Economics Department

Economics Working Papers

The University of Auckland — *Year* 1997

A LAD Regression Under Non-Standard Conditions

Alan Rogers
University of Auckland, a.rogers@auckland.ac.nz

LAD REGRESSION UNDER NON-STANDARD CONDITIONS

Alan J. Rogers*

Department of Economics
The University of Auckland

ABSTRACT

Most work on the asymptotic properties of least absolute deviations (LAD) estimators makes use of the assumption that the common distribution of the disturbances has a density which is finite and positive at zero. We consider the implications of weakening this assumption in a regression setting. We see that the results obtained are similar in flavor to those obtained in a least squares context when the disturbance variance is allowed to be infinite: both the shape of the limiting distribution and the rate of convergence to it is affected in reasonably simple and intuitive ways. As well as conventional regression models we outline results for some simple autoregressive models which may have a unit root and/or infinite error variance.

* Address for correspondence: Department of Economics, University of Auckland, Private Bag 92019, Auckland, New Zealand; email: a.rogers@auckland.ac.nz

1. Introduction

We consider the linear regression model

$$y_i = x_i'\beta^\circ + u_i, \quad i = 1,...,n$$

where the u_i are independently and identically distributed with distribution function F, and $\{x_i\}$ is a sequence of $k \times 1$ regressor vectors. We are interested in the LAD (least absolute deviations) estimator of β°, which has long been of interest, especially as an alternative to the ordinary least squares estimator. More recently the asymptotic properties of the estimator have come under closer scrutiny, with quite a lot of attention being paid to alternative ways of obtaining the limiting distribution of the estimator. Two recent approaches are due to Pollard (1991) and Phillips (1991). The standard result here is that under some conditions

$$n^{1/2}(b_n^{LAD}-\beta^\circ) \rightarrow_d N(0,M^{-1}/(2f(0))^2),$$

where b_n^{LAD} is the LAD estimator of β°, f is the common probability density function of the u_i, M is the probability limit (as $n \rightarrow \infty$) of $\Sigma_{i=1,n}\, x_i x_i'/n$; $\rightarrow_d$ denotes convergence in distribution. In order to get this result it is assumed that the distribution of the u_i has zero median and possesses a density, f, which is both positive and continuous in a neighborhood of 0. (Phillips (1991) adopts the slightly stronger assumption that f is analytic on such a neighborhood .)

Such an assumption is an appealing one to make in many settings, but it seems worthwhile to consider the consequences of weakening it. One reason for doing so comes from inspection of the formula for the asymptotic covariance matrix for $n^{1/2}(b_n^{LAD}-\beta^\circ)$: the (asymptotic) efficiency of b_n^{LAD} relative to, say, the OLS estimator depends only on $(2f(0))/\sigma^2$, where $\sigma^2 = var(u_i)$, provided the latter exists and it is also true that $E[u_i] = 0$. So, it seems natural to want to know what will happen when $f(0) = 0$ or $f(0) = \infty$. Such cases are certainly quite consistent with conventional assumptions about the distribution of the u_i in regression contexts, such as symmetry about 0. Indeed, a comprehensive investigation of the location model (for which $x_i = 1$ for all i) has been undertaken by Smirnov (1952), although his work does not seem to have attracted much attention in the econometric literature on quantile-like estimates, or indeed in the statistical literature (one exception being Serfling (1980)). This is somewhat surprising given the amount of attention that has been paid to the behavior of ordinary least squares and related estimators under non-standard conditions, such as infinite variance disturbances.

The purpose of this paper is to provide some indications of what can happen under weaker conditions on the nature of F near zero. The conditions we employ are motivated partly by the ease with which they yield some definite results which are easy to understand. Some of these results are outlined in the next section. More importantly, the conditions we impose on F are closely related to conditions on domains of attraction that arise in Smirnov's study of the location model. This motivation is provided in more detail in Section 3. In Section 4 we see how relaxing the usual condition on F affects the rate of convergence of coefficient estimators in some simple time series models, including ones where F exhibits non-standard behavior in the tails as well as near zero.

2. Preliminary Ideas and Results

Here we suppose that $\{x_i\}$ is a deterministic sequence, and, as, in the preceding section, that the u_i are independently and identically distributed with distribution function F.

The usual argument leading to a limiting distribution for the LAD estimator is roughly as follows. We note first that

$$sgn(y_i - x_i'b_n^{LAD}) - sgn(y_i - x_i'\beta^o) = -I\{|u_i| \le |x_i'\delta_n|\} \times [sgn(x_i'\delta_n) + sgn(u_i)]$$

where

$$\delta_n \equiv b_n^{LAD} - \beta^o,$$

$$I(|u_i| \le |x_i'\delta_n| \equiv \begin{cases} 1 \text{ if } |u_i| \le |x_i'\delta_n| \\ 0 \text{ if } |u_i| > |x_i'\delta_n| \end{cases}$$

and $sgn(z) = 1$ if $z > 0$, $sgn(z) = 0$ for $z = 0$, $sgn(z) = -1$ if $z < 0$. Multiplying through by x_i, summing over the i, and making use of the first order condition $\Sigma_{i=1,n}\, x_i sgn(y_i - x_i'b_n^{LAD}) = 0$, yields

$$0 - \Sigma_{i=1,n}\, x_i sgn(u_i) = -\Sigma_{i=1,n}\, x_i \phi(u_i; x_i'\delta_n)[sgn(x_i'\delta_n) + sgn(u_i)].$$

The next step is not trivial, and amounts to establishing that asymptotically the summands on the right hand side can be replaced by suitably evaluated expectations. Noting that for any $\beta \in \mathbf{R}^k$, $E[I\{|u_i| \le |x_i'\beta|\}[sgn(x_i(\beta-\beta^o)) + sgn(u_i)]] = 2[F(x_i'(\beta-\beta^o)) - F(0)]$, it turns out that under some conditions on the x_i,

$$\eta_n \Sigma_{i=1,n}\, x_i sgn(u_i) \approx_a \eta_n \Sigma_{i=1,n}\, 2x_i[F(x_i'\delta_n) - F(0))]. \qquad \textbf{(1)}$$

where $\approx_a$ denotes equality in distribution in the limit as $n \to \infty$, and $\{\eta_n\}$ is a sequence of normalizing constants which is such that $\eta_n \Sigma_{i=1,n}\, x_i sgn(u_i)$ has a limiting distribution as $n \to \infty$; at this stage use is made of $F(0) = 1/2$, i.e., the u_i have zero median.

The next step typically taken is to apply first order Taylor series approximations to the terms $F(x_i'\delta_n) - F(0)$ appearing in (1) to obtain

$$\eta_n \Sigma_{i=1,n}\, x_i sgn(u_i) \approx_a \eta_n \Sigma_{i=1,n}\, 2x_i[F(0) + F'(0)(x_i'\delta_n) + o(x_i'\delta_n) - F(0)],$$

which after a little more work yields, in the case where $\eta_n = n^{-1/2}$,

$$n^{-1/2} \Sigma_{i=1,n} x_i \mathrm{sgn}(u_i) \approx_a [n^{-1}\Sigma_{i=1,n} 2x_i x_i' F'(0)] \times n^{1/2}\delta_n,$$

from which the standard result mentioned in the introduction follows, given $n^{-1}\Sigma_{i=1,n} x_i x_i' \rightarrow M$, $n^{-1/2}\Sigma_{i=1,n} x_i \mathrm{sgn}(u_i) \rightarrow_d N(0,M)$ as $n \rightarrow \infty$, with M being positive definite.

The last step in the foregoing relies on a preliminary consistency argument to validate the Taylor approximations. The novel approach adopted by Phillips (1991) proceeds from the same starting point but avoids (1) and the Taylor series expansions applied to it, by treating $\phi(\cdot)$ initially as if it were differentiable and then arguing in terms of generalized functions, and it is at this stage that the (local) differentiability of F is invoked. This approach does not seem amenable to the sort of generalization we are interested in this paper. Instead we will rely for a rigorous development on a modification of the elegant convexity argument of Pollard (1991). This modification is presented in the Appendix. One advantage of Pollard's approach is that it is self contained in the sense that there is no need to establish consistency of the LAD estimator prior to examining its limiting distribution (after suitable centering and scaling).

The derivation sketched out above requires for its final step that the Taylor argument be applicable, with F being required to possess a positive derivative in a neighborhood of 0. But even if the Taylor series argument is applicable, it is not clear that it is the only or most natural approximation to apply.

Suppose, for example, that

$$F(z) - F(0) = L(z) \times \mathrm{sgn}(z)|z|^{\gamma}, \quad \gamma \geq 0 \tag{2}$$

where L is non-negative, $L(z) > 0$ for all $z \neq 0$, and is slowly varying at zero from above and below. The slow variation (at zero) assumption means that for each $\lambda > 0$,

$$L(\lambda z)/L(z) \to 1 \text{ as } z \to 0.$$

Note that (2) implies that $F(z)$ is regularly varying at zero from above and below, because

$$[F(\lambda z)\text{-}F(0)]/[F(z)\text{-}F(0)] \to \lambda^\gamma,\ [F(-\lambda z)\text{-}F(0)]/[F(-z)\text{-}F(0)] \to \lambda^\gamma \text{ as } z \to +0,$$

for each $\lambda > 0$.

At this stage we motivate (2) simply as one of way of weakening the usual assumption (i.e., $0 < F'(0) < \infty$), while at the same time preserving some smoothness in F at zero. In fact (2) is of far greater importance in the asymptotic theory of LAD estimation than it might appear, and we return to this point in the next section. Here we recall (see, e.g., Seneta (1976)), that F is regularly varying at zero whenever

$$\psi(\lambda) \equiv \lim_{z\to 0} [F(\lambda z)\text{-}F(0)]/[F(z)\text{-}F(0)]$$

exists and is positive and finite for each $\lambda > 0$, so regular variation is obtained if we are prepared to make mild assumptions on the limit function $\psi(\lambda)$. In addition, $\psi(\lambda)$ can be thought of as a type of derivative of F at zero, with

$$\psi(\lambda) = \lim_{t\to 0} [F(\lambda\theta(t))\text{-}F(0)]/t, \qquad \theta(t) \equiv F^{-1}(t+F(0))$$

and viewed in this way regular variation is a natural weakening of the assumption that $F'(0)$ is positive and finite. If it is, then $\theta(t) = 0 + t/F'(0) + o(t)$ as $t \to 0$, and $\psi(\lambda) =$ $\lim_{t\to 0} [F(\lambda t/F'(0)) + \lambda o(t) - F(0)]/t$, so $\psi(\lambda) = F'(0)$ for $\lambda = F'(0)$, from which $F'(0)$ is seen to be a fixed point of ψ in this special case.

If we use (2) in (1) we now have simply

$$\eta_n \Sigma_{i=1,n} x_i \text{sgn}(u_i) \approx_a 2\, \eta_n \Sigma_{i=1,n} x_i[L(x_i'\delta_n)\, \text{sgn}(x_i\delta_n)\, |x_i'\delta_n|^\gamma], \tag{3}$$

a result which can be established rigorously under a mild condition on the x_i by adapting a proof of Pollard (1991). How this can be done is explained in the Appendix. We now explore some implications of (3) by means of a number of examples.

Example 1. (Location Model.) We consider the location model for which $k = 1, x_i = 1$ for all i. In this case $\eta_n = n^{-1/2}$, and (3) reduces to

$$n^{-1/2}\Sigma_{i=1,n}\ \mathrm{sgn}(u_i) \approx_a 2n^{1/2}L(\delta_n)\mathrm{sgn}(\delta_n)\,|\,\delta_n\,|^{\gamma};$$

here the limiting distribution of the term on the left hand side is N(0,1), and so the limiting distribution (after normalization) of δ_n is non-normal in general: normality clearly occurs in the special case where

$$\gamma = 1 \quad \text{and} \quad L(z) \rightarrow c > 0 \ \text{ as } \ z \rightarrow 0$$

for some constant c. If we impose the second of these conditions but do not constrain γ to equal one, we obtain

$$n^{1/2\gamma}\delta_n \approx_a \mathrm{sgn}(v) \times (\,|\,v\,|/2c)^{1/\gamma}, \qquad \text{where} \ \ v \sim N(0,1).$$

Here, letting γ differ from one has two effects. The first is to induce non-normality in the limiting distribution, as we have already noted, and the second causes the rate of convergence of b_n^{LAD} to β^o to differ from the usual $O_p(n^{-1/2})$ rate: the rate of convergence is faster whenever $\gamma < 1$, and slower for $\gamma > 1$. The shape of the limiting distributions also clearly depends on γ in a very simple way. The smaller is γ the more heavy tailed is this distribution, but with greater concentration near zero, and the larger γ is the thinner tailed it is with less concentration near zero (with the distribution concentrated entirely at -1 and 1 in the limit as $\gamma \rightarrow \infty$).

For this example, in the special case where $\gamma = 1$, b_n^{LAD} has a $O_p(n^{-1/2})$ convergence rate to $\beta°$, and $n^{1/2}(b_n^{LAD}-\beta°)$ has a normal limiting distribution with variance given by $1/4c^2$, given $L(z) \to c > 0$ as $z \to 0$. So, we obtain asymptotic normality with a $O_p(n^{-1/2})$ convergence rate without imposing the usual differentiability condition on F. On the other hand, if F is differentiable at 0, and the derivative (density) there is positive, F must be regularly varying at zero with exponent equal to unity, so that (3) holds with $\gamma = 1$, i.e., $F(z) - F(0) = z\,L(z)$. Moreover, in this case, $F'(z) = zL'(z) + L(z)$, so $F'(0) = L(0) > 0$, and (3) yields $n^{1/2}\delta_n \approx_a v/2L(\delta_n) \approx_a v/2L(0) = v/2F'(0)$, for $v \sim N(0,1)$, which is the familiar result.

Before we proceed we note that if we relax the requirement that $L(z)$ have a positive, finite limit as $z \to 0$, the shape of the limiting distribution is not affected if a symmetry condition such as $L(z)/L(-z) \to 1$ as $z \to 0$ is imposed, provided $\gamma > 0$, but the rate of convergence of δ_n is affected. This can be seen quite easily by means of examples such as $L(z) \approx -\mathit{ln}(|z|)$ or $L(z) \approx -1/\ln(|z|)$ for $z \approx 0$. Another possibility consistent with (2) is $L(z) \to c^+$ as $z \to +0$, and $L(z) \to c^-$ as $z \to -0$ with $c^+ \neq c^-$, $c^+, c^- \in (0,\infty)$. In this case, the rate of convergence of the estimator is unaffected, but the limiting distribution is asymmetric. There are other possibilities as well, and (2) may also be relaxed. With the exception of Example 5 of the next section we will not pursue these issues here.

Example 2. (Univariate Regression.) Take $k = 1$, as in Example 1, but let the x_i be conventional non-stochastic regressors. Assuming that $L(z) \to c$ as $z \to 0$, (3) yields

$$\eta_n \textstyle\sum_{i=1,n} x_i \mathrm{sgn}(u_i) \approx_a 2\, \eta_n \sum_{i=1,n} x_i\, c\, \mathrm{sgn}(x_i\delta_n)\, |x_i\delta_n|^{\gamma},$$

and the fact that x_i, δ_n are scalars allows us to write this as

$$\eta_n \Sigma_{i=1,n} x_i sgn(u_i) \approx_a 2\, \eta_n \Sigma_{i=1,n} c\, |x_i|^{1+\gamma} sgn^2(x_i) sgn(\delta_n)\, |\delta_n|^{\gamma}.$$

Now, if, $n^{-1/2} \Sigma_{i=1,n} x_i sgn(u_i) \rightarrow_d N(0,M_2)$, with $M_2 \equiv \lim_{n\to\infty} n^{-1} \Sigma_{i=1,n} x_i^2$, then $\eta_n = n^{1/2}$, and

$$sgn(\delta_n) \times n^{1/2} |\delta_n|^{\gamma} \approx_a n^{-1/2} \Sigma_{i=1,n} x_i sgn(u_i) / 2n^{-1} \Sigma_{i=1,n} c\, |x_i|^{1+\gamma},$$

or

$$n^{1/2\gamma} |\delta_n| \approx_a |[n^{-1/2} \Sigma_{i=1,n} x_i sgn(u_i) / 2n^{-1} \Sigma_{i=1,n} c\, |x_i|^{1+\gamma}]|^{1/\gamma},$$

and

$$n^{1/2\gamma}\delta_n \approx_a sgn(v) \times |v|^{1/\gamma}, \quad v \sim N(0, M_2/4c^2M_\gamma^2),$$

where $M_\gamma \equiv \lim_{n\to\infty} n^{-1} \Sigma_{i=1,n} |x_i|^{1+\gamma}$, assuming this limit exists.

Note here that the magnitude of γ controls the rate of convergence of the estimator, and determines the shape of its limiting distribution (after normalisation), as in Example 1, but it also plays a central role in determining the size of the variance of the distribution of the variate v via M_γ. We note in passing that it is possible to obtain simple bounds on the relative magnitudes of the sums appearing in the definitions of M_γ and M_2. For example, if $\gamma > 1$,

$$\min_i\{|x_i|\}^{\gamma-1} \leq \Sigma_{i=1,n} |x_i|^{1+\gamma} / \Sigma_{i=1,n} x_i^2 \leq \max_i\{|x_i|\}^{\gamma-1};$$

for $\gamma < 1$, the same bounds apply, but with the equalities reversed.

Example 3. (Time Trend.) Here we consider the case of simple time trend with $k = 1$, and

$$x_i = i^{\rho},\ i = 1,2,\ldots,n.$$

It is well known that the OLS estimator is consistent under conventional error assumptions (such as existence of $E[u_i^2]$) if and only if $\rho \geq -1/2$. For the LAD estimator the conditions consistency turn out to be $\rho \geq -1/(1+\gamma)$ and $\rho > -1/2\gamma$, and which of these binds clealry depends on the sign of $(\gamma - 1)$. Given $\rho \geq -1/(1+\gamma)$, the condition $\rho > -1/2\gamma$ emerges from easily from the arguments presented in the Appendix. And, if $\rho \geq 0$, so that both the LAD and OLS estimators are consistent, the relationship between the rates of convergence is similar to that in Example 2 above: for OLS it is $O_p(n^{1/2+\rho})$, for LAD it is $O_p(n^{1/2\gamma+\rho})$, and using arguments similar to those used in Example 2 (and noting that $\eta_n = n^{-1/2-\rho}$) we can obtain $n^{1/2\gamma+\rho}(b_n^{LAD}-\beta^o) \rightarrow_d \text{sgn}(v) \times |v|^{1/\gamma}$, where $v \sim N(0,(\rho(1+\gamma)+1)^2/4c(2\rho+1)^2)$, if we invoke the assumption that $|L(z)| \rightarrow c$ as $z \rightarrow 0$.

Although we will be concerned with the case for which $k = 1$, we mention the following generalization, partly to indicate the nature of the complications that can arise. These are similar to issues involved in moving from the single to multiple parameter case in the theory of M-estimates: see, e.g., Huber (1981).

Example 4. (Multivariate Regression.) We consider the same sort of model as in Example 3 but with $k > 1$. Assuming that $L(z) \rightarrow c$ as $z \rightarrow 0$, we have, as in Example 2,

$$n^{-1/2} \Sigma_{i=1,n} x_i \text{sgn}(u_i) \approx_a 2n^{-1} \Sigma_{i=1,n} c\, x_i \,\text{sgn}(x_i'\delta_n)\, n^{1/2} |x_i'\delta_n|^{\gamma},$$

assuming $\eta_n = n^{-1/2,}$ but this is harder to deal with now that δ_n is a $k \times 1$ vector, except of course when $\gamma = 1$, in which case the standard result is obtained, or when the x_i have certain simple structures. Another way of viewing this difficulty is via the adaptation of Pollard's approach explained in the Appendix. If we suppose that there are constants, a_n, such that $a_n\delta_n$ has a limiting distribution with density at $d \in R^k$ denoted by $\lambda(d)$, then $\lambda(d) = \phi(Q(d))$, where ϕ is the limiting density function of $n^{-1/2} \Sigma_{i=1,n} x_i \text{sgn}(u_i)$, and

$$Q(d) = \lim_{n\rightarrow\infty} 2n^{-1} \Sigma_{i=1,n} c x_i \,\text{sgn}(x_i'd/a_n)\, n^{1/2} |x_i'd/a_n|^{\gamma};$$

for the last limit to exist we will typically require that $n^{1/2} = a_n^{\gamma}$, so that δ_n has a $O_p(n^{-1/2\gamma})$ convergence rate, as in Examples 1 and 2.

3. A Motivation for Regular Variation

We have considered the implications of a condition on F which, when combined with some restrictions on $\{x_i\}$, is sufficient to ensure that $a_n(b_n^{LAD}-\beta^\circ)$ has a (non-degenerate) limiting distribution, where $\{a_n\}$ is a sequence of normalizing constants. The class of limiting distributions encountered in the last section is quite narrow. The question that arises naturally is how broad the class of possible limiting distributions can be. This question has been comprehensively answered by Smirnov (1952) for the location model of Example 1, where b_n^{LAD} is the sample median of the y_i.

Result 1 (Smirnov). Suppose that in the location model for which $x_i = 1$ for all $i = 1,...,n$, and the u_i are independently and identically distributed, there exist sequences of normalizing constants $\{a_n\}$, $\{c_n\}$ and a non-degenerate distribution function Λ such that

$$\Pr[(b_n^{LAD}-c_n)/a_n \le z] \rightarrow \Lambda(z).$$

for all z at which $\Lambda(z)$ is continuous. Then, aside from adjustments for scale and location,

$$\Lambda(z) = \Phi(Q(z)),$$

at all z for which $\Lambda(z) \in (0,1)$, Φ is the standard normal distribution function, and Q(z) takes one of the following forms

(a) $$Q(z) = \begin{cases} -\mu^\circ |z|^{\gamma^\circ} & \text{for } z \in (-\infty,0) \\ 0 & \text{for } z = 0 \\ \mu^* z^{\gamma^*} & \text{for } z \in (0,\infty) \end{cases}$$

$$\begin{cases} -\infty & \text{for } z \in (-\infty,1] \end{cases}$$

(b) $\quad Q(z) \quad = \quad \begin{cases} 0 & \text{for } z \in (-1,1] \\ \infty & \text{for } z \in (1,\infty) \end{cases}$

for μ°, μ^* positive and one of these (but not both) possibly infinite, and $\gamma^\circ,\gamma^* > 0$; if both μ° and μ^* are finite, then $\gamma^\circ = \gamma^*$.

Note that (b) can be regarded as a limiting case of (a) as $\gamma^\circ, \gamma^* \to \infty$, for finite μ°, μ^*. The range of possible limiting distributions in Result 1 is clearly narrow. In fact, if we confine attention to those Λ which are symmetric around zero, the only possibility is $\Lambda(z) = \Phi(\mu\ \mathrm{sgn}(z)|z|^\gamma)$, $\gamma, \mu \in (0,\infty)$, which is precisely the limiting distribution we encountered in Example 1.

The next question concerns conditions on F which are such that $(b_n^{LAD}-c_n)/a_n$ has one of the limiting distributions in Result 1, where we continue to confine attention to the location model. This is the content of the next result, also due to Smirnov (1952).

Result 2 (Smirnov).

(a) Result 1(a) holds with $\mu^\circ = \infty, \mu^* < \infty$ if and only if F satisfies

$F(0+) = 1/2$, (where $F(0+) = \lim_{z\to+0} F(z)$),

$[F(z) - 1/2]/[1/2 - F(-z)] \to 0$ as $z \to +0$,

$[F(\lambda z)-1/2]/[F(z)-1/2] \to \lambda^{\gamma^*}$ as $z \to +0$, for each $\lambda > 0$.

(b) Result 1(a) holds with $\mu^\circ < \infty, \mu^* = \infty$ if and only if F satisfies

$F(0-) = 1/2$, (where $F(0-) = \lim_{z\to+0} F(z)$),

$[F(z) - 1/2]/[1/2 - F(-z)] \to \infty$ as $z \to +0$,

$[F(-\lambda z)-1/2]/[F(-z)-1/2] \to \lambda^{\gamma^\circ}$ as $z \to +0$ for each $\mu > 0$.

(c) Result 1(a) holds with $\mu^\circ < \infty$, $\mu^* < \infty$, and $\gamma^\circ = \gamma^* \equiv \gamma$ if and only if F satisfies

$F(0-) = F(0+) = 1/2$, so 0 is a continuity point of F,

$F(-\varepsilon) < F(0) < F(\varepsilon)$ for all $\varepsilon > 0$,

$[F(z) - 1/2]/[1/2 - F(-z)] \to q$ as $z \to +0$, where q is positive and finite,

$[F(\lambda z)-1/2]/[F(z)-1/2] \to \lambda^\gamma$ as $z \to +0$ for each $\mu < 0$.

(d) It is sufficient for Result 1(b) to hold that F be continuous at 0 (i.e., $F(0-) = F(0+) = 1/2$) and

$[F(\lambda z) - 1/2]/F(\mathrm{sgn}(\lambda)z) - 1/2] \to \infty$ as $z \to +0$, whenever $|\lambda| > 1$.

The sufficient condition in (d) is that F be rapidly varying at zero from both the left and right, so that $F(z) - 1/2$ approaches zero as $z \to 0$ more rapidly than it would do if F were regularly varying at zero.

The cases (a) - (c) all feature regular variation of F at zero prominently.

For (a), the third condition is that F be regularly varying at zero from the right with exponent γ^*, the second being that if $F(-0) = 1/2$, F(z) approaches 1/2 from below more slowly than from above. The conditions in (b) mirror those in (a).

The last condition in (c) is that F be regularly varying at zero from the right with exponent γ, and the third condition in (c) when combined with this is slightly stronger than the requirement F be regularly varying from the left with the same exponent, γ.

Therefore, the assumptions on F invoked in the preceding section are closely related to the necessary and sufficient conditions for Result 2(c). If we restrict L(z) in (2) to satisfy $L(z)/L(-z) \to q > 0$ as $z \to +0$, these two sets of conditions are identical.

The main point here is that the regular variation assumption embodied in (2) has a motivation beyond that of a convenient approximation which allows extensions to existing results to be obtained in a straightforward manner.

It is instructive at this stage to compare the foregoing results for the LAD estimator, b_n^{LAD} - the sample median for the location model - with the results for the OLS estimator (i.e., the sample mean) for the same model. If we denote this estimator by b_n^{OLS}, we know that if there exist sequences of constants $\{a_n\}$, $\{c_n\}$ such that $(b_n^{OLS}-c_n)/n$ has a non-degenerate limiting distribution (as $n \to \infty$) then this limiting distribution is stable. This is the analog of Result 1. The analog of Result 2 is the following. It is necessary and sufficient for the limiting distribution of $(b_n^{OLS}-c_n)/n$ to be stable with index $\alpha \in (0,2]$ that F satisfy

$$\begin{aligned} 1-F(z) &= [c_u+\iota_u(z)] \times z^{-\alpha} L^\circ(z), \quad z > 0 \\ F(z) &= [c_l+\iota_\lambda(z)] \times (-z)^{-\alpha} L^\circ(-z), \quad z \leq 0, \end{aligned} \tag{4}$$

with $\iota_u(z) \to 0$, $\iota_l(-z) \to 0$ as $z \to +\infty$, and $L^\circ(z)$ is slowly varying at infinity, c_u, $c_l \geq 0$, $c_u+c_l > 0$. (See, e.g., Ibragimov and Linnik (1971), p.76.) Regular variation of 1-F at $+\infty$ and F at $-\infty$ clearly features here prominently, as does regular variation of F at zero in Result 2.

Now we return to the regression model $y_i = x_i'\beta + u_i$. A considerable amount of attention has been paid to the consequences for least squares estimators of heavy tailed distributions for the u_i. The assumption most commonly made here is that these distributions lie in the domain of attraction of a stable law with index $\alpha \in (0,2)$, meaning that F satisfies (4), since this is a necessary and sufficient for there to be a limiting distribution, after normalization for scale and location, for the OLS estimator in the simplest case, namely the location model.

Given that (4) is a commonly adopted assumption in regression contexts, particularly when one of the estimators under consideration is the OLS estimator, it seems natural to impose an analogous condition based on similar considerations for the LAD estimator. That, is, Result 2 defines domains of attraction for the limiting distributions of the LAD estimator for the location model, so the question arises as to what results we get for the LAD estimator in more general regression settings when the disturbances have a distribution in one of these domains of attraction. The results of the last section suggest some of the forms these results can take. In the next section, we sketch some additional results for certain time series models.

Before we proceed it may be worth mentioning the consequences of F obeying a condition similar to that in Result 2 (a) where the behaviour of F near zero does not exhibit the symmetry we imposed in Examples 1-3 of the last section.

Example 5. Suppose that F satisfies

$$F(0) - F(z) = \begin{cases} z^{\gamma}L(z) & \text{for } z \geq 0 \\ -(-z)^{\gamma-\varepsilon}L(z) & \text{for } z < 0, \end{cases}$$

where $\gamma > 0$ and $\varepsilon \in (0,\gamma)$, and $L(z) \rightarrow c$ as $z \rightarrow 0$, $F(0) = 1/2$. We distinguish two cases.

(a) If $x_i = 1$ for all i, we know from Result 1(a) that the limiting distribution of given by

$$n^{1/2\gamma}(b_n^{LAD}-\beta^{\circ}) \approx_a \max\{0,v\} \times (|v|/2c)^{1/\gamma}, \quad v \sim N(0,1),$$

which is a one-sided version of the result in Example 1. To see how this works in terms of the approach of the preceding section we note that (1) yields

$$n^{1/2}2c|\delta_n^+|^{\gamma} \approx_a v \quad \text{if} \quad v \geq 0$$

$$-n^{1/2}2c|\delta_n^-|^{\gamma-\varepsilon} \approx_a v \quad \text{if} \quad v < 0$$

where $v \sim N(0,1)$ and $\delta_n^+ = \max\{0,\delta_n\}$, $\delta_n^- = \min\{0,\delta_n\}$. Then δ_n^+ has an $O_p(n^{-1/2\gamma})$ convergence rate and δ_n^- has the more rapid rate of $O_p(n^{-1/2(\gamma-\varepsilon)})$. In this example it is the larger of the two exponents of regular variation which determines the rate of convergence of δ_n to zero in probability, since $\delta_n = \delta_n^+ + \delta_n^-$.

(b) Suppose $x_i = -1$ for $i = 1,...,m$, $x_i = 1$ for $i = m+1,...,n$. Now, if $\delta_n \geq 0$,

$$n^{-1/2}\Sigma_{i=1,n} 2x_i[F(x_i\delta_n)-F(0)] \approx_a 2cn^{-1/2}\{m|\delta_n|^{\gamma-\varepsilon} + (n-m)|\delta_n|^{\gamma}\}$$

and if $\delta_n < 0$,

$$n^{-1/2}\Sigma_{i=1,n} 2x_i[F(x_i\delta_n)-F(0)] \approx_a 2cn^{-1/2}\{-m|\delta_n|^{\gamma} - (n-m)|\delta_n|^{\gamma-\varepsilon}\}$$

so from (1) we have, letting $\delta_n^+ = \max\{0,\delta_n\}$ and $\delta_n^- = \min\{0,\delta_n\}$ as in (a)

$$2cn^{-1/2}[m|\delta_n^+|^{\gamma-\varepsilon} + (n-m)|\delta_n^+|^{\gamma}] \approx_a \max\{0,v\},$$

$$2cn^{-1/2}[-m|\delta_n^-|^{\gamma} - (n-m)|\delta_n^-|^{\gamma-\varepsilon}] \approx_a \min\{0,v\}, \qquad v \sim N(0,1).$$

So if $m/n \to r \in (0,1)$ as $n \to \infty$, then it follows from $|\delta_n|^{\varepsilon} \to_p 0$, that

$$n^{1/2(\gamma-\varepsilon)}\delta_n \approx_a \begin{cases} |v/2r|^{1/(\gamma-\varepsilon)} & \text{if } v \geq 0 \\ |v/2(1-r)|^{1/(\gamma-\varepsilon)} & \text{if } v < 0 \end{cases}$$

which is evidently asymmetric unless $r = 1/2$ but exhibits a more rapid convergence rate than in (a) whenever $r \neq 0$ (or 1), with this rate now being determined by the smaller of the two exponents of regular variation.

The difference between (a) and (b) in this example is at first glance a surprising consequence of an apparently minor difference between regressors, but has a simple interpretation in terms of contamination. Suppose that $x_i = 1$ for all i, as in (a), but that for each i, $u_i = v_i$ with probability 1-r, and $u_i = -v_i$ with probability $r \in (0,1)$, where v_i has the asymmetric distribution in the preceding example. Then it is now the case that u_i is regularly varying at zero with exponent $\gamma\text{-}\varepsilon$ from both the left and right, and it is this exponent that controls the rate of convergence of b_n^{LAD} to $\beta°$. Since contaminating a given distribution can never thin its "waist" (or tails), such contamination can only improve the rate of convergence of the LAD estimator in a simple location model, but, as is well known can never improve that of the OLS estimator.

4. **Some Further Results**

Here we concentrate on the simple autoregressive model for which $x_i = y_{i-1}$, so we have

$$y_i = \beta y_{i-1} + u_i, \quad i = 1,\ldots,n$$

$$y_0 = 0, \quad \beta \geq 0.$$

The u_i have zero median ($F(0) = 1/2$), and are independently and identically distributed. Throughout we assume for simplicity that the common distribution function of the u_i, F, lies in the normal domain of attraction of a stable distribution with index $\alpha \in (0,2]$, so that F satisfies (4) with $L°(z) = 1$, for $z > 0$. We also assume for expositional ease that in (4) $c_l = c_u = 1$. Assume that F is symmetric around zero, so that $E[u_i] = 0$ whenever $\alpha > 1$. We will also assume that F satisfies (2), for $\gamma > 0$ with $L(z)$ in (2) satisfying $L(z) \rightarrow c$ as $z \rightarrow 0$. So α measures the fatness of F's tails, and γ that of its "waist".

First we review some results for the standard case where $\gamma = 1$. When $\beta < 1$, the presence of fat tailed errors (i.e., $\alpha \in (0,2)$) causes the OLS estimator to converge at a faster rate than in the traditional model with finite variance disturbances. Moreover, the rate of convergence is faster the smaller is the index α. Compared to OLS, the LAD estimator has the same convergence rate in the finite variance case, but a marginally faster rate in the heavy tailed case. This last point is explained more fully below in Example 7 where we consider the consequences of allowing γ to differ from one. When the disturbances have finite variance, the $O_p(n^{-1/2})$ convergence rate for the OLS estimator in the stationary autoregressive model with $\beta < 1$ changes abruptly to a $O_p(n^{-1})$ rate for the non-stationary case with $\beta = 1$. The same is true for the LAD estimator. And when $\beta = 1$, the convergence rate for OLS remains rather remarkably at $O_p(n^{-1})$, when the disturbances are heavy tailed, i.e., for any $\alpha \in (0,2)$. This result is mentioned by Knight (1989) and a derivation appears in Chan and Tran (1989), who also derive the limiting distribution of $n^{-1}(b_n^{OLS}-1)$. (Phillips (1990) generalizes their results to integrated processes where the disturbances are weakly dependent as well as having heavy tailed marginal distributions.) So far as the LAD estimator is concerned, the combination of $\beta = 1$ and a fat tailed error distribution gives rise to an $O_p(n^{-1/2-1/\alpha})$ rate of convergence, which is faster than the $O_p(n^{-1})$ rate for OLS, and more rapid the smaller is the tail index α. This result is due to Knight (1989), and is noteworthy to the extent that we are used to thinking of the tails of error distributions as being irrelevant for the limiting behaviour of LAD estimators.

Now we consider the consequences of allowing the exponent γ in (2) to differ from unity. This does not affect the limiting behaviour of the OLS estimator, but can affect that of the LAD estimator. What follows is based on

$$a_n^{LAD}|b_n^{LAD} - \beta^\circ| = a_n^{LAD}[|\Sigma_{i=2,n} y_{i-1}\mathrm{sgn}(u_i) / \Sigma_{i=2,n} 2c|y_{i-1}|^{1+\gamma}|]^{1/\gamma} + o_p(1) \qquad (5)$$

in which $\{a_n^{LAD}\}$ is a suitable sequence of normalizing constants such that the first term on the right hand side has a non-degenerate limiting distribution. We do not provide a formal justification for (5) here: a heuristic defence is available by analogy with (3) which we know works for the deterministic regressor case. Tighter derivations of the results of this section appear in the Appendix. For the OLS estimator, $\{a_n^{OLS}\}$ is the counterpart of $\{a_n^{LAD}\}$ - a sequence of constants such that

$$a_n^{OLS}(b_n^{OLS} - \beta^\circ) = a_n^{OLS}\{\Sigma_{i=2,n} y_{i-1}u_i / \Sigma_{i=2,n} y_{i-1}^2\}$$

has a non-degenerate limiting distribution.

Example 6. Consider the random walk case where $\beta^\circ = 1$. Then, $y_i = \Sigma_{j=1,i} u_j$, and for $r \in [0,1]$, $y_{\lfloor rn \rfloor} = \Sigma_{j=1,\lfloor rn \rfloor} u_j$, and $y_{\lfloor rn \rfloor}/ n^{1/\alpha} \rightarrow_d S_\alpha(r)$, where $S_\alpha(r)$ is a α-stable process, and $\lfloor \cdot \rfloor$ denotes integer part. Now, as in Phillips (1991),

$$n^{-1/2}\Sigma_{i=2,n} (y_{i-1}/n^{1/\alpha})\mathrm{sgn}(u_i) \rightarrow_d \int_{0,1} S_\alpha^-(r)\, dW(r),$$

where $S_\alpha^-(r)$ is the left limit of $S_\alpha(\cdot)$ at r, and W(r) is a standard Wiener process. In addition

$$n^{-1}\Sigma_{i=2,n} 2c|y_{i-1}/n^{1/\alpha}|^{1+\gamma} \rightarrow_d \int_{0,1} 2c|S_\alpha(r)|^{1+\gamma} dr$$

So, from (5),

$$a_n^{LAD}|b_n^{LAD} - 1| \approx_a a_n^{LAD} \times [(n^{1/2}/n)n^{1/\alpha} / (n^{1/\alpha})^{1+\gamma}]^{1/\gamma}$$
$$\times\ [|n^{-1/2}\Sigma_{i=2,n} (y_{i-1}/n^{1/\alpha})\mathrm{sgn}(u_i) / n^{-1}\Sigma_{i=2,n} 2c|y_{i-1}/n^{1/\alpha}|^{1+\gamma}|]^{1/\gamma}$$

and we can conclude that

$$a_n^{LAD} = 1/[(n^{1/2}/n)n^{1/\alpha}/(n^{1/\alpha})^{1+\gamma}]^{1/\gamma} = 1/[n^{-1/2} \times n^{-\gamma/\alpha}]^{1/\gamma} = n^{1/2\gamma + 1/\alpha},$$

and

$$n^{1/2\gamma + 1/\alpha} \,|\, b_n^{LAD} - 1 \,| \rightarrow_d [\,|\int_{0,1} S_\alpha^-(r)dW(r) / \int_{0,1} 2c\,|S_\alpha(r)|^{1+\gamma}dr\,|\,]^{1/\gamma}$$

which coincides with Knight's (1989) result when $\gamma = 1$. (See also Phillips (1991).) For the OLS estimator, $a_n^{OLS} = n$, and

$$n(b_n^{OLS} - 1) \rightarrow_d \int_{0,1} S_\alpha^-(r)dS(r) / \int_{0,1} S_\alpha(r)^2 dr.$$

Evidently, the same sort of pattern as encountered in sections 2 and 3 emerges: the larger is γ, the slower the rate of convergence of the LAD estimator. The LAD and OLS estimators have the same rate of convergence when $1 = 1/2\gamma + 1/\alpha$, which can only be satisfied if $\alpha > 1$. That is, F cannot be so thin "waisted" that the LAD estimator is dominated by the OLS estimator (in rate of convergence terms) when F has tails which are as fat or fatter than those of the Cauchy distribution.

Example 7. Here we consider the stationary case for which $\beta° \in [0,1)$. For simplicity we outline the argument for $\beta° = 0$, so that $\{y_i\}$ is an i.i.d. sequence. So far as the OLS estimator is concerned, it is a simple matter to show that in

$$a_n^{OLS}(b_n^{OLS} - 0) = a_n^{OLS} \times \Sigma_{i=2,n} u_{i-1}u_i \,/\, \Sigma_{i=2,n} u_{i-1}^2.$$

the normalized quantities $n^{-2/\alpha}\Sigma_{i=2,n} u_{i-1}^2$, $(n\,ln(n))^{-1/\alpha}\Sigma_{i=2,n} u_{i-1}u_i$ have non-degenerate limiting distributions for $\alpha < 2$. (The factor $n\,ln(n)$ arises because of the presence of the product of the independent variates u_i and u_{i-1}, which have the same heavy-tailed distributions. See Phillips (1990, Appendix A). So,

$$a_n^{OLS} = 1/[(n\,ln(n))^{-1/\alpha} / n^{-2/\alpha}] = [n/ln(n)]^{1/\alpha}.$$

For the LAD estimator, from (5),

$$a_n^{LAD}|b_n^{LAD} - 0| \approx_a a_n^{LAD}[|\Sigma_{i=2,n} u_{i-1}sgn(u_i) / \Sigma_{i=2,n} 2c|u_{i-1}|^{1+\gamma}|]^{1/\gamma}.$$

Now, for any $\alpha \in (0,2]$, the numerator in the right side here is

$$\Sigma_{i=2,n} u_{i-1}sgn(u_i)/n^{1/\alpha} \to_d S_\alpha(1),$$

while $\Sigma_{i=2,n} |u_{i-1}|^{1+\gamma}$ is the sum of n-1 i.i.d. variates the distribution of which lies in the domain of attraction of a stable law with index $\alpha/(1+\gamma)$ which is less than 2 whenever $\alpha < 2$. This follows from

$$Pr[|u_i|^{1+\gamma} \leq z] = Pr[|u_i| \leq z^{1/(1+\gamma)}] = \max\{z, 0\} \times F(z^{1/(1+\gamma)}),$$

and the assumptions on F made above. There are two cases here with quite different outcomes.

(a) Suppose that $1+\gamma < \alpha$ (< 2) which requires $\gamma < 1$ as well as $\alpha > 1$ (so that $E[|u_i|] < \infty$): in this case $E[|u_i|^{1+\gamma}] < \infty$, $n^{-1}\Sigma_{i=2,n} |u_{i-1}|^{1+\gamma}$ has probability limit equal to this expectation, and

$$a_n^{LAD}|b_n^{LAD} - \beta°| \approx_a a_n^{LAD} \times [n^{1/\alpha}/n]^{1/\gamma} \times [|n^{-1/\alpha}\Sigma_{i=2,n} u_{i-1}sgn(u_i)/n^{-1}\Sigma_{i=2,n} 2c|u_{i-1}|^{1+\gamma}|]^{1/\gamma};$$

so $a_n^{LAD} = 1/n^{(1-\alpha)/\alpha\gamma} = n^{(\alpha-1)/\alpha\gamma}$, recalling that $1-\alpha < 0$, in view of $\alpha > 1$. In addition $1+\gamma < \alpha$ implies $(\alpha-1)/\gamma > 1$, so $a_n^{LAD} > n^{1/\alpha}$ in this case, a point worth noting for (b) below.

(b) Now reverse the inequality in (a) so $1+\gamma > \alpha$. Then $E[|u_i|^{1+\gamma}] = \infty$, and

$$a_n^{LAD}|b_n^{LAD} - \beta°| \approx_a a_n^{LAD} \times [n^{1/\alpha}/(n^{1/\alpha})^{1+\gamma}]^{1/\gamma}$$
$$\times [|n^{-1/\alpha}\Sigma_{i=2,n} u_{i-1}sgn(u_i) / (n^{-1/\alpha})^{1+\gamma} \Sigma_{i=2,n} 2c|u_{i-1}|^{1+\gamma}|]^{1/\gamma};$$

from which we see that $a_n^{LAD} = n^{1/\alpha}$. Here

$$(n^{-1/\alpha})^{1+\gamma}\Sigma_{i=2,n} |u_{i-1}|^{1+\gamma} \to_d S_{\alpha/(1+\gamma)}(1),$$

and the fact that $\alpha/(1+\gamma) < 1$ means that we do not need to bother with a sequence of centering constants.

Hence the rate of convergence of the LAD estimator in Example 7(b) is independent of γ, the parameter that is critical in determining the rate of convergence of the LAD estimator in all the preceding examples considered in this paper. There is a parallel here of sorts with the OLS estimator in Example 6 (where $\beta^\circ = 1$) with fat tailed errors, because there the rate of convergence of the OLS estimator is independent of the magnitude of the tail parameter α.

If, in Example 7, $1+\gamma = \alpha$, we get a result similar to that for (b) in the sense that $b_n^{LAD}-\beta^\circ = O_p(n^{-1/\alpha})$ but we need to be a little more careful about a choice of sequence of centering constants. (See, e.g., Loeve (1977, p.402).)

Comparing the OLS and LAD estimators in Example 7 we see that the LAD estimator converges at slightly faster rate whenever $1+\gamma \geq \alpha$, with the rates of convergence of both estimators dependent on α but not on γ. When $1+\gamma < \alpha$, the LAD estimator converges at a faster rate, with this rate now dependent on the waist thickness parameter γ.

Parts (a) and (b) of Example 7 together say in effect that the magnitude of α limits the extent to which a reduction in γ can improve the convergence rate of the LAD estimator (absolutely and relatively to the OLS estimator): no improvement is ever possible when $\alpha \in (0,1)$, and when $\alpha \in (1,2)$ such improvement is also impossible if $\gamma \geq 1$.

And, if $\alpha = 2$, so the u_i have finite variance, the OLS estimator has the usual $O_p(n^{-1/2})$ convergence rate, and the LAD estimator exhibits behavior similar to that encountered in Examples 1 and 2 of the Section 2, i.e., it has an $O_p(n^{-1/2\gamma})$ convergence rate.

We conclude this section with a case which provides a parallel for Example 7, but not for Example 6.

Example 8. Suppose that in $y_i = x_i'\beta^\circ + u_i$, the $\{x_i\}$, $\{u_i\}$ are mutually independent i.i.d. sequences, the distribution of the u_i satisfies the assumptions made above (with $\gamma > 0$, $\alpha \in (0,2)$) and the distribution of the x_i has tails with the same properties but with index $\kappa \in (0,2)$. If $a_n^{LAD}(b_n^{LAD}-\beta^\circ)$ and $a_n^{OLS}(b_n^{OLS}-\beta^\circ)$ have non-degenerate limiting distributions, then

$$a_n^{OLS} = \begin{cases} n^{1/\kappa} & \text{if } \kappa < \alpha \\ n^{(2\alpha-\kappa)/\kappa\alpha} & \text{if } \kappa > \alpha \end{cases}$$

(with $\kappa < 2\alpha$ being required for the consistency of b_n^{OLS}), and

$$a_n^{LAD} = \begin{cases} n^{1/\kappa} & \text{if } 1+\gamma > \kappa \\ n^{(\kappa-1)/\kappa\gamma} & \text{if } 1+\gamma < \kappa. \end{cases}$$

As we might expect, the tail parameter, α, of the u_i's does not affect the convergence rate of the LAD estimator, nor is that of the OLS estimator affected by γ. But we again have a situation in which the rate of convergence of the LAD estimator does not depend on the parameter γ, provided $1+\gamma > \kappa$. Moreover, this rate is the same as that for the OLS estimator when $\kappa < \alpha$. So

$$a_n^{OLS} = a_n^{LAD} = n^{1/\kappa} \quad \text{if } \kappa < \alpha < 1+\gamma,$$

which is a condition involving the three parameters κ, α and γ.

The intuition here seems straightforward: if the distribution of the regressors has sufficiently thick tails (i.e., small enough κ) then this swamps the effect a large γ has in reducing the convergence rate of the LAD estimator. A similar phenomenon is present in Example 7. This explanation is probably too simplistic, however, because the parameter γ always affects the convergence rate of the LAD estimator in the random walk case of

Example 6, no matter how heavy tailed are the innovations. The same is true for the deterministic trend case in Example 3, because there the tails of the disturbance distribution are irrelevant for the asymptotics of the LAD estimator, and the trend parameter ρ can never be large enough to obliterate the effect of γ in these asymptotics.

Finally, the results of this section concern only rates of convergence of the estimators: the shapes of their limiting distributions after normalization for scale will generally also depend on the parameters of the models.

5. Conclusion

We have attempted to indicate how the asymptotic behavior of the LAD estimator is affected when the conventional assumption on the disturbance distribution is relaxed in a way which seems intuitively attractive, and which also has a compelling justification in the simple case of a location model. We have seen (in Section 2) that the results for a location model originally due to Smirnov (1952) apply with some modifications that are not especially surprising in the single regressor case, but the range of results is substantially richer in the case of stochastic regressors, including simple stable and unstable autoregressive models. At the very least, these results should serve as a warning that the usual rule of thumb to the effect that the LAD estimator is more robust to "outliers" can be seriously wrong, and can also be interpreted in a way which understates the possible superiority of the LAD method. This is so simply because this method of estimation is usually - but not always - sensitive to "inliers".

Appendix

A.1 The primary objective here is to provide a formal justification for results on the LAD estimator in the text. We first define

$$M(z) \equiv E[\,|u_i - z| - |u_i|\,],$$

and establish the following.

Lemma. Suppose that F satisfies (2) where $L(z)$ is non-negative and slowly varying and either (i) $\gamma > 0$, or (ii) $\mathrm{sgn}(L(z))$ is monotone non-decreasing in some neighborhood of 0. Then

$$M(z) = (1+o(1)) \times 2\,|z|^{1+\gamma} L(z)/(1+\gamma) \quad \text{as } z \to 0.$$

Proof. Take $z > 0$, and note that

$$M(z) = -z[1-F(z)] - 2\int_{0,z} t dF(t) + z\int_{0,z} dF(t) + zF(0)$$

$$= -z[1-F(0)+F(0)-F(z)] - 2\int_{0,z} t dF(t) + z[F(z)-F(0)] + zF(0)$$

$$= 2z[F(z)-F(0)] - 2\int_{0,z} t dF(t) \quad (\text{using } 1-F(0) = F(0) = 1/2)$$

which on integrating by parts yields

$$M(z) = 2z[F(z)-F(0)] + 2\int_{0,z} [F(t)-F(0)]dt - 2z[F(z)-F(0)]$$

$$= 2\int_{0,z} [F(t)-F(0)]dt.$$

Now using the regular variation assumption (2), we have, again for $z > 0$,

$$M(z) = 2\int_{0,z} L(t)t^{\gamma} dt = 2z^{1+\gamma}\int_{0,1} L(vz)v^{\gamma} dv.$$

The next step involves establishing that

$$\int_{0,1} L(vz)v^{\gamma} dv \approx L(z)\int_{0,1} v^{\gamma} dv \quad \text{as } z \to 0.$$

But this follows from an adaptation of Theorem 2.6 of Seneta (1976) to deal with regular variation at zero rather than at infinity. Therefore,

$$M(z) \approx 2\, z^{\gamma+1}L(z) \int_{0,1} v^{\gamma}\, dv = 2\, L(z)z^{\gamma+1}/(1+\gamma) \quad \text{as} \quad z \to +0.$$

The case $z < 0$ is handled analogously, and the result follows immediately.

Note that when $L(z) \to c$ as $z \to 0$, the Lemma gives the simple convex power law approximation $M(z) \approx 2c|z|^{1+\gamma}/(1+\gamma)$, which is quadratic when $\gamma = 1$; if, in addition, F is differentiable in a neighborhood of zero with $0 < F'(0) < \infty$, this in turn reduces to precisely the quadratic approximation used by Pollard (1991), since then $c = F'(0)$.

A.2 We are concerned here with the results based on (3) where the x_i are non-stochastic. The following follows Pollard's (1991) proof of his Theorem 1 closely, and much of the notation is also his. Let

$$G_n(\delta) = \Sigma_{i=1,n} [\,|u_i - x_i'\delta/a_n| - |u_i|\,],$$

with $\{a_n\}$ being a sequence to be defined below. Then $G_n(\delta)$ is a convex function of δ which is minimized at $\delta = a_n(b_n^{LAD} - \beta^o)$. Define

$$\Gamma_n(\delta) = E[G_n(\delta)] = \Sigma_{i=1,n} M(x_i'\delta/a_n),$$

(recalling the definition of $M(\cdot)$ in A.1) and note that if we let

$$R_{i,n}(\delta) = |u_i - x_i'\delta/a_n| - |u_i| + (x_i'\delta/a_n) \times \mathrm{sgn}(u_i),$$

$$W_n = - \Sigma_{i=1,n} x_i\, \mathrm{sgn}(u_i),$$

then $E[W_n] = 0$ if $F(0+) = F(0-)$, and

$$G_n(\delta) = \Gamma_n(\delta) + W_n'\delta/a_n + \Sigma_{i=1,n} \{R_{i,n}(\delta) - E[R_{i,n}(\delta)]\}$$

$$= \Gamma_n(\delta) + W_n'\delta/a_n + Q_n(\delta)$$

where

$$Q_n(\delta) = \Sigma_{i=1,n} R_{i,n}(\delta) - E[R_{i,n}(\delta)].$$

We now confine attention to the case where k = 1, and define $\{a_n\}$ by

$$L^*(\max_i\{|x_i|\}/a_n \times \Sigma_{i=1,n} |x_i/a_n|^{1+\gamma} = 1,$$

where

$$L^*(z) = \max\{L(z), L(-z)\}.$$

We assume that F satisfies the condition of the Lemma in A.1, and that for all $\upsilon \geq 1$,

$$\max_i\{|x_i|\} / (\Sigma_{i=1,n} |x_i|^{\upsilon})^{1/\upsilon} \to 0 \text{ as } n \to \infty.$$

This last condition implies that $\max_i\{|x_i|\}/a_n \to 0$; and if $L(z) \to c > 0$ as $z \to 0$, $\Sigma_{i=1,n} |x_i/a_n|^{1+\gamma} \to 1/c$ as $n \to \infty$, so $a_n^{1+\gamma} \approx c\, \Sigma_{i=1,n} |x_i|^{1+\gamma}$ for large n in this case. Now, from the Lemma in A.1,

$$\Gamma_n(\delta) = \Sigma_{i=1,n} M(x_i'\delta/a_n) = \Sigma_{i=1,n} [2L(x_i'\delta/a_n)|x_i'\delta/a_n|^{1+\gamma}/(1+\gamma)](1 + o(1)),$$

since $\max_i\{|x_i|\}/a_n \to 0$ ensures that $|x_i'\delta/a_n| \to 0$ for fixed δ uniformly for all x_i. So

$$\Gamma_n(\delta) = [|\delta|^{1+\gamma}/(1+\gamma)] \times \Sigma_{i=1,n} [2L(x_i/a_n)|x_i/a_n|^{1+\gamma}](1+o(1))$$

$$= [|\delta|^{1+\gamma}/(1+\gamma)] \times 2L^*(\max_i\{|x_i|/a_n\}) \times \Sigma_{i=1,n} |x_i/a_n|^{1+\gamma}(1+o(1))$$

by adapting for sums rather than integrals the modification of Seneta's Theorem 2.6 used in A.1. Hence, from the definition of $\{a_n\}$,

$$\Gamma_n(\delta) = 2\,[|\delta|^{1+\gamma}/(1+\gamma)] + o(1) \text{ as } n \to \infty.$$

and we can therefore write

$$G_n(\delta) = 2\,[|\delta|^{1+\gamma}/(1+\gamma)] + o(1) + W_n'\delta/a_n + Q_n(\delta).$$

Next we follow Pollard (1991) in writing

$$E[Q_n(\delta)^2] = E\,[\Sigma_{i=1,n} R_{i,n}(\delta) - E[R_{i,n}(\delta)]\,]^2$$

$$\leq E\,[\Sigma_{i=1,n} R_{i,n}(\delta)^2\,]$$

$$\leq 4\,\Sigma_{i=1,n}\,|x_i'\delta/a_n|^2\,\Pr[\,|u_i|\,\leq\,|x_i'\delta/a_n|\,]$$

$$= 4\,\Sigma_{i=1,n}\,|x_i'\delta/a_n|^2[F(|x_i'\delta/a_n|) - F(-|x_i'\delta/a_n|)]$$

$$= 4\,\Sigma_{i=1,n}\,|x_i'\delta/a_n|^2[\,|x_i'\delta/a_n|^\gamma[L(|x_i'\delta/a_n|) + L(-|x_i'\delta/a_n|)]$$

$$\approx 4\times 2\ \ |\delta|^{2+\gamma}\,L^*(\max_i\{|x_i|\}/a_n)\times\Sigma_{i=1,n}\,|x_i/a_n|^{2+\gamma}(\,1+o(1))$$

$$\leq 4\times 2\ \ \max_i\{|x_i/a_n|\}\,L^*(\max_i\{|x_i|\}/a_n)\times\Sigma_{i=1,n}\,|x_i/a_n|^{1+\gamma}(1+o(1)).$$

So $E[Q_n^2]\rightarrow 0$ as $n\rightarrow\infty$, from $\max_i\{|x_i/a_n|\}\rightarrow 0$ and the definition of $\{a_n\}$, $Q_n(\delta)\rightarrow_p 0$ as $n\rightarrow\infty$, for each as a consequence, and we can write

$$G_n(\delta) = 2[\,|\delta|^{1+\gamma}/(1+\gamma)] + o(1) + W_n'\delta/a_n + o_p(1).$$

Next, recall that $G_n(\delta)$ is minimized at $\delta = a_n(b_n^{LAD}\text{-}\beta^o)$, let $\mu = \delta/a_n$, and write

$$G_n(a_n\mu)/\eta_n = 2[\,|a_n\mu|^{1+\gamma}/(1+\gamma)]/\eta_n + W_n'\mu/\eta_n + o_p(1/\eta_n),$$

where $\eta_n = (\Sigma_{i=1,n}\ x_i^2)^{1/2}$, so $W_n/\eta_n \rightarrow_d N(0,1)$ given $1\text{-}F(0) = F(0) = 1/2$ and $\max_i\{|x_i|\}/(\Sigma_{i=1,n}\ |x_i|^\upsilon)^{1/\upsilon}\rightarrow 0$ as $n\rightarrow\infty$ (setting $\upsilon = 2$ here). The minimizer of the approximation, $2c[\,|a_n\mu|^{1+\gamma}/(1+\gamma)]/\eta_n + W_n'\mu/\eta_n$, to $G_n(a_n\mu)/\eta_n$, satisfies the first order condition

$$2\times(a_n^{1+\gamma}/\eta_n)\mathrm{sgn}(\mu)\,|\mu|^\gamma = -\,W_n/\eta_n,$$

and it is from this that (3) and the results in Examples 1 - 3 follow, once additional arguments, including an appeal to Pollard's convexity lemma, are completed. Note that

$$a_n^{1+\gamma}/\eta_n\ = L(\max_i\{|x_i|/a_n\}\ \Sigma_{i=1,n}\,|x_i|^{1+\gamma}/(\Sigma_{i=1,n}\,x_i^2)^{1/2},$$

and the right hand side here is $O_p(n^{1/2})$ under conventional regressor assumptions (such as invoked in Example 2). The requirement for consistency here is $a_n^{1+\gamma}/\eta_n\rightarrow\infty$ as $n\rightarrow\infty$.

A.3 For the case of stochastic regressors, the argument in A.2 sometimes applies with little modification. This is true for Example 8 if $\kappa > 1+\gamma$, for Example 7(a) (where $\alpha > 1+\gamma$) and for Example 6, as explained in A.4 below.

However, the argument does not work for Example 8 when $\kappa < 1+\gamma$. But we can adapt Pollard's (1991) proof of his Theorem 3 to handle this case in the following way. Let

$$G_n(\delta) = \Sigma_{i=1,n} [\,|u_i - x_i'\delta/a_n| - |u_i|\,],$$

where now we choose $a_n = n^{1/\kappa}$, in anticipation of the rate of convergence we wish to demonstrate. $G_n(\delta)$ is minimized at $\delta = a_n(b_n^{LAD} - \beta^o)$, $G_n(0) = 0$, and

$$G_n(\delta) = \Sigma_{i=1,n} (x_i'\delta/a_n)\mathrm{sgn}(u_i) + 2\Sigma_{i=1,n} |x_i'\delta/a_n - u_i|\, I\{|u_i| \leq |x_i'\delta/a_n|\}$$

$$= G_{1n}(\delta) + G_{2n}(\delta), \text{ say,}$$

with $I\{\cdot\}$ being the indicator function used in section 2 (i.e., $I\{z \leq w\} = 1$ if $z \leq w$, $I\{z \leq w\}$ otherwise). Following Pollard's (1991) idea, we want to show that, for large $|d|$, $\Pr[G_n(d) > 0] \to 1$ as $n \to \infty$, since this forces the minimizer of $G_n(\delta)$, (i.e., $a_n(b_n^{LAD} - \beta^o)$) to lie in [-d,d] with arbitrarily high probability for all n sufficiently large. Here, in $G_n(\delta)$,

$$G_{1n}(\delta) = \Sigma_{i=1,n} (x_i'\delta/a_n)\mathrm{sgn}(u_i) = \delta \times n^{-1/\kappa}\Sigma_{i=1,n} x_i \mathrm{sgn}(u_i) = O_p(1),$$

so we want to show that $\Pr[G_{2n}(\delta) > q] \to 1$ as $n \to \infty$, for $|\delta| > d$, for large d, q. Now,

$$G_{2n}(\delta) = 2\,\Sigma_{i=1,n} |x_i'\delta/a_n - u_i| \times I\{|u_i| \leq |x_i'\delta/a_n|\}$$

$$\geq 2\,\Sigma_{i=1,n} |x_i'\delta/a_n - u_i| \times I\{|u_i| \leq \theta|x_i'\delta/a_n|\},$$

$$\geq 2\,\Sigma_{i=1,n} (1-\theta)|x_i'\delta/a_n| \times I\{|u_i| \leq \theta|x_i'\delta/a_n|\}$$

for $\theta \in (0,1)$. Note that the expectation of the ith summand here conditional on x_i is

$$(1-\theta)|x_i'\delta/a_n| \times [F(\theta|x_i'\delta/a_n|) - F(0) + F(0) - F(-\theta|x_i'\delta/a_n|)]$$

$$\approx 2\,c(1-\theta)\theta^{\gamma}|x_i'\delta/a_n|^{1+\gamma} \text{ for small } \theta|x_i'\delta/a_n|,$$

by virtue of the regularly varying property of F at zero, i.e., (2) with $L(z) \to c > 0$ as $z \to 0$. To make use of this we consider x_i for which $\theta|x_i'\delta/a_n| \leq \varepsilon > 0$, for small ε. So,

$$G_{2n}(\delta) \geq 2\, \Sigma_{i=1,n} z_i,$$

$$z_i \equiv (1-\theta)|x_i'\delta/a_n| \times I\{|u_i| \leq \theta|x_i'\delta/a_n|\} \times I\{|x_i| \leq a_n\varepsilon/(\theta|\delta|)\},$$

and the expectation of the z_i is approximately

$$E[2c(1-\theta)\, \theta^{\gamma}\, |\delta/a_n|^{1+\gamma}\, |x_{i,n}^*|^{1+\gamma}]$$

where $x_{i,n}^* = x_i$ if $x_i \leq a_n\varepsilon/(\theta|\delta|)$, $x_{i,n}^* = 0$ otherwise. Now, from the assumption on the common distribution of the x_i,

$$E[|x_i^*|^{1+\gamma}] = B_n \times (\varepsilon/(\theta|\delta|))^{1+\gamma-\kappa}\, a_n^{1+\gamma-\kappa}, \qquad B_n \to B > 0 \text{ as } n \to \infty,$$

noting that $1+\gamma-\kappa > 0$ for the case we are interested in here. So, for large n,

$$E[z_i] \approx B_n \times 2c(1-\theta)\, \theta^{\gamma}\, |\delta/a_n|^{1+\gamma}(\varepsilon/(\theta|\delta|))^{1+\gamma-\kappa}\, a_n^{1+\gamma-\kappa}$$

$$\approx \{B\, 2c(1-\theta)\, \theta^{\gamma}\, |\delta|^{\kappa}\, (\varepsilon/\theta)^{1+\gamma-\kappa}\}/n$$

since $a_n = n^{1/\kappa}$. This is increasing in $|\delta|$. Similarly,

$$E[z_i^2] \approx \{C_n \times 2c(1-\theta)^2\, \theta^{\gamma}\, |\delta|^{\kappa}\, (\varepsilon/\theta)^{2+\gamma-\kappa}\}/n, \qquad C_n \to C > 0 \text{ as } n \to \infty,$$

so $\text{var}[\Sigma_{i=1,n} z_i]$ has a finite limit as $n \to \infty$. So, by Chebychev's inequality, for large n, $\Pr[\Sigma_{i=1,n} z_i \leq q]$ can be made as close to zero as desired for fixed $q > 0$ by making $|\delta|$ large. In view of $G_{2n}(\delta) \geq 2\, \Sigma_{i=1,n} z_i$ it follows that the rate of convergence of $b_n^{LAD}-\beta^o$ is $O_p(n^{-1/\kappa})$. Note that this approach, which can be adapted for Example 7(b), does not seem suitable for showing that $b_n^{LAD}-\beta^o$ has a faster rate of convergence when $1+\gamma < \kappa$.

A.4 Here we first see how the arguments in A.2 can be modified to obtain the convergence rate for the LAD estimator in Example 6. Let $G_n(\delta) = \Sigma_{i=1,n}[|u_i-x_i'\delta/a_n| - |u_i|]$, as before with

$$a_n = n^{1/\alpha} n^{1/(1+\gamma)},$$

and recalling that $x_i = y_{i-1} = \Sigma_{i=1,i-1}\, u_j$, we let $X_i = \{u_1,\dots,u_{i-1}\}$ and redefine

$$M(x_i'\delta/a_n) = (E[(|u_i - x_i'\delta/a_n| - |u_i|) \mid X_i]),$$

$$\Gamma_n(\delta) = E[G_n(\delta) \mid X_i] = \Sigma_{i=2,n} M(x_i'\delta/a_n)$$

so, if, as before

$$R_{i,n}(\delta) = |u_i - x_i'\delta/a_n| - |u_i| + (x_i'\delta/a_n) \times \mathrm{sgn}(u_i), \quad W_n = -\Sigma_{i=2,n}\, x_i\, \mathrm{sgn}(u_i),$$

then,

$$G_n(\delta) = \Gamma_n(\delta) + W_n'\delta/a_n + Q_n(\delta) \quad \text{for} \quad Q_n(\delta) = \Sigma_{i=2,n} R_{i,n}(\delta) - E[R_{i,n}(\delta) \mid X_i].$$

Next note that

$$R_{i,n}(\delta) = 2|u_i - x_i'\delta/a_n| I\{|u_i| \le |x_i'\delta/a_n|\}$$

$$\le 2|x_i'\delta/a_n|\, I\{|u_i| \le |x_i'\delta/a_n|\}$$

and, following the same lines as Knight (1989, pp 270 - 272), introduce the truncation

$$R_{i,n}^*(\delta) = 2|x_i'\delta/a_n|\, I\{|u_i| \le |x_i'\delta/a_n|\}\, I\{|x_i'\delta/a_n| \le M\}, \; M < \infty,$$

so that the summands in

$$Q_n^*(\delta) = \Sigma_{i=2,n} \{R_{i,n}^*(\delta) - E[R_{i,n}^*(\delta) \mid X_i]\}$$

are martingale differences. To show that $Q_n^*(\delta) \to_p 0$, we therefore want to show that

$$\Sigma_{i=2,n} E[R_{i,n}^*(\delta)^2 \mid X_i]] \to_p 0.$$

Now, from (2) with $L(z) \to c$ as $z \to 0$,

$$\Sigma_{i=2,n} E[R_{i,n}^*(\delta)^2 \mid X_i]] \approx_a \Sigma_{i=2,n} 4|x_i'\delta/a_n|^2\, 2c|x_i'\delta/a_n|^\gamma\, I\{|x_i'\delta/a_n| \le M\}$$

$$\le \max_i\{|x_i'\delta/a_n|\}\, \Sigma_{i=2,n} 4|x_i'\delta/a_n|^{1+\gamma}\, 2c\; I\{|x_i'\delta/a_n| \le M\}$$

$$\to_p 0 \text{ as } n \to \infty$$

because

$$\max_i\{|x_i/a_n|\} = \max_i\{|x_i/n^{1/\alpha}|\}/n^{1/(1+\gamma)} \to_p 0, \text{ as } n \to \infty,$$

and

$$\Sigma_{i=2,n} |x_i'\delta/a_n|^{1+\gamma} = \Sigma_{i=2,n} |x_i'\delta/n^{1/\alpha}|^{1+\gamma}/n = O_p(1), \text{ as } n \to \infty$$

from the definition of $\{a_n\}$ and the assumption on the common distribution of the u_i. Hence $Q_n^*(\delta) \to_p 0$ as $n \to \infty$. To show this also true of $Q_n(\delta)$, we use

$$\Pr[\Sigma_{i=2,n} 2|x_i'\delta/a_n| I(|u_i| \le |x_i'\delta/a_n|) I\{|x_i'\delta/a_n| > M\} > \varepsilon]$$

$$\le \Pr[\max_i\{|x_i'\delta/a_n|\} > M] \to 0 \text{ as } n \to \infty.$$

Next, from the Lemma in A.1,

$$\Gamma_n(\delta) = \Sigma_{i=1,n} M(x_i'\delta/a_n)$$

$$= [|\delta|^{1+\gamma}/(1+\gamma)] \times \Sigma_{i=1,n} 2c\, |x_i/a_n|^{1+\gamma}(1+o_p(1))$$

$$= 2\,[|\delta|^{1+\gamma}/(1+\gamma)]O_p(1) + o_p(1) \text{ as } n \to \infty,$$

again from the definition of $\{a_n\}$, with the $O_p(1)$ term independent of δ and so we can write

$$G_n(\delta) = 2\,[|\delta|^{1+\gamma}/(1+\gamma)]O_p(1) + o_p(1) + W_n'\delta/a_n$$

or,

$$G_n(a_n\mu)/\eta_n = 2[|a_n\mu|^{1+\gamma}/(1+\gamma)]O_p(1)/\eta_n + W_n'\mu/\eta_n + o_p(1/\eta_n),$$

where we now choose $\eta_n = n^{1/2}n^{1/\alpha}$ so $W_n/\eta_n = O_p(1)$, and the result follows essentially as in A.2: note that the convergence rate of the LAD estimator here is of order in probability

$$1/(a_n^{1+\gamma}/\eta_n)^{1/\gamma} = 1/[(n^{1/\alpha}n^{1/(1+\gamma)})^{1+\gamma}/\, n^{1/2}n^{1/\alpha}]^{1/\gamma} = 1/(n^{(1+\gamma)/\alpha}n^{1/2}/\,n^{1/\alpha})^{1/\gamma} = 1/n^{(1/\alpha + 1/2\gamma)},$$

as stated in Example 6.

Note that if we apply the same argument to Example 7, we redefine $X_i = u_{i-1} = x_i$, and first choose a_n so that $\Sigma_{i=2,n} |x_i/a_n|^{1+\gamma} = O_p(1)$. In (a), where $1+\gamma > \alpha \in (0,2)$, we can set

$a_n^{1+\gamma} = n$ because $\Sigma_{i=2,n} |x_i|^{1+\gamma}/n \to_p E[|x_i|^{1+\gamma}] < \infty$, and $\max_i\{|x_i/a_n|\} = \max_i\{|x_i/n^{1/(1+\gamma)}|\} \to_p 0$ because $\max_i\{|x_i/n^{1/\alpha}|\} = O_p(1)$, and these two points drive the derivation just given above. For (b) $1+\gamma < \alpha \in (0,2)$ we set $a_n^{1+\gamma} = (n^{1/\alpha})^{1+\gamma}$ to get $\Sigma_{i=2,n}|x_i/a_n|^{1+\gamma} = O_p(1)$ but now $\max_i\{|x_i/a_n|\} = \max_i\{|x_i/n^{1/\alpha}|\} = O_p(1)$ and so we need to make use of an alternative derivation, such as that in A.3 above. The same ideas apply for Example 8, once we redefine X_i suitably.

References

Chan, N.H., and L.T. Tran (1989), "On the first Order Autoregressive Process with Infinite Variance," *Econometric Theory*, 5, 354-362.

Huber, P (1981), *Robust Statistics*, Wiley, New York.

Ibragimov, I.A., and Y.V. Linnik (1971), *Independent and Stationary Sequences of Random Variables*,

Knight, K. (1989), "Limit Theory for Autoregressive Parameters in an Infinite Variance Random Walk," *Canadian Journal of Statistics*, 17, 261-278.

Loeve, M. (1977), *Probability Theory I*, 4th edition, Springer-Verlag, New York.

Phillips, P.C.B. (1990), "Time Series Regression with a Unit Root and Infinite Variance Errors," *Econometric Theory*, 6, 44-62.

Phillips, P.C.B. (1991), "A Shortcut to LAD Asymptotics," *Econometric Theory*, 7, 450-463.

Pollard, D. (1991), "Asymptotics for Least Absolute Deviation Regression Estimators," *Econometric Theory*, 7, 186-199.

Seneta, E (1976), *Regularly Varying Functions*, Springer-Verlag, New York.

Serfling, R.J. (1980), *Approximation Theorems of Mathematical Statistics*, Wiley, New York.

Smirnov, N.V. (1952), "Limit Distributions for the Terms of a Variational Series," *American Mathematical Society*, Translation Number 67.