p(1)$ error in (6); and (iii) the weak law of large numbers (8).

Let us start with (i). Obviously, (5) has no meaning as an ordinary equation or as an ordinary Taylor series expansion. But it can be interpreted in terms of generalized functions and as a generalized Taylor series. Thus, suppose φ is a suitable test function for linear functionals of a generalized function g . Suppose, for instance, that φ belongs to the space S of entire functions which, together with their derivatives, approach zero more rapidly than any power of $1/|u|$ as $|u| \rightarrow \infty$ (e.g. e^{-u^2}) and that the linear functional

$$(10) \qquad (g(u+h), \ \varphi) = \int_{\mathbb{R}} g(u+h)\varphi(u)du = \int_{\mathbb{R}} g(s)\varphi(s-h)ds$$

is an ordinary analytic function of h in some neighborhood of $h = 0$ for all φ . Then, $g_h(u) = g(u+h)$ is a generalized analytic function of h (GS, pp. 149–150). Indeed, by expanding $\varphi(s-h)$ about its value at $h = 0$ in (10) and noting that $(g^{(j)}, \varphi) = (g, (-1)^j\varphi^{(j)})$, we obtain

$$(g(u+h), \ \varphi) = \Sigma_{j=0}^\infty (1/j!)(-h)^j(g, \ \varphi^{(j)}) = \Sigma_{j=0}^\infty (1/j!)h^j(g^{(j)}, \ \varphi)$$

which we can write in formal terms as

$$g_h = \Sigma_{j=0}^{\infty}(1/j!)g_h^{(j)}h^j \ .$$

Setting $g_h(u) = \mathrm{sgn}(u+h)$, $h = x_t'(\hat{\beta}_n - \beta^0)$ and noting that $\int_{-\infty}^{\infty}\mathrm{sgn}(u+h)\varphi(u)du$ is

analytic in h for all $\varphi \in S$, we deduce the expansion given in (5) above.

Next consider (iii). The limit given on the right of (8) is well defined because the generalized function $\delta(\cdot)$ arises only through the linear functional $E(\delta(u)) = \int\delta(u)f(u)du$ $= f(0)$. Note that this last expression remains true even though the density $f(\cdot)$ may exist only in a neighborhood of zero. This is because generalized functions like $\delta(\cdot)$ may be defined locally in terms of their operation on test functions with support in arbitrarily small given neighborhoods of every point (see GS, p. 140).

To be more complete in deducing (8) we may replace $\delta(\cdot)$ with its inverse Fourier transform representation, which we signify by " $F^{-1}(\cdot)$ ", i.e.

$$\delta(u_t) = F^{-1}(1) = (2\pi)^{-1}\int_{-\infty}^{\infty}e^{-iu_t\sigma}d\sigma \ .$$

Again, the integral is formal and the correspondence produces the Fourier transform pair of generalized functions $(\delta(\cdot), 1)$ (GS, p. 168). In place of (8), we may now show that

$$(11) \qquad n^{-1}\Sigma_1^n e^{-iu_t\sigma}x_t x_t' \to_p \lim_{n\to\infty} n^{-1}\Sigma_1^n E(e^{-iu_t\sigma})x_t x_t' = cf_u(-\sigma)Q \ .$$

But this follows immediately because $(e^{-iu_t\sigma} - E(e^{-iu_t\sigma}))$ is iid with zero mean and finite variance. Upon inversion of (11) we get (8), since $f(0) = (1/2\pi)\int_{-\infty}^{\infty}cf_u(\sigma)d\sigma$. In fact, (11) may be regarded as the appropriate way to interpret (8) as a weak law for generalized random variables.

This leaves us with (ii). Working from (5) we have

$$(12) \qquad n^{-1/2}\Sigma_1^n \mathrm{sgn}(u_t)x_t = 2A_n n^{1/2}(\hat{\beta}_n - \beta^0) \ ,$$

where

$$A_n = n^{-1}\Sigma_1^n \delta(u_t)x_t x_t' + \Sigma_{j=1}^{\infty}(-1)^{j+1}((j+1)!)^{-1}n^{-1}\Sigma_1^n \delta^{(j)}(u_t)x_t x_t' \left[x_t'(\hat{\beta}_n - \beta^0)\right]^j$$

and we have to show that the second term of A_n is $o_p(1)$ as $n \to \infty$.

As with (8) we have

$$n^{-1}\Sigma_1^n \delta^{(k)}(u_t)x_t x_t' \to_p \lim_{n \to \infty}\left[n^{-1}\Sigma_1^n E(\delta^{(k)}(u_t))x_t x_t'\right] = (-1)^k f^{(k)}(0)Q ,$$

where we use the fact that

$$(\delta^{(k)}, f) = \int_{-\infty}^{\infty} \delta^k(u)f(u)du = (\delta, (-1)^k f^{(k)}) = (-1)^k f^{(k)}(0) ,$$

(GS, p. 26). Moreover, since $f(\cdot)$ is analytic at zero, power series such as

$$\Sigma_{k=2}^{\infty}(1/k!)f^{(k)}(0)\epsilon^{k-1} = (f(\epsilon) - f(0) - f'(0)\epsilon)/\epsilon$$

are convergent and of $O(\epsilon)$ for all ϵ in the vicinity of zero. However, $\hat{\beta}_n \to_p \beta^0$ and $x_t'(\hat{\beta}_n - \beta^0) \to_p 0$ uniformly in t. It follows that the second term of A_n converges in probability to zero as required.

4. AN ASYMPTOTIC EXPANSION

One advantage of the above approach is that it lends itself to the development of higher order asymptotic expansions. To see how to proceed we set $q_n = n^{1/2}(\hat{\beta}_n - \beta^0)$, $\ell_1 = n^{-1/2}\Sigma_1^n \text{sgn}(u_t)x_t$, scale the expansion (5) by $n^{1/2}$ and write it in the form

$$(13) \qquad 0 = \ell_{1i} + \ell_{2ij}q_{nj} + n^{-1/2}\ell_{3ijk}q_{nj}q_{nk} + n^{-1}\ell_{4ijk}q_{nj}q_{nk}q_{nm} + o_p(n^{-1}) ,$$

where we use the summation convention. Inverting (13), we have up to $O_p(n^{-1/2})$

$$(14) \qquad q_{ni} = \ell_2^{ij}\ell_{1j} + n^{-1/2}\ell_2^{ij}\ell_{3jk\ell}(\ell_2^{km}\ell_{1m})(\ell_2^{\ell n}\ell_{1n}) + O_p(n^{-1}) .$$

Next observe that the distribution of $\text{sgn}(u_t)$ is symmetric, that the distribution of ℓ_1 will admit a valid Edgeworth expansion and that, because of the symmetry of the distribution of $\text{sgn}(u_t)$, there will be no skewness term in this expansion. Thus, we may write

symbolically

$$\ell_1 =_d N(0,Q) + O_p(n^{-1}) = \underset{\sim}{\ell}_1 + O_p(n^{-1}) \, , \text{ say.}$$

The second term of (14) is $O_p(n^{-1/2})$ and

$$\ell_{3jk\ell} = n^{-1}\Sigma_1^n \delta'(u_t)x_{tj}x_{tk}x_{t\ell} \rightarrow_p (-1)f'(0)r_{jk\ell} = \underset{\sim}{\ell}_{3jk\ell} \, , \text{ say,}$$

where

$$r_{jk\ell} = \lim_{n\to\infty} \frac{1}{n}\Sigma_1^n x_{tj}x_{tk}x_{t\ell} \, .$$

Thus (14) would *appear* to yield a conventional Edgeworth expansion for the distribution of q_n. However, there is an additional complication that arises from the components ℓ_{2ij} and the elements ℓ_2^{ij} of its inverse. We have

$$\ell_{2ij} = 2n^{-1}\Sigma_1^n \delta(u_t)x_{ti}x_{tj} \rightarrow_p 2f(0)q_{ij} = \underset{\sim}{\ell}_{2ij} \, , \text{ say.}$$

Similarly, define $\underset{\sim}{\ell}_2^{ij} = (1/2\, f(0))q^{ij}$. Then (14) may be written as

$$(15) \qquad q_{ni} = \underset{\sim}{\ell}_2^{ij}\ell_{1j} + (\ell_2^{ij} - \underset{\sim}{\ell}_2^{ij})\ell_{1j} + n^{-1/2}\underset{\sim}{\ell}_2^{ij}\underset{\sim}{\ell}_{3jk\ell}(\underset{\sim}{\ell}_2^{km}\ell_{1m})(\underset{\sim}{\ell}_2^{\ell m}\ell_{1m}) + o_p(n^{-1/2}) \, .$$

The order of magnitude of the second term of (15) depends on that of

$$\ell_{2ij} - \underset{\sim}{\ell}_{2ij} = 2n^{-1}\Sigma_1^n(\delta(u_t) - f(0))x_t x_t' + o(1) \, .$$

This term is much more difficult to analyze and is larger than $O_p(n^{-1/2})$.

To see what is involved, set $x_t = 1$ and consider

$$(16) \qquad n^{-1}\Sigma_1^n(\delta(u_t) - f(0)) = \int_{\mathbb{R}} \delta(u)d(F_n(u) - F(u)) \, ,$$

where F_n is the empirical distribution function of u_t and F is the cdf of u_t. In general, we have the weak convergence (Billingsley, 1968, p. 141)

$$(17) \qquad \sqrt{n}(F_n(u) - F(u)) \rightarrow_d Y(u) \, ,$$

where $Y(\cdot)$ is a Gaussian process with covariance kernel $F(u)(1 - F(u))$, $u \leq v$. However, we cannot employ (17) in (16) because the implied limit variate, viz.

$\int_{\mathbb{R}} \delta(u) dY(u)$, does not exist. For instance, if u_t were uniformly distributed then $Y(u)$ would be a Brownian bridge process and $\int_{\mathbb{R}} \delta(u) dY(u)$ would be its "derivative," which, like the derivative of Brownian motion, does not exist as an ordinary random variable.

There is another way to proceed. Note that (16) is the derivative at $s = 0$ of

$$n^{-1}\Sigma_1^n((1/2)\text{sgn}(u_t+s) - F(s)) - n^{-1}\Sigma_1^n((1/2)\text{sgn}(u_t) - F(0))$$
$$= \{F_n(0) - F_n(-s)\} - \int_0^s dF .$$

For $s = -tn^{-1/3}$ this expression is

$$-\{F_n(tn^{-1/3}) - F_n(0) - f(0)tn^{-1/3} - (1/2)f'(0)t^2 n^{-2/3}\} + o(n^{-2/3})$$
$$= -X_n(t) + o(n^{-2/3}) , \text{ say.}$$

Now Kim and Pollard (1990, theorem 4.7 and example 6.5, p. 216) show that

$$(18) \qquad n^{2/3}\{F_n(tn^{-1/3}) - F_n(0) - f(0)tn^{-1/3}\} \to_d (1/2)t^2 f'(0) + f(0)^{1/2} W(t)$$

with $W(t)$ a two–sided Brownian motion. Thus,

$$n^{2/3}X_n(t) \to_d f(0)^{1/2}W(t) .$$

Define $\hat{f}_n(0)$ to be the left derivative of the concave majorant of F_n (i.e. the smallest concave function on $[0,\infty)$ that is everywhere greater than or equal to F_n). Kim and Pollard further show that $n^{1/3}(\hat{f}_n(0) - f(0))$ has a limit distribution given by that of the slope at the origin of the concave majorant of Brownian motion with quadratic drift, a result that is originally due to Prakasa Rao (1969). Treating $\tilde{f}_n(0) = n^{-1}\Sigma_1^n \delta(u_t)$ as an estimate of $f(0)$, these results suggest that (16) and, hence, $\ell_{2ij} - \tilde{\ell}_{2ij}$ are both $O_p(n^{-1/3})$. But the second term in the expansion (15) involves the product $(\ell_2^{ij} - \tilde{\ell}_2^{ij})\ell_{1j}$, where $\ell_1 =_d N(0,Q)$. Because ℓ_1 has zero mean this term will contribute to the asymptotic expansion of the distribution of q_n only through the variance of ℓ_1 and, hence, will produce an adjustment of $O(n^{-2/3})$. It follows that only the first and third members on the right side of (15) contribute to the expansion up to $O(n^{-1/2})$. In

particular, setting $\xi = \underline{\zeta}_1 =_d N(0,Q)$ we have

$$(19) \quad q_{ni} = (1/2\ f(0))q^{ij}\xi_j - n^{-1/2}(1/2\ f(0))^3 f'(0)q^{ij}r_{jk\ell}(q^{km}\xi_m)(q^{\ell m}\xi_m) + o_p(n^{-1/2})\ .$$

Observe that when $f'(0) = 0$, which will be the case for symmetric error distributions, the term of $O_p(n^{-1/2})$ itself drops out, leaving only the first order asymptotic term. In the general case, we can derive an asymptotic expansion of the density of some linear combination such as $r = c'q_n = c_i q_{ni}$ of the standardized error vector q_n. The formulae for the expansion of the density up to $O(n^{-1/2})$ may then be deduced from those in the literature (e.g. Sargan (1976) and Phillips (1977, 1982), where the last reference puts them in a form that is especially simple to interpret). With a little algebra, we obtain the following explicit expansion to $O(n^{-1/2})$:

$$\text{pdf}(r) = (1/\omega)\varphi(r/\omega)\Big[1 + n^{-1/2}\{a_1(r/\omega) + a_3(r/\omega)^3\}\Big] + o(n^{-1/2})\ ,$$

where $\varphi(x) = (2\pi)^{-1/2}\exp(-(1/2)x^2)$ and the parameters are:

$$\omega^2 = (1/2\ f(0))^2 c_i q^{ij} c_j\ ,$$

$$a_1 = (1/2\ \omega)q_{st}f_{st} - 3a_3\ ,$$

$$a_3 = (1/2\ \omega^3)(1/2\ f(0))^2 f_{st}c_s c_t\ ,$$

$$f_{st} = -2(1/2\ f(0))^3 f'(0)c_i q^{ij}r_{jk\ell}q^{ks}q^{\ell t}\ .$$

5. ESTIMATING THE COVARIANCE MATRIX

In conventional Taylor expansions of the first order conditions, the Hessian is often used to produce an estimate of the asymptotic covariance matrix of the estimator. The same idea may be applied here. Taking the dominant member of the "Hessian," $2A_n$, in the expansion (12) to be $2n^{-1}\Sigma_1^n \delta(u_t)x_t x_t'$ we simply replace the delta function in this expression by a *delta sequence* $\delta_m(\cdot)$ for which

$$(20) \qquad \lim_{m\to\infty} \int_{-\infty}^{\infty} \delta_m(x)\varphi(x)dx = \varphi(0)$$

for all $\varphi \in S$. For example, the delta sequence $\delta_m(x) = (m/\pi)^{1/2} e^{-mx^2}$ satisfies (20) and corresponds to a density estimate based on a normal kernel. The covariance matrix estimate arising from this delta sequence takes the form

$$2n^{-1} \Sigma_1^n \delta_m(u_t) x_t x_t' \ .$$

The errors, u_t, in this expression can be replaced by residuals to produce a feasible covariance matrix estimate. The parameter m is like a bandwidth (for the normal kernel given above we would have $h = 1/m^{1/2}$) and must be chosen so that $m \to \infty$, $nm^{-1/2} \to \infty$ as $n \to \infty$ (i.e. $h \to 0$, $nh \to \infty$) for consistency (e.g. Silverman, 1986, p. 71).

Since delta sequences encompass most density estimates (e.g. Walter and Blum, 1979), this approach to the estimation of the covariance matrix is really quite general.

6. MODEL EXTENSIONS

It would appear that the approach suggested here remains valid for a large class of models with weakly dependent, rather than iid, errors. All that is required is that $\ell_1 = n^{-1/2} \Sigma_1^n \, \text{sgn}(u_t) x_t$ satisfy some central limit theorem for weakly dependent errors. For example, suppose that u_t satisfies

(A_2) *The sequence* (u_t) *is strictly stationary and strong mixing with mixing coefficients* β_k *that satisfy* $\Sigma_1^\infty \beta_k < \infty$, *and* u_t *has zero median and probability density* $f(\cdot)$ *that is positive and analytic at zero.*

Then $\text{sgn}(u_t)$ is also strong mixing with mixing coefficients that satisfy the same summability condition and we have the central limit theorem

$$n^{-1/2} \Sigma_1^n \, \text{sgn}(u_t) x_t \to_d N(0,V) \ ,$$

where

$$V = \Sigma_{k=-\infty}^\infty R_k E\{\text{sgn}(u_0)\text{sgn}(u_k)\} \ , \text{ with } R_k = \lim_{n\to\infty} n^{-1} \Sigma_1^n x_t x_{t+k}' \ .$$

When $x_t = 1$ we have

$$V = E(\text{sgn}(u_0))^2 + 2\Sigma_{k=1}^{\infty} E(\text{sgn}(u_0)\text{sgn}(u_k))$$
$$= 1 + 2\Sigma_{k=1}^{\infty} [\{P(u_0 < 0, u_k < 0) + P(u_0 > 0, u_k > 0)\}$$
$$- \{P(u_0 > 0, u_k < 0) + P(u_0 < 0, u_k > 0)\}] .$$

Using the same expansion (5) as before, we deduce the following limit theory for the LAD estimator

$$n^{1/2}(\hat{\beta}_n - \beta^0) \to_d N(0, (1/2 \, f(0))^2 Q^{-1} V Q^{-1}) .$$

For inference V, as well as $f(0)$, now needs to be estimated. The situation is entirely analogous to the estimation of autocorrelation consistent covariance matrix estimation as it arises in conventional regression contexts (cf. White (1984), Newey and West (1987)).

Models in which the carrier variables x_t are random may also be accommodated. The details are close to those given in Pollard (1989a) so we shall only touch on them briefly here. Suppose the x_t are strictly stationary, ergodic, square integrable and $\mathcal{F}_{t-1}$ measurable for some increasing sequence of σ–fields $\mathcal{F}_t$, suppose u_t is independent of $\mathcal{F}_{t-1}$ and satisfies $(\mathcal{A}_1)$ and let $V_n = \Sigma_1^n E(z_t z_t') = nE(x_t x_t') = n\Omega$, where $z_t = \text{sgn}(u_t)x_t$. Then the following two conditions hold:

$$\max_{t \leq n} \|V_n^{-1/2} z_t\| = \max_{t \leq n} \|\Omega^{-1/2}(n^{-1/2} z_t)\| \to_p 0 ,$$
$$V_n^{-1/2}(\Sigma_1^n z_t z_t')V_n^{-1/2} = \Omega^{-1/2}(n^{-1}\Sigma_1^n z_t z_t')\Omega^{-1/2} \to_p I_k$$

and we may use a martingale central limit theorem (e.g. Hall and Heyde (1980), theorem 3.2, p. 58) to establish that

$$V_n^{-1/2}\Sigma_1^n \, \text{sgn}(u_t)x_t' \to_d N(0,I) .$$

Moreover,

$$n^{-1}\Sigma_1^n e^{-iu_t \sigma} x_t x_t' \to_p E(e^{-iu_t \sigma} x_t x_t') = cf_u(-\sigma)\Omega$$

and so

$$n^{-1}\Sigma_1^n \delta(u_t)x_t x_t' \to_p f(0)\Omega ,$$

where $f(\cdot)$ is the probability density of u_t. These results enable us to deduce the limit law

$$n^{1/2}(\hat{\beta}_n - \beta^0) \to_d N(0, (1/2\, f(0))^2 \Omega^{-1})$$

in the same way as before. This covers the case of strictly exogenous stationary regressors x_t with $\Omega = E(x_t x_t')$. It also covers the case of stable autoregressions with $x_t' = (y_{t-1}, \ldots, y_{t-k})$ with $(u_t) = iid(0, \sigma^2)$ (the finite variance σ^2 ensures that the elements of x_t are square integrable) and $\Omega = (\omega_{ij})$ with $\omega_{ij} = E(y_{t-i} y_{t-j})$.

7. AUTOREGRESSIONS WITH INFINITE VARIANCE ERRORS

An autoregression with infinite variance errors is quite difficult to analyze by conventional methods and has so far resisted a complete study. An and Chen (1982) and Pollard (1989a) considered the prototypical case of an AR(1) with iid Cauchy errors. Pollard showed that the LAD estimator $\hat{\beta}_n$ lies within $O_p(n^{-1})$ of β^0, improving slightly on the $O_p(n^{-1+\epsilon})$, $\epsilon > 0$, result of An and Chen. Pollard was unable to find the limit distribution of $n(\hat{\beta}_n - \beta^0)$ and left this as an unsolved problem. Davis, Knight and Liu (1990) provide a solution to this problem as an optimization functional in a recent paper that came to the author's attention after the present work was in its second draft. The heuristic development that follows provides an alternative solution to the problem.

We shall consider the stable AR(1)

(21) $\qquad y_t = \beta^0 y_{t-1} + u_t$, with $|\beta^0| < 1$

and iid errors with infinite variance. To develop a limit theory we shall assume that u_t is in the domain of attraction of a stable law with characteristic exponent α. Specifically we say that $u \in \mathcal{D}(\alpha)$ if

(22) $\qquad P(u > x) = c_1 x^{-\alpha} L(x)(1 + \alpha_1(x))$, $x > 0$, $c_1 \geq 0$

and

$$(23) \qquad P(u < -x) = c_2 x^{-\alpha} L(x)(1 + \alpha_2(x)) \, , \quad x > 0 \, , \quad c_2 \geq 0$$

with $0 < \alpha < 2$, $L(x)$ a slowly varying function at ∞ and $\alpha_i(x) \to 0$ as $|x| \to \infty$. If $L(x) = 1$ in (22) and (23), then u is in the normal domain of attraction of a stable law with parameter α and we write $u \in \mathit{ND}(\alpha)$. With this terminology in hand, we assume

(A_3) *The sequence* (u_t) *is iid with* $u_t \in \mathcal{D}(\alpha)$, *where* $0 < \alpha < 2$, *with zero median and with probability density* $f(\cdot)$ *that is positive and analytic at zero. If* $\alpha > 1$, $E(u_t) = 0$ *and if* $\alpha = 1$ *then* $u_t =_d -u_t$ *(i.e.* u_t *is symmetrically distributed).*

Define the normalizing sequence

$$a_n = \inf\{x : P(|u| > x) \leq n^{-1}\} \, .$$

For $u \in \mathcal{D}(\alpha)$ we have $a_n = n^{1/\alpha} L(n)$ for some slowly varying function $L(n)$. For $u \in \mathit{ND}(\alpha)$ we have $a_n = c n^{1/\alpha}$ for some constant $c > 0$. With this construction we have the following weak convergence results for $0 < \alpha < 2$:

$$(24) \qquad a_n^{-1} \Sigma_1^n u_t \to_d U_\alpha(1) \, , \quad a_n^{-1} \Sigma_1^{[nr]} u_t \to_d U_\alpha(r) \, ,$$

$$(25) \qquad (a_n^{-1} \Sigma_1^{[nr]} u_t, \, a_n^{-2} \Sigma_1^{[nr]} u_t^2) \to_d (U_\alpha(r), \int_0^r (dU_\alpha)^2) \, .$$

Here $U_\alpha(r)$ is a Lévy α–stable process and $\int_0^r (dU_\alpha)^2 = [U_\alpha]_r$ is its quadratic variation process. The first result of (24) is classical (e.g. Ibragimov and Linnik (1971), Ch. 2); the second is its functional version; and (25) is a joint functional limit law for first and second sample moments that is proved in Resnick (1986, pp. 94–95).

To find the limit distribution of the LAD estimator we work from expansion (5), setting $x_t = y_{t-1}$. We scale (5) by $n a_n^{-1}$, giving

$$(26) \qquad a_n^{-1} \Sigma_1^n \, \mathrm{sgn}(u_t) y_{t-1} = 2[a_n^{-2} \Sigma_1^n \delta(u_t) y_{t-1}^2] a_n(\hat{\beta}_n - \beta^0)$$

$$- 2 \Sigma_{k=2}^\infty (-1)^k (1/k!)[a_n^{-k-1} \Sigma_1^n \delta^{(k-1)}(u_t) y_{t-1}^{k+1}] \left[a_n(\hat{\beta}_n - \beta^0) \right]^k$$

Observe that $\mathrm{sgn}(u_t)$ and y_{t-1} are independent, $\mathrm{sgn}(u_t)$ is bounded and the product variate $X_t = \mathrm{sgn}(u_t) y_{t-1} \in \mathcal{D}(\alpha)$ —see Appendix A of Phillips (1990). Note also that

u_t , y_{t-1} and X_t all have the same normalizing sequence a_n , so that by (24) we have

$$a_n^{-1} \Sigma_1^n \, \text{sgn}(u_t) y_{t-1} \to_d S_\alpha(1) \; ,$$

where $S_\alpha(r)$ is an α–stable process. The right side of (26) may be treated term by term. First observe that $y_{t-1}^2 = X_t^2 \in \mathcal{D}(\alpha/2)$ and by (25)

$$a_n^{-2} \Sigma_1^n y_{t-1}^2 \to_d \int_0^1 (dS_\alpha)^2 \; .$$

However, our interest is in the factor in square brackets in the first term on the right side of (26), and this involves the generalized random variable $\delta(u_t)$. Since u_t is independent of y_{t-1}^2 , the effect of $\delta(u_t)$ in the summation is to randomly select elements that will remain in the summation according to subsequent realizations of the independent draw u_t . The remaining elements in the sum are weighted according to $E(\delta(u_t))$ $= E(\delta(u_t)|\mathcal{F}_{t-1}) = f(0)$ and we obtain

$$a_n^{-2} \Sigma_1^n \delta(u_t) y_{t-1}^2 \to_d f(0) \int_0^1 (dR_\alpha)^2 \; ,$$

where $R_\alpha(r)$ is a Lévy α–stable process independent of $S_\alpha(r)$. Similarly, for the higher order terms on the right side of (26) we have $E(\delta^{(k-1)}(u_t)) = E(\delta^{(k-1)}(u_t)|\mathcal{F}_{t-1})$ $= (-1)^{k-1} f^{(k-1)}(0)$ and

$$a_n^{-k-1} \Sigma_1^n \delta^{(k-1)}(u_t) y_{t-1}^{k+1} \to_d (-1)^{k-1} f^{(k-1)}(0) \int_0^1 (dR_\alpha)^{k+1} \; .$$

This leads to the following limit theory for the LAD estimator of β^0 in (21) under errors that satisfy $(\mathcal{A}_3)$

$$(27) \qquad a_n(\hat{\beta}_n - \beta^0) \to_d \xi$$

where the limit variate ξ is a solution of the equation

$$S_\alpha(1) = 2\{f(0) \int_0^1 (dR_\alpha)^2\} \xi + 2\Sigma_{k=2}^\infty \{(1/k!) f^{(k-1)}(0) \int_0^1 (dR_\alpha)^{k+1}\} \xi^k \; .$$

In the above expression $\int_0^t (dR_\alpha)^s$ is the s–variation process of R_α $s \geq 2$. Observe that for distributions $f(\cdot)$ for which $f^{(j)}(0) = 0$, $\forall j \geq 1$, we have the simple result that the

limit variate is

$$\xi = S_\alpha(1)/2f(0)\int_0^1 (dR_\alpha)^2 \ .$$

For example, if (v_1, v_2) are independent uniform on $[-a, a]$ for some fixed $a > 0$, then $u = v_1/v_2$ has density

$$f(u) = \begin{cases} 1/4 & |u| \leq 1 \\ 1/4u^2 & |u| > 1 \end{cases} .$$

Such a density obviously satisfies $(\mathcal{A}_3)$ and $f^{(j)}(0) = 0$, $\forall j \geq 1$.

It is of some interest to compare the LAD estimator $\hat{\beta}_n$ with the OLS estimator $\tilde{\beta}_n$. Bloomfield and Steiger (1983, p. 105) speculate that $\hat{\beta}_n$ is asymptotically more efficient than $\tilde{\beta}_n$ and conjecture that $\hat{\beta}_n$ converges at a faster rate for the case where $0 < \alpha < 2$. This can, in fact, be demonstrated by considering the limit distribution of $\tilde{\beta}_n$. Let us suppose that $u \in \mathcal{ID}(\alpha)$, as it is in the cases studied experimentally by Bloomfield and Steiger. Then, $u_t y_{t-1} \in \mathcal{D}(\alpha)$ rather than $\mathcal{ID}(\alpha)$ and the norming sequence for sums of $u_t y_{t-1}$ has the form

$$b_n = c(n \log n)^{1/\alpha}, \quad c = \text{const.}$$

(see Appendix A of Phillips (1990)). Since $\tilde{\beta}_n - \beta^0 = \Sigma y_{t-1} u_t / \Sigma y_{t-1}^2$, it follows that

$$(28) \quad (n/\log n)^{1/\alpha}(\tilde{\beta}_n - \beta^0) = \left[a_n^{-2} \Sigma_1^n y_{t-1}^2 \right]^{-1} \left[(n \log n)^{-1/\alpha} \Sigma_1^n y_{t-1} u_t \right] \to_d P_{\alpha/2}^{-1} Q_\alpha \,,$$

where Q_α is a stable variate of exponent α, independent of the positive stable variate $P_{\alpha/2} = \int_0^1 (dS_\alpha)^2$. Comparing (28) with (27) we see that the LAD estimator has a factor of $(\log n)^{1/\alpha}$ in its favor, confirming the conjecture of Bloomfield and Steiger concerning the higher relative efficiency of LAD. The difference arises because, $u_t y_{t-1} \in \mathcal{D}(\alpha)$ and sample covariances of u_t and y_{t-1} form the numerator of $\tilde{\beta}_n - \beta^0$, whereas $\text{sgn}(u_t)y_{t-1} \in \mathcal{ID}(\alpha)$ because $\text{sgn}(u_t)$ is bounded and the LAD estimator depends on sample covariances of $\text{sgn}(u_t)$ and y_{t-1}. The difference in rates of convergence of $\hat{\beta}_n$

and $\tilde{\beta}_n$ to β^0 is then due to the differences in the tail thickness of the distributions of $u_t y_{t-1}$ and $sgn(u_t)y_{t-1}$. By bounding the effect of the error through the presence of the $sgn(u_t)$ factor in $sgn(u_t)y_{t-1}$, the LAD estimator successfully achieves a faster rate of convergence. This helps to explain the simulation results of Bloomfield and Steiger.

8. AUTOREGRESSIONS WITH INFINITE VARIANCE ERRORS AND A UNIT ROOT

The case $\beta^0 = 1$ in (21) has recently been studied by Knight (1989). It may also be treated by our approach here. When $\beta^0 = 1$, the output y_t of (21) is an integrated process. In consequence, we have the weak convergence

$$(29) \qquad a_n^{-1} y_{[n \cdot]} \to_d S_\alpha(\cdot) ,$$

where $S_\alpha(\cdot)$ is an α–stable process. Further we have

$$n^{-1/2} \Sigma_1^{[n \cdot]} sgn(u_t) \to_d W(\cdot)$$

and

$$(30) \qquad n^{-1/2} a_n^{-1} \Sigma_1^n sgn(u_t) y_{t-1} = n^{-1/2} \Sigma_1^n (a_n^{-1} y_{t-1}) sgn(u_t) \to_d \int_0^1 S_\alpha^- dW ,$$

where $W(\cdot)$ is standard Brownian motion and $S_\alpha^-(r)$ signifies the left limit of the process $S_\alpha(\cdot)$ at r. Next observe that

$$n^{-1} a_n^{-2} \Sigma_1^n y_{t-1}^2 = n^{-1} \Sigma_1^n \left[a_n^{-1} y_{t-1} \right]^2 \to_d \int_0^1 S_\alpha^2$$

and

$$n^{-1} \Sigma_1^n (\delta(u_t) - f(0)) \left[a_n^{-1} y_{t-1} \right]^2 \to_p 0$$

so that

$$(31) \qquad n^{-1} \Sigma_1^n \delta(u_t) \left[a_n^{-1} y_{t-1} \right]^2 \to_d f(0) \int_0^1 S_\alpha^2 .$$

Combining (30) and (31) we have

$$(32) \qquad n^{1/2}a_n(\hat{\beta}_n - 1) \sim \frac{n^{-1/2}a_n^{-1}\Sigma_1^n \ \text{sgn}(u_t)y_{t-1}}{2n^{-1}a_n^{-2}\Sigma_1^n \delta(u_t)y_{t-1}^2} \to_d \frac{\int_0^1 S_\alpha^- dW}{2f(0)\int_0^1 S_\alpha^2},$$

as given in Theorem 3 of Knight (1989). Since $a_n = n^{1/\alpha}L(n)$ we have the remarkable result, due to Knight, that the LAD estimator converges at a faster rate in the unit root model for $0 < \alpha < 2$ than the OLS estimator. As remarked by Knight, robust estimators such as LAD retain the advantages of the strong signal from y_{t-1} (due to integration and thick tailed errors) but alternate the effects of outliers in the error u_t to the extent that they occur in the sample covariance between $\text{sgn}(u_t)$ and y_{t-1} (in the LAD case).

9. CONCLUSION

The methods outlined here seem to offer some promise as tools for the analysis of regression asymptotics in non—regular cases like that of the LAD estimator. Our approach has not always been rigorous and in Sections 7 and 8, in particular, is quite heuristic. Nevertheless, it is hoped that the results obtained point to the usefulness of the approach. One would like to hope that the approach can be made entirely rigorous by providing a tight probabilistic framework for the use of the generalized random variables that appear here. In the meantime it seems reasonable to conclude that the classical approach warrants more attention than it has yet received. It may indeed offer some advantages that empirical process methods do not seem to presently enjoy, viz. the capacity to develop higher order asymptotics.

10. REFERENCES

Amemiya, T. (1985). *Advanced Econometrics*. Cambridge: Harvard University Press.

An, H–Z. and Z–G Chen (1982). "On convergence of LAD estimates in autoregression with infinite variance," *Journal of Multivariate Analysis*, 12, 335–345.

Andrews, D. W. K. (1989). "Asymptotics for semiparametric econometric models: I. Estimation," Cowles Foundation Discussion Paper No. 908.

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

Billingsley, P. (1968). *Convergence of Probability Measures*. New York: Wiley.

Bloomfield, P. and W. L. Steiger (1983). *Least Absolute Deviations: Theory, Applications and Algorithms*. Boston: Birkhauser.

Cramér, H. (1946). *Mathematical Methods of Statistics*. Princeton: Princeton University Press.

Daniels, H. E. (1961). "The asymptotic efficiency of a maximum likelihood estimator," *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 1. Berkeley: University of California Press.

Davis, R. A., K. Knight and J. Liu (1990). "M–estimation for autoregressions with infinite variance," mimeographed.

Gelfand, I. M. and G. E. Shilov (1964). *Generalized Functions*, Vol. 1. New York: Academic Press.

Hall, P. and C. C. Heyde (1980). *Martingale Limit Theory and its Application*. New York: Academic Press.

Huber, P. J. (1967). "The behaviour of maximum likelihood estimates under nonstandard conditions," *Fifth Berkeley Symposium on Mathematical Statistics and Probability*. Berkeley: University of California Press.

Ibragimov, I. A. and Y. V. Linnik (1971). *Independent and Stationary Sequences of Random Variables*. Groningen: Wolters–Noordhof.

Kim, J. and D. Pollard (1990). "Cube root asymptotics," *Annals of Statistics*, 18, 191–219.

Knight, K. (1989). "Limit theory for autoregressive parameters in an infinite variance random walk," *Canadian Journal of Statistics*, 261–278.

McFadden, D. (1989). "A method of simulated moments for estimation of discrete response models without numerical integration," *Econometrica*, 57, 995–1026.

Newey, W. K. and K. D. West (1987). "A simple positive definite heteroskedasticity and autocorrelation consistent covariance matrix," *Econometrica*, 55, 703–708.

Pakes, A. and D. Pollard (1989). "Simulation and the asymptotics of optimization estimators," *Econometrica*, 57, 1027–1058.

Phillips, P. C. B. (1977). "Approximations to some finite sample distributions associated with a first order stochastic difference equation," *Econometrica*, 45, 463–486.

_____ (1982). "Small sample distribution theory in econometric models of simultaneous equations," Cowles Foundation Discussion Paper No. 617.

_____ (1990). "Time series regression with a unit root and infinite variance errors," *Econometric Theory*, 6, 44–62.

Pollard, D. (1984). *Convergence of Stochastic Processes*. New York: Springer–Verlag.

_____ (1985). "New ways to prove central limit theorems," *Econometric Theory*, 1, 295–314.

_____ (1989a). "Asymptotics for least absolute deviation estimators," mimeographed, Yale University.

_____ (1989b). "Asymptotics via empirical processes," *Statistical Science*, 4, 341–366.

Prakasa Rao, B. L. S. (1969). "Estimation of a unimodal density," *Sankhya Series A*, 31, 23–36.

Resnick, S. I. (1986). "Point processes regular variation and weak convergence," *Advances in Applied Probability*, 18, 66–138.

Sargan, J. D. (1976). "Econometric estimators and the Edgeworth Approximation," *Econometrica*, 44, 421–428.

Silverman, B. W. (1986). *Density Estimation*. London: Chapman Hall.

White, H. (1984). *Asymptotics Theory for Econometricians*. New York: Academic Press.

Walter, G. and J. Blum (1979). "Probability density estimation using delta sequences," *Annals of Statistics*, 7, 328–340.