Econometric Theory, **17**, 2001, 820–852. Printed in the United States of America.

LEAST ABSOLUTE DEVIATIONS REGRESSION UNDER NONSTANDARD CONDITIONS

ALAN J. ROGERS
University of Auckland

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 that is both positive and finite at zero. We consider the implications of weakening this assumption in a number of regression settings, primarily with a time series orientation. These models include ones with deterministic and stochastic trends, and we pay particular attention to the case of a simple unit root model. The way in which the conventional assumption on the error distribution is modified is motivated in part by N.V. Smirnov's work on domains of attraction in the asymptotic theory of sample quantiles. The approach adopted usually allows for simple characterizations (often featuring a single parameter, γ), of both the shapes of the limiting distributions of the LAD estimators and their convergence rates. The present paper complements the closely related recent work of K. Knight.

1. INTRODUCTION

The least absolute deviations (LAD) estimator of the coefficient vector in a linear regression model has long been of interest, especially as an alternative to the ordinary least squares (OLS) estimator. More recently, the asymptotic properties of the LAD estimator have come under closer scrutiny, with quite a lot of attention being paid to alternative ways of obtaining the limiting distribution of this estimator. Under the assumptions conventionally made, this estimator has an $O_p(n^{1/2})$ rate of convergence and, upon normalization, a Gaussian limiting distribution with covariance matrix proportional to $1/(2f(0))^2$ where $f(0)$ is the common density of the independent disturbances at the median, zero, of their distribution. Two new approaches to obtaining these asymptotics in a variety of linear regression contexts are the convexity arguments of Knight (1989, 1991, 1993), Pollard (1991), and Davis, Knight, and Liu (1992) and the use of generalized functions by Phillips (1991).

An early version of this paper was presented to the first meeting of the New Zealand Econometrics Study Group held at the University of Auckland in January 1997. Thanks go to the participants for their comments. I am especially grateful to two referees, one of whom made detailed suggestions about the structure of the proof of Proposition 2. I am also indebted to the editor for his guidance and specific suggestions and for making preprints of K. Knight's papers (1998a, 1998b) available. Address correspondence to: Alan J. Rogers, Department of Economics, University of Auckland, Private Bag 92019, Auckland, New Zealand; e-mail: a.rogers@auckland.ac.nz.

 © 2001 Cambridge University Press 0266-4666/01 $9.50

https://doi.org/10.1017/S0266466601174074 Published online by Cambridge University Press

This paper is concerned with the consequences of weakening the usual assumption on the disturbance distribution in a number of simple regression settings. Given the presence of $f(0)$ in the formula for the limiting distribution of the LAD estimator in the standard case, 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 disturbances in regression contexts, such as symmetry, existence of variance, and so on.

In fact, for the location model, where the LAD estimator is just the sample median, the analytics have been comprehensively developed by Smirnov (1952), although his work does not seem to have attracted the attention it deserves. (An exception is Serfling, 1980.) It is interesting from a number of perspectives, one of which is the parallels it yields between the effects of nonstandard distributional assumptions on the asymptotics for the LAD estimator and the well known effects of other nonstandard distributional assumptions on the asymptotics of the ordinary least squares (OLS) estimator. For the LAD estimator, the nonstandard features that matter are, not surprisingly, at the center of the distribution, whereas for the OLS estimator it is the tails of the distribution that are important.

The idea of relaxing the usual distributional assumption in regression models has been pursued very recently by Knight (1998a, 1998b). The present paper is concerned with the same general problem and is based on work originally done independently of Knight's. To avoid duplication we concentrate on models not covered by his work, with particular emphasis given to a simple random walk, with both finite and infinite variance errors. This last aspect of the paper extends the results of Knight (1989) and Herce (1996).

The direction in which this paper weakens the usual assumption on the distribution of the disturbances was initially motivated by the ease with which it yields some definite and easily understood results, but, perhaps more important, it was also motivated by the connection it has to the domains of attraction for the limiting distribution of the sample median (and other sample quantiles) in Smirnov's (1952) study of the location model. This also means that the extent to which the error distribution deviates from the conventional requirements is captured by a single parameter, γ. The approach adopted by Knight (1998a) is one suggested quite naturally by his analytics, and the latter provide a powerful and versatile way of obtaining a range of results. We see, however, that the two approaches are very closely related.

2. PRELIMINARIES

We are concerned with the least absolute deviations (LAD) estimator of β_o in the regression model

$$y_i = x_i' \beta_o + u_i, \qquad i = 1, \ldots, n,$$

https://doi.org/10.1017/S0266466601174074 Published online by Cambridge University Press

in which the u_i are independently and identically distributed and the x_i are $k \times 1$ vectors of observations on k regressors. The distribution function of the u_i is denoted by F. We begin with the observation that expressions of the form $F(z) - F(0)$ for z near zero arise frequently in the study of LAD estimators. One way in which this happens is as follows. The LAD estimator of β_o, denoted by b_n^{LAD}, is any $\beta \in \Re^k$ that minimizes

$$\sum_{i=1}^{n} |y_i - x_i' \beta|,$$

but it is convenient to work instead with the objective function

$$Z_n(\delta) = (\phi_n / \upsilon_n) \sum_{i=1}^{n} (|u_i - x_i' \delta / \phi_n| - |u_i|), \tag{1}$$

where $\{\phi_n\},\{\upsilon_n\}$ are suitably chosen sequences of constants; for each n, $Z_n(\delta)$ is a convex function of δ, and any $\delta \in \Re^k$ that minimizes $Z_n(\delta)$ is equal to $\phi_n(b_n^{LAD} - \beta_o)$. In what follows we let

$$d_n = b_n^{LAD} - \beta_o.$$

Next, we let $I(\cdot)$ denote the indicator function and note that we can make use of the identities

$$\begin{aligned} |w - z| - |w| &= -z \operatorname{sgn}(w) + 2(z - w)\{I(z > w \geq 0) - I(z < w < 0)\} \\ &= -z \operatorname{sgn}(w) + 2 \int_0^z \{I(w \leq s) - I(w < 0)\} ds, \end{aligned}$$

in which $\operatorname{sgn}(w) = w/|w|$ for $w \neq 0$, $\operatorname{sgn}(w) = 1$ for $w = 0$; these two identities are almost the same as those in, respectively, Knight (1989) and Knight (1998a, 1998b), with the first being equivalent to one used by Pollard (1991). Applying the second to the summands in $Z_n(\delta)$ enables us to write

$$Z_n(\delta) = A_n(\delta) + B_n(\delta), \tag{2}$$

where

$$\begin{aligned} A_n(\delta) &= (\phi_n / \upsilon_n) \sum_{i=1}^{n} (-x_i' \delta / \phi_n) \operatorname{sgn}(u_i) \\ &= -(1/\upsilon_n) \sum_{i=1}^{n} x_i' \delta \operatorname{sgn}(u_i) \end{aligned} \tag{3}$$

and

$$B_n(\delta) = 2(\phi_n / \upsilon_n) \sum_{i=1}^{n} \int_0^{x_i' \delta / \phi_n} \{I(u_i \leq s) - I(u_i < 0)\} ds. \tag{4}$$

https://doi.org/10.1017/S0266466601174074 Published online by Cambridge University Press

Here $A_n(\delta)$ is clearly linear in δ, and $\{v_n\}$ can usually be chosen so $A_n(\delta)$ has known behavior as $n \to \infty$. (In many cases, $v_n = n^{1/2}$.) The dependence of this behavior on F is, from (3), evidently minimal. The nature of the dependence on δ of $B_n(\delta)$ is more involved, but we note that if, e.g., the x_i are nonstochastic, then, whenever F is continuous at zero (implying $E[I(u_i < 0)] = F(0)$),

$$E[B_n(\delta)] = 2(\phi_n/v_n)\sum_{i=1}^{n}\int_0^{x_i'\delta/\phi_n}\{F(s) - F(0)\}ds$$

$$= 2v_n^{-1}\sum_{i=1}^{n}\int_0^{x_i'\delta}\{F(s/\phi_n) - F(0)\}ds,$$

so that if $B_n(\delta) - E[B_n(\delta)]$ is negligible in some sense when n is large, the dependence of $B_n(\delta)$ on the shape of F near zero is clear. To proceed further we need to impose more structure. The approach taken by Knight (1998a) is to assume that $E[B_n(\delta)]$ tends to a convex limit function that takes on values (not all infinite or all zero) in $[0,\infty]$.

Alternatively, but in a similar vein, one can make assumptions on the sequence of functions

$$\Psi_n(z) = n^q\int_0^z\{F(s/\rho_n) - F(0)\}ds$$

for fixed $q > 0$ and sequence of constants, $\{\rho_n\}$, as in Knight (1998a), or assumptions on

$$\psi_n(z) = n^q\{F(z/\rho_n) - F(0)\},$$

as in Knight (1998b). The potential for such assumptions to be applicable to an examination of the limiting behavior of $B_n(\delta)$ should be clear. (For example, if $v_n = n^{1/2}$ it is natural to choose $q = \frac{1}{2}$ and $\rho_n = \phi_n$.) In particular, if the x_i are stochastic (with x_i and u_i independent for each i), and if we denote the summands in (4) by $b_{i,n}(\delta)$, so $B_n(\delta) = \sum_{i=1}^{n} b_{i,n}(\delta)$, then

$$E[b_{i,n}(\delta)|x_i] = 2v_n^{-1}\int_0^{x_i'\delta}\{F(s/\phi_n) - F(0)\}ds,$$

and the limiting behavior of $B_n(\delta)$ can sometimes be explored via these these conditional expectations.

The approach adopted in an early version of the present paper was to assume that $F(z) - F(0)$ is regularly varying at zero. This assumption can be motivated by the way in which it yields simple results in many important cases and also by its historical importance in the theory of sample quantiles (see Smirnov, 1952), and because it is a natural way of relaxing the assumption that F has a finite, positive density at zero, while retaining some smoothness in F. Specifically, we will often make use of the following assumption.

https://doi.org/10.1017/S0266466601174074 Published online by Cambridge University Press

A1. F is continuous at zero with $F(0) = \frac{1}{2}$, and defining

$$G(z) \equiv F(z) - F(0),$$
$$M(z) \equiv \max\{|G(-z)|, |G(z)|\},$$

then $\lim_{z\to 0}\{-G(z)/G(-z)\}$ exists although it may be infinite, and $M(z)$ satisfies

$$M(z) = L(z)|z|^{\gamma},$$

where $\gamma > 0$ and L is an even, slowly varying function at zero (i.e., $L(z) = L(-z)$ for all z, and $L(\lambda z)/L(z) \to 1$ as $z \to 0$ for any $\lambda \neq 0$) with $L(z) > 0$ for all $z \neq 0$. That is, M is regularly varying at zero with exponent $\gamma > 0$.

A1 is equivalent to the requirement that F belong to one of three out of four possible domains of attraction for the sample median in the location model (where $k = 1$ and $x_i' = 1$ for all i). See Smirnov (1952). Under A1, $F(z) - F(0)$ is regularly varying at zero (with positive exponent, γ) from above and/or below and need not be symmetric around zero, unlike M and L, which have this property by construction. It is interesting to note the role regular variation at zero plays in the (nonstandard) limit theory for the sample median, just as regular variation at infinity has a vital role in the limit theory for the sample mean in situations featuring infinite variances. In view of its prominence in what follows it will be convenient to refer to γ as the center exponent (of F).

The following result provides a linkage between A1 and the limiting behavior of functions like ψ_n and Ψ_n defined previously. Its proof and that of Lemma 2, which follows, appear in Appendix A.

LEMMA 1. *For $G(z) \equiv F(z) - F(0)$ and $z \geq 0$, the following are equivalent:*

(i) $G(z) = L(z)z^{\gamma}$ *for some* $\gamma > 0$, *where L is slowly varying at zero from above;*

(ii) for a given $q > 0$, *there exists a sequence* $\{\rho_n\}$, *with* $\rho_n \to \infty$ *as* $n \to \infty$, *such that for each* $z \geq 0$, $\psi_n(z) \equiv n^q G(z/\rho_n) \to \psi(z)$ *as* $n \to \infty$, *where ψ is continuous at zero,* $\psi(0) = 0$, *and* $\psi(z^o) \in (0,\infty)$ *for a continuity point,* $z^o > 0$, *of ψ;*

(iii) for a given $q > 0$, *there exists a sequence* $\{\rho_n\}$, *with* $\rho_n \to \infty$ *as* $n \to \infty$, *such that for each* $z \geq 0$

$$\Psi_n(z) \equiv n^q \int_0^z G(s/\rho_n)ds \to \Psi(z) \quad \text{as } n \to \infty$$

and there is an open interval $(z_1, z_2) \subset \Re^+$ *on which Ψ is finite and strictly convex.*

So, requiring that F satisfy A1 is less restrictive than it might first appear and is closely related to weak assumptions concerning the existence and nature of limits of the functions ψ_n, Ψ_n. We note that q and $\{\rho_n\}$ are connected via γ by $\rho_n = n^{q/\gamma}\ell(1/n)$ where ℓ varies slowly at zero. The implication running from (ii) to (i) has a similar flavor to a result of Feller (1966). The result of most interest in the present context is the equivalence between (i) and (iii). The strict

https://doi.org/10.1017/S0266466601174074 Published online by Cambridge University Press

convexity of the limit function Ψ in (iii) means that typically $B_n(\delta)$ is also approximated for large n by a strictly convex function if (iii) holds, and the equivalence between (i) and (iii) means that this approximating function has a particularly simple form. Moreover, this approximation can be obtained by exploiting a uniform convergence property of regularly varying functions.

Because we are interested in generalizing the conventional assumption on F, it is perhaps natural to ask what happens when A1 is not satisfied. A partial answer is provided by the following lemma.

LEMMA 2. *For $G(z) \equiv F(z) - F(0)$, suppose that $G(z)$ is strictly increasing at all $z \in (-z^*, z^*)$ for some $z^* > 0$ and $G(z) \to 0$ as $z \to 0$.*

(a) If G is slowly varying at zero from above then for any $q > 0$ there exists a sequence $\{\rho_n\}$ such that, for each $z \geq 0$,

$$\Psi_n(z) \equiv n^q \int_0^z G(s/\rho_n)ds \to z \quad \text{as } n \to \infty,$$

with $\rho_n \to \infty$ and $\rho_{\lfloor gn \rfloor}/\rho_n \to \infty$ as $n \to \infty$ for any $g > 1$; here $\lfloor \cdot \rfloor$ denotes integer part.

(b) If G is rapidly varying at zero from above in the sense that $G(\lambda z)/G(z) \to 0$ as $z \searrow 0$ for all $\lambda \in (0,1)$, then for any given $q > 0$ there exists a sequence $\{\rho_n\}$ such that $\psi_n(z) \equiv n^q G(z/\rho_n) \to \psi(z)$ as $n \to \infty$ for each $z \geq 0$, where

$$\psi(z) = \begin{cases} 0 & \text{if } z \in [0,1) \\ 1 & \text{if } z = 1 \\ \infty & \text{if } z \in (1,\infty) \end{cases}$$

with $\rho_n \to \infty$ and $\rho_{\lfloor gn \rfloor}/\rho_n \to 1$ as $n \to \infty$ for any $g > 0$. And, $\Psi_n(z) \equiv n^q \int_0^z G(s/\rho_n)ds$, $\Psi_n(z) \to \Psi(z)$ as $n \to \infty$, where

$$\Psi(z) = \begin{cases} 0 & \text{if } z \in [0,1] \\ \infty & \text{if } z \in (1,\infty) \end{cases}.$$

We provide an illustration of the application of Lemma 2 in Proposition 3, which follows. The condition of part (b) is closely related to the exact characterization of one of Smirnov's (1952) four domains of attraction for the sample median, this being the only one not covered by A1. The behavior of the sequences $\{\rho_n\}$ in Lemmas 1 and 2 turns out to be intimately related to the convergence rates for the LAD estimator under the associated conditions on F.

Taken together, condition (i) of Lemma 1 and the conditions of parts (a) and (b) of Lemma 2 cover all possibilities for the behavior of $F(z) - F(0)$ that are likely to be of interest, at least if we are prepared to confine attention to those F that are strictly increasing and reasonably smooth at zero.

For the models considered in the following sections the x_i take a form sufficiently specific that there is no need to impose additional assumptions on them. In each of these cases we assume for simplicity that F is symmetric around

https://doi.org/10.1017/S0266466601174074 Published online by Cambridge University Press

zero. This is inessential, and interesting things can happen when it is relaxed, but no issues of substance are raised that are not present in asymmetric cases in the location model. See Smirnov (1952) and also Knight (1998a). And in each of the models considered there is only one regressor (i.e., $k = 1$). Significant complications arise in extracting the limiting distributions of the LAD estimators of the individual coefficients when this is relaxed. Again, see Knight (1998a, 1998b). These difficulties do not arise in the standard case (subsumed by $\gamma = 1$ in A1) because then $Z_n(\delta)$ has a quadratic approximation. Nevertheless, the case $k = 1$ yields a wide enough range of results to be interesting. Proofs of all propositions appear in Appendix B.

3. DETERMINISTIC TREND

The following result is an application of the ideas of the previous section to the case of a single nonstochastic regressor that takes the form of a deterministic trend.

PROPOSITION 1. *Suppose that* $x_i = i^{\omega}$. *If* F *is symmetric around zero and satisfies A1, and*

(a) if $\omega > -\frac{1}{2}$, $\omega + \frac{1}{2}\gamma > 0$, *then* $\phi_n d_n \to_d \operatorname{sgn}(\xi)|\xi/[2(1+\gamma)]|^{1/\gamma}$, *where* $\phi_n = \ell^o(1/n)n^{\omega+1/2\gamma}$, ℓ^o *is a function that is slowly varying at zero, and* $\xi \backsim N(0,1)$;

(b) if $\omega < -\frac{1}{2}$ *and* $\omega(1+\gamma) > -1$, *then* $\phi_n d_n = O_p(1)$ *as* $n \to \infty$, *where* $\phi_n = \ell^o(1/n)n^{[1+\omega(1+\gamma)]/\gamma}$ *and* ℓ^o *varies slowly at zero, provided that whenever* $\omega < 0$, $L(z)$ *in A1 is nonincreasing in* $|z|$ *at each* $z \in (-z^*, z^*)$ *for* $z^* > 0$ *sufficiently close to zero.*

Considering part (a) of this result first, we see that the center exponent γ affects both the shape of the limiting distribution of $\phi_n d_n$ and the rate of convergence of d_n to zero as $n \to \infty$. This is as for the location model that can be obtained as a special case by setting $\omega = 0$ and also for some other regression models with well-behaved nontrending regressors. For the OLS estimator the rate of convergence is $O_p(n^{\omega+1/2})$ whenever $E[u_i^2] < \infty$ and $\omega > -\frac{1}{2}$. So in this case, the LAD estimator converges more rapidly when $\gamma < 1$; when $\gamma > 1$ the convergence rate is slower, and the condition of part (a) on ω is more stringent. The form of the limiting distribution of the estimator in part (a) follows from the fact that the components of $Z_n(\delta)$ defined in the preceding section satisfy

$$A_n(\delta) \to_d -\delta\xi, \qquad B_n(\delta) \to_p 2|\delta|^{1+\gamma},$$

where in the definitions of $A_n(\delta)$ and $B_n(\delta)$, in (3) and (4), $\{\phi_n\}$ is as defined in part (a) and $\{\upsilon_n\}$ is defined by $\upsilon_n = (\sum_{i=1}^n i^{2\omega})^{1/2}$.

https://doi.org/10.1017/S0266466601174074 Published online by Cambridge University Press

Proposition 1(b) is interesting in two respects. First, if $\gamma < 1$, the LAD estimator is consistent in circumstances where the OLS estimator is not. Second, writing

$$Z_n(\delta) = A_n(\delta) + E[B_n(\delta)] + \{B_n(\delta) - E[B_n(\delta)]\}$$

for suitable choices of $\{v_n\}$, $\{\phi_n\}$, $A_n(\delta)$ has a non-normal, nondegenerate limiting distribution, and, as for part (a), $E[B_n(\delta)] \to 2|\delta|^{1+\gamma}$, but we do not show that $2|\delta|^{1+\gamma}$ is also the probability limit of $B_n(\delta)$; hence a limiting distribution is not given in part (b).

We note also that the conditions $\omega > -\frac{1}{2}$ and $\omega > -\frac{1}{2}\gamma$ of part (a) imply $\omega(1 + \gamma) > \max\{-(1 + \gamma)/2, -(1 + \gamma)/2\gamma\} \geq -1$, so the condition $\omega(1 + \gamma) > -1$ is present in part (a) and also in part (b) (and in fact is used in the proof of part (a)). And, the condition on L that is imposed when $\omega < 0$ can easily be replaced by others, such as the requirement that L be bounded above and below away from zero on a small neighborhood of zero.

4. RANDOM WALK

We are concerned here with the model

$$y_i = \beta y_{i-1} + u_i, \qquad y_0 = 0$$

so that $x_i = y_{i-1}$ for $i = 2, \ldots, n$ and $x_1 = 0$. We will be interested in the LAD estimator in the random walk case for which $\beta = 1$. In the standard case where F has a finite, positive density at zero, some interesting results emerge from this model, especially when F is heavy tailed. See Chan and Tran (1989), Knight (1989, 1991), and Phillips (1991) and, for the case of finite variance disturbances, Herce (1996). To see how these and other results are affected by allowing for more general behavior of F at zero, we make the following assumption on the tails of F.

A2. F is symmetric around zero and lies in the normal domain of attraction of a symmetric stable law with index $\alpha \in (0,2]$, so that either $E[u_i^2] < \infty$ for $\alpha = 2$ or $E[u_i^2] = \infty$ for $\alpha \in (0,2)$, in which case

$$1 - F(z) = F(-z) \sim c_o z^{-\alpha} \quad \text{as } z \to \infty$$

with $c_o > 0$.

The symmetry assumption is, as before, adopted for convenience, and the same applies to the restriction to normal domains of attraction, because it simplifies expressions for sequences of normalizing constants (see Ibragimov and Linnik, 1971; Feller, 1966).

https://doi.org/10.1017/S0266466601174074 Published online by Cambridge University Press

PROPOSITION 2. *Suppose that the distribution function, F, of the u_i satisfies A1, A2, and $x_i = y_{i-1}$ for $i = 2,\ldots,n$, with $x_1 = y_0 = 0$. When $\beta = 1$,*

(a) if $E[u_i^2] = \sigma^2 < \infty$ (i.e., $\alpha = 2$),

$$\phi_n d_n \to_d \frac{1}{\sigma}\,\mathrm{sgn}\left(\int_0^1 W_1(s)dW_2(s)\right) \times \left(\frac{\left|\int_0^1 W_1(s)dW_2(s)\right|}{2(1+\gamma)\int_0^1 |W_1(s)|^{1+\gamma}ds}\right)^{1/\gamma},$$

in which W_1 and W_2 are dependent standard Wiener processes, and

$$\phi_n = n^{1/2\gamma+1/2}\ell^o(1/n),$$

where ℓ^o varies slowly at zero;

(b) if $E[u_i^2] = \infty$ (i.e., $\alpha \in (0,2)$),

$$\phi_n d_n \to_d \mathrm{sgn}\left(\int_0^1 S_\alpha(s)dW(s)\right) \times \left(\frac{\left|\int_0^1 S_\alpha(s)dW(s)\right|}{2(1+\gamma)\int_0^1 |S_\alpha(s)|^{1+\gamma}ds}\right)^{1/\gamma},$$

in which W is a standard Wiener process, S_α is a stable process with index α, independent of W, and

$$\phi_n = n^{1/2\gamma+1/\alpha}\ell^o(1/n),$$

where ℓ^o varies slowly at zero.

Part (b) of Proposition 2 generalizes a result of Knight (1989) to allow for the sort of behavior of F at zero in which we are interested in this paper. His result can be obtained by setting $\gamma = 1$ in A1 and requiring that the slowly varying function, L, in A1 satisfy $L(z) \to c \in (0,\infty)$ as $z \to 0$, this being necessary but not sufficient for F to have a finite, positive density at zero. In this case, the convergence rate of the LAD estimator is $O_p(n^{1/2+1/\alpha})$, which is more rapid than the $O_p(n)$ rate for the OLS estimator, with the latter being independent of the tail thickness parameter, α. See also Phillips (1991). We see from Proposition 2(b) that if γ in A1 is less than one, the convergence rate of the LAD estimator is even more rapid. If γ is greater than one, whether or not the LAD estimator converges more rapidly than the OLS estimator depends on the magnitudes of α and γ. If $\alpha \leq 1$ the LAD estimator converges more rapidly for any γ, and this superiority also holds if $\alpha \in (1,2)$ and $\gamma < \alpha/2(\alpha-1)$. However, when this inequality is reversed, i.e., $\gamma > \alpha/2(\alpha-1)$ (and $\alpha \in (1,2)$) the OLS estimator dominates the LAD estimator in terms of its rate of convergence. The borderline situation for which $\gamma = \alpha/2(\alpha-1)$ and $\alpha \in (1,2)$ is resolved by the behavior of the slowly varying function L in A1: if $L(z) \to c \in (0,\infty)$ as $z \to 0$, the LAD estimator has the $O_p(n)$ rate of the OLS estimator, a more rapid rate if $L(z) \to \infty$ as $z \to 0$ and a slower rate if $L(z) \to 0$ as $z \to 0$.

https://doi.org/10.1017/S0266466601174074 Published online by Cambridge University Press

This is another reflection of the intuition that the higher the concentration of probability assigned by F to a neighborhood of zero, the more efficient the LAD estimator is likely to be.

Part (a) of Proposition 2 generalizes the work of Herce (1996). Here there is a more direct parallel with Proposition 1(a) because, given $E[u_i^2] < \infty$, (i.e., $\alpha = 2$) any other characteristics of the tails of F are irrelevant for the asymptotics. When $\gamma < 1$, the LAD estimator has a more rapid convergence rate than the $O_p(n)$ rate of the OLS estimator, and it has a less rapid rate when $\gamma > 1$. When $\gamma = 1$, the relative performance is decided by the behavior of L at zero.

There is an important difference between the limiting distributions in parts (a) and (b) that has been emphasized and explained by Herce (1996): in (a) W_1 and W_2 are dependent Wiener processes, whereas the processes S_α, W in (b) are independent. The reasons given for this in the papers by Knight (1989, 1991) and Herce (1996) apply without qualification to the cases of interest here because the question revolves around the limiting behavior of $A_n(\delta) = -\delta n^{-1/2} \sum_{i=2}^{n} (y_{i-1}/n^{1/\alpha}) \operatorname{sgn}(u_i)$ $(\alpha \in (0,2])$, which is entirely unaffected by the generalization of interest in this paper, because the influence of this generalization is felt solely in $B_n(\delta)$, as indicated in Section 2. The processes W_2 and W in parts (a) and (b), respectively, are limits of the partial sums of $\{\operatorname{sgn}(u_i)\}$ upon normalization by $n^{1/2}$, whereas W_1 and S_α are limits of partial sums of $\{u_i\}$ upon normalization by $n^{1/\alpha}$, with $\alpha = 2$ for part (a), $\alpha \in (0,2)$ in part (b): as Herce (1996) notes, this last sequence of normalizing constants is responsible for the absence of any dependence between W and S_α that might be expected to be present in view of the dependence, for each i, between u_i and $\operatorname{sgn}(u_i)$.

The effects of the dependence in part (a) on the one hand and the independence in part (b) on the other emerge quite clearly from an examination of Figures 1–4, which depict simulated limiting distribution functions and associated estimates of the limiting densities.

Finally, we note that the shapes of the limiting distributions in parts (a) and (b) of Proposition 2 are affected by the center exponent, γ, as is apparent from the presence of the exponent $1/\gamma$ in the formulae. The effect is quite similar to that in Proposition 1(a) and is clearly discernible in Figures 1–4. As expected, the magnitude of γ has a dramatic effect on the nature of these distributions near zero.

The next set of results is concerned with the same random walk model where F satisfies A2 but not A1. With some caution they can be thought of as limiting cases obtained by letting γ in Proposition 2 tend either to infinity, for parts (a) and (b), or to zero, in part (c).

PROPOSITION 3. *Suppose that the distribution function, F, of the u_i satisfies A2 for some $\alpha \in (0,2]$, is strictly increasing at all $z \in (-z^*, z^*)$ for some $z^* > 0$, is continuous at zero with $F(0) = \frac{1}{2}$, and $x_i = y_{i-1}$ for $i = 2, \ldots, n$, with $x_1 = y_0 = 0$. When $\beta = 1$,*

https://doi.org/10.1017/S0266466601174074 Published online by Cambridge University Press

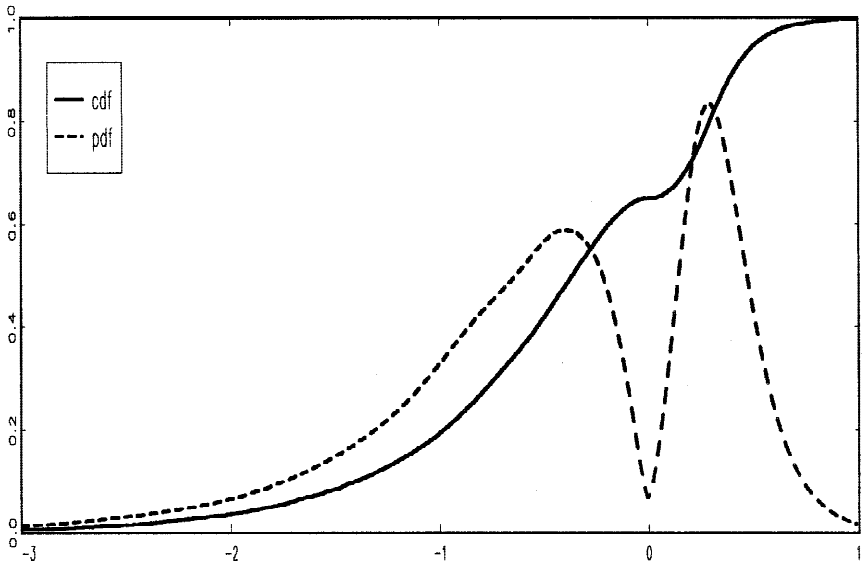


FIGURE 1. Limiting distribution for Proposition 2(a): $\gamma = 2$.

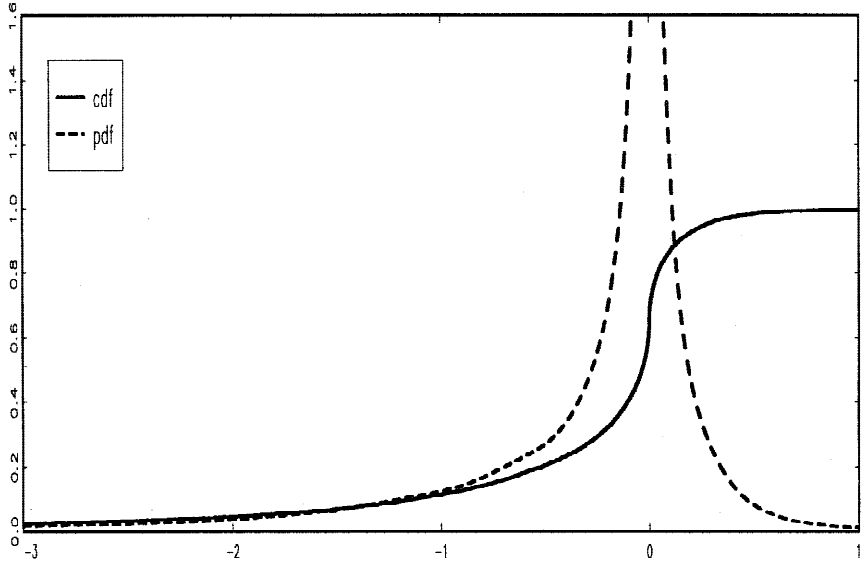


FIGURE 2. Limiting distribution for Proposition 2(a): $\gamma = 0.5$.

https://doi.org/10.1017/S0266466601174074 Published online by Cambridge University Press

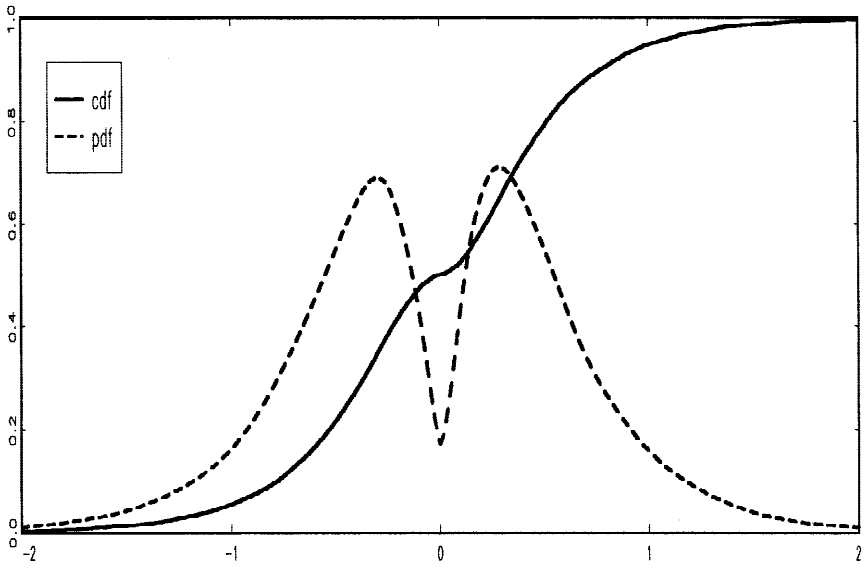


FIGURE 3. Limiting distribution for Proposition 2(b): $\gamma = 4.0$, $\alpha = 1.5$.

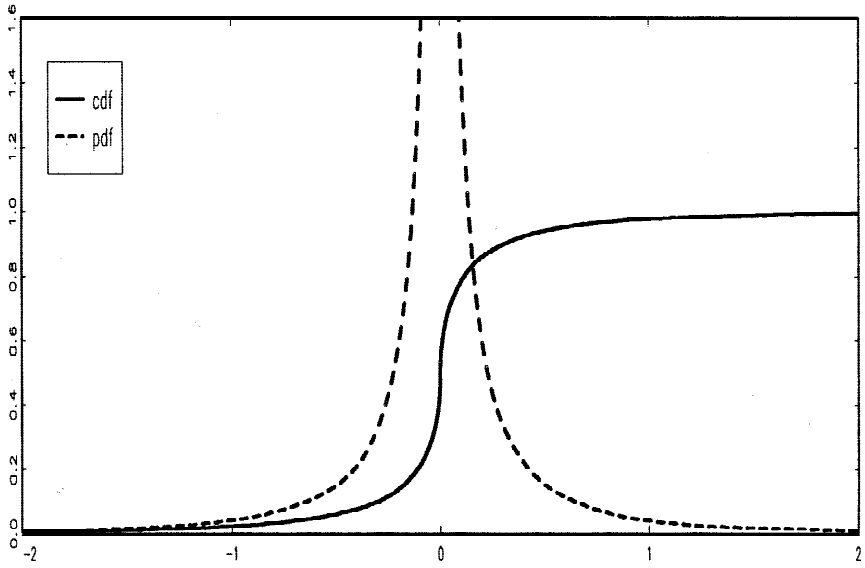


FIGURE 4. Limiting distribution for Proposition 2(b): $\gamma = 0.5$, $\alpha = 1.5$.

https://doi.org/10.1017/S0266466601174074 Published online by Cambridge University Press

(a) *if F is rapidly varying at zero in the sense of Lemma 2(b) and $E[u_i^2] = \sigma^2 < \infty$ (i.e., $\alpha = 2$), then*

$$\phi_n d_n \to_d \frac{\operatorname{sgn}\left(\int_0^1 W_1(r)dW_2(r)\right)}{\max_{0\le r\le 1}\{|\sigma W_1(r)|\}},$$

where W_1 and W_2 are dependent Wiener processes; and

$$\phi_n = n^{1/2}\rho_n,$$

$\{\rho_n\}$ being a sequence of constants satisfying $\rho_n/\rho_{\lfloor gn\rfloor} \to 1$ as $n \to \infty$, for any $g > 0$;

(b) *if F is rapidly varying at zero in the sense of Lemma 2(b) and $E[u_i^2] = \infty$ (i.e., $\alpha \in (0,2)$), then*

$$\phi_n d_n \to_d \frac{\operatorname{sgn}\left(\int_0^1 S_\alpha(r)dW(r)\right)}{\max_{0\le r\le 1}\{|S_\alpha(r)|\}},$$

where W is a Wiener process, and S_α is a stable process independent of W, and

$$\phi_n = n^{1/\alpha}\rho_n,$$

$\{\rho_n\}$ being a sequence of constants satisfying $\rho_n/\rho_{\lfloor gn\rfloor} \to 1$ as $n \to \infty$, for any $g > 0$;

(c) *if F is slowly varying at zero, then, for $\alpha \in (0,2]$,*

$$n^{1/\alpha}\rho_n d_n \to_p 0 \quad \text{as } n \to \infty$$

for any sequence, $\{\rho_n\}$, of constants satisfying $n^{1-1/v}[F(1/\rho_n) - F(0)] \sim 1$ as $n \to \infty$, for some $v < 2$.

In parts (a) and (b) of this result the limiting distributions take quite unusual forms, and these arise because of the nature of the stochastic processes followed by the regressors in each case. They do however have a parallel with a result of Smirnov's (1952) for the location model (i.e., the model of Proposition 1 with $\omega = 0$ but with F rapidly varying at zero). In this case the limiting distribution is bimodal (attaching probability of $\frac{1}{2}$ to each of the points -1 and 1), and the convergence rate is very slow. A related result with independently and identically distributed regressors is Example 4 of Knight (1998a). In parts (a) and (b) the convergence rate of the estimator is close to $O_p(n^{1/\alpha})$, $\alpha \in (0,2]$, a result that we might expect from parts (a) and (b) of Proposition 2, because the latter is approximately the convergence rate obtained there by making γ large. Part (a) of Proposition 3 in particular is of some interest as a curiosum because the convergence rate for the LAD estimator is close to the familar $O_p(n^{1/2})$ rate associated with many estimators in a wide variety of models with stationary regressors.

https://doi.org/10.1017/S0266466601174074 Published online by Cambridge University Press

The simulated limiting distribution functions and the associated estimated densities that appear in Figures 5 and 6 provide some insight into the shape of the limiting distributions in parts (a) and (b) of Proposition 3. The distribution for part (a) depicted in Figure 5 resembles an extreme version of Figure 3 (rather than, interestingly, Figure 1), whereas the infinite variance disturbances in part (b) have a notable smoothing effect, apparent in Figure 6.

In part (c) of Proposition 3, no limiting distribution is provided. Nor is a limiting distribution obtainable in a simple location model if $F(z) - F(0)$ is slowly varying at zero, because then there is no domain of attraction of the sample median that contains F; this follows from Smirnov (1952). In the present random walk model the reasoning is as follows. If we choose the sequences of constants $\{v_n\}$, $\{\phi_n\}$ in the definition of $Z_n(\delta)$ in (2) so that both $A_n(\delta)$ and $B_n(\delta)$ (in (3) and (4), respectively) have nondegenerate limiting distributions, then

$$\{A_n(\delta), B_n(\delta)\} \rightarrow_d \left\{-\delta \int_0^1 S_\alpha(r)dW(r),\ 2|\delta| \int_0^1 |S_\alpha(r)|dr\right\}$$

when $E[u_i^2] = \infty$ (i.e., $\alpha \in (0,2)$), and so whereas $Z_n(\delta) = A_n(\delta) + B_n(\delta)$ converges in distribution to a convex random function, $Z(\delta)$, of δ, this limiting function is not strictly convex: it is uniquely minimized at $\delta = 0$ if

$$\left|\int_0^1 S_\alpha(r)dW(r)\right| < 2\int_0^1 |S_\alpha(r)|dr$$

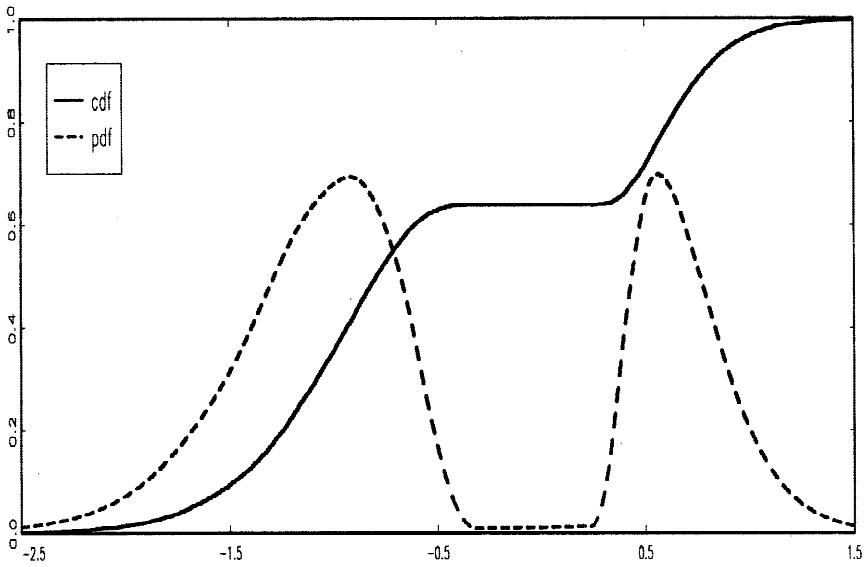


FIGURE 5. Limiting distribution for Proposition 3(a).

https://doi.org/10.1017/S0266466601174074 Published online by Cambridge University Press

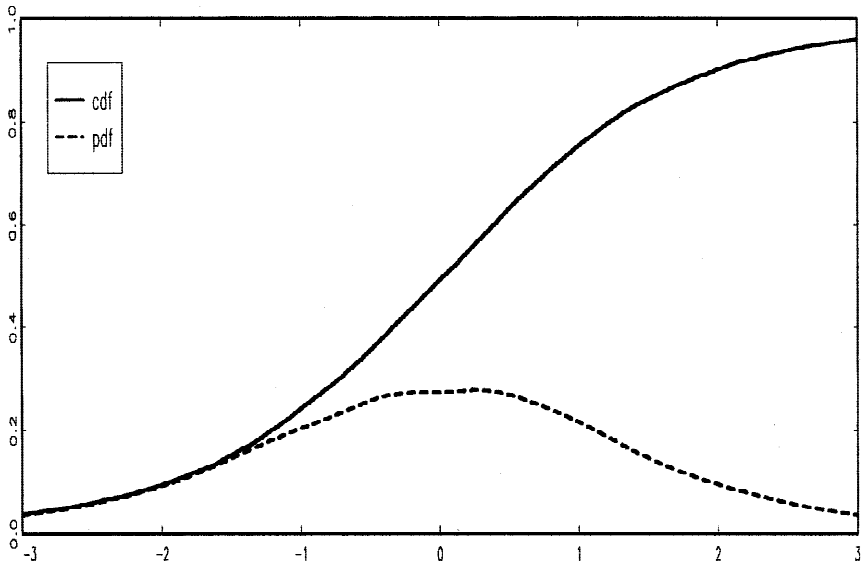


FIGURE 6. Limiting distribution for Proposition 3(b): $\alpha = 1$.

but at $\delta = \infty \times \text{sgn}(\int_0^1 S_\alpha(r)dW(r))$ if the inequality is reversed. So, with positive probability $\arg\min\{Z(\delta)\}$ does not exist. Here W, S_α are Wiener and stable processes, respectively. The argument for the case where $E[u_i^2] < \infty$ (i.e., $\alpha = 2$) is essentially the same. Hence no non-degenerate limiting distribution is obtainable, at least by the methods used here. As might be expected, the convergence rate of the LAD estimator is extremely rapid, and in particular, from Lemma 2 and the proof of Proposition 3(c), $n^{1/\alpha}d_n = o_p(n^{-\kappa})$ as $n \to \infty$, for arbitrarily large κ. As an example, suppose that $F(z) - F(0) \sim -\text{sgn}(z)/\ln(|z|)$ as $z \to 0$: in the statement of Proposition 3(c), we can choose $\{\rho_n\}$ so that $n^{1-1/v}[-1/\ln(1/\rho_n)] \sim 1$ as $n \to \infty$, for any $v < 2$, which implies that $d_n = o_p(n^{-1/\alpha}\exp(-n^{1/2-\varepsilon}))$ for any $\varepsilon > 0$.

5. STATIONARY AUTOREGRESSION

For the random walk model considered in the preceding section, the flavor of Proposition 2 is similar to that of Proposition 1(a), because in both cases the parameter γ plays a similar role: $\gamma < 1$ always means that the LAD estimator has a faster rate of convergence than the OLS estimator, and this is also true of models with more conventional regressors, such as Example 1 of Knight (1998a). In this section we examine the model $y_i = \beta y_{i-1} + u_i$, $y_0 = 0$ of the last section but with $|\beta| < 1$. We make the following assumption on the tails of F.

https://doi.org/10.1017/S0266466601174074 Published online by Cambridge University Press

A3. F is symmetric around zero, and either $E[|u_i|^\kappa] < \infty$ for any $\kappa > 0$ or, for some $\eta > 0$,

$$1 - F(z) = F(-z) \sim c_o z^{-\eta} \quad \text{as } z \to +\infty$$

with $c_o > 0$.

Note that if in A2 $\alpha < 2$, then α coincides with η in A3. If $\eta = 2$ in A3, then $E[u_i^2] = \infty$, and $\eta = 2$ has no counterpart in A2 in view of the restriction there to normal domains of attraction. In A3, if $\eta > 2$ then $E[u_i^2] < \infty$, but $E[|u_i|^\kappa] = \infty$ for some $\kappa > 2$. In general, η is the maximal moment exponent of F, i.e., $\eta = \sup\{\kappa : E[|u_i|^\kappa] < \infty\}$. There is an obvious parallel here with the center exponent, γ, of F in A1: γ is the maximal moment exponent of the distribution of $1/|u_i|$, because $\gamma = \sup\{\kappa : E[|u_i|^{-\kappa}] < \infty\}$.

PROPOSITION 4. *Suppose that the distribution function, F, of the u_i satisfies A3 and, for parts (a)–(d), A1. When $x_i = y_{i-1}$ for $i = 2, \ldots, n$, with $x_1 = y_0 = 0$ and $\beta = 0$,*

(a) if $E[u_i^2] < \infty$ and $E[|u_i|^{1+\gamma}] < \infty$, then

$$\phi_n d_n \to_d |\xi|^{1/\gamma} \operatorname{sgn}(\xi),$$

where $\phi_n = n^{1/2\gamma} \ell^o(1/n)$, ℓ^o is slowly varying at zero, and ξ is normally distributed with zero expectation and variance equal to

$$\frac{E[u_i^2]}{\{2(1+\gamma)E[|u_i|^{1+\gamma}]\}^{2/\gamma}};$$

(b) if $E[u_i^2] < \infty$, and $E[|u_i|^{1+\gamma}] = \infty$, so $1 + \gamma \geq \eta > 2$, then $n^\kappa d_n \to_p 0$ as $n \to \infty$ for any $\kappa < \frac{1}{2}(\eta - 1)$;

(c) if $E[u_i^2] = \infty$, and $E[|u_i|^{1+\gamma}] < \infty$, so $1 + \gamma < \eta \leq 2$, then $n^\kappa d_n \to_p 0$ as $n \to \infty$ for any $\kappa < (\eta - 1)/\eta\gamma$;

(d) if $E[u_i^2] = \infty$, $E[|u_i|] < \infty$, and $E[|u_i|^{1+\gamma}] = \infty$, so $1 + \gamma \geq \eta > 1$, $\eta \leq 2$, then $n^\kappa d_n \to_p 0$ as $n \to \infty$ for any $\kappa < 1/\eta$;

(e) if $E[|u_i|] = \infty$ with $\eta < 1$, then $d_n = O_p(n^{-1/\eta})$ as $n \to \infty$.

The condition $\beta = 0$ is invoked to simplify the derivations and because we are concerned here only with providing an indication of the sort of results that can be obtained. Part (a) is a special case of a result due to Knight (1998b), except that we verify $E[B_n(\delta)] \to_p 2|\delta|^{1+\gamma} E[|u_i|^{1+\gamma}]$ directly using the regular variation of F in A1. The parameter γ plays the same role here as in Propositions 1 and 2: it affects the rate of convergence and the shape of the limiting distribution, with the standard result again obtained by setting $\gamma = 1$ and requiring that L in A1 satisfy $L(z) \to c \in (0, \infty)$ as $z \to 0$.

Parts (b)–(d) require that $E[|u_i|^\kappa] = \infty$ for certain values of κ and different configurations of the parameters η and γ. The first point to note about these results is they provide some information only on the convergence rates of the LAD estimator. For part (c) this is so essentially for the same reason Proposi-

https://doi.org/10.1017/S0266466601174074 Published online by Cambridge University Press

tion 1(b) is incomplete in a similar sense: we can determine the limiting behavior of $A_n(\delta)$ and $E[B_n(\delta)]$ for choices of the sequences $\{v_n\}$ and $\{\phi_n\}$ that would deliver a limiting distribution of $\phi_n d_n$, but difficulties arise in establishing $B_n(\delta) - E[B_n(\delta)] \rightarrow_p 0$ as $n \rightarrow \infty$ for such sequences. For parts (b), (d), and (e) there are related technical difficulties concerning the limiting behavior of $B_n(\delta)$.

The second feature of these results is the role played by the center exponent, γ, in (b) and (d): there the convergence rates are determined entirely by the tail parameter η, whenever η, γ satisfy the inequality

$$1 + \gamma \geq \eta,$$

which says that F has a relatively low concentration of probability near zero compared to the magnitude of the tail probabilities. Note that in the standard case where $\gamma = 1$, this condition is simply $\eta \leq 2$, implying that the u_i have infinite variance. When the inequality does not hold, so that $1 + \gamma < \eta$, the amount of probability near zero is large enough to influence the convergence rate, and, as in Propositions 1 and 2, a smaller value of γ yields a faster convergence rate, for a given value of η. But this influence of γ is again constrained by the magnitude of η: if $\eta < 1$, so that the u_i do not have finite expectation, then the nature of F near zero is entirely irrelevant for the convergence rate of the estimator, and this is the content of part (e). Notice that part (e) does not require that the median, zero, of F be unique, and the symmetry requirement in A3 can also easily be dispensed with.

In fact, the results of Davis et al. (1992) can be applied to the model of Proposition 4 to show that $d_n = O_p(n^{-1/\eta})$ whenever $\eta \in [1,2)$ with the only condition imposed on F near zero also being $F(0) = 1 - F(0) = \frac{1}{2}$; this result is an implication of their Theorem 4.2, and it subsumes parts (c) and (d) of Proposition 4, but part (c) is a substantially stronger result because of the more rapid convergence rate given there, obtained at the expense of imposing A1. And, when $\eta \in (1,2)$, their Theorem 4.1 implies that $Z_n(\delta) \rightarrow_d Z(\delta)$, where $Z(\delta)$ has a nondegenerate limiting distribution whenever $E[|u_i|^{-\xi}] < \infty$ for some $\xi > \eta - 1$. This moment condition reduces in the present context to the requirement $1 + \gamma > \eta$, which is a strict version of the inequality encountered earlier. Under an additional condition that amounts to $\gamma < 1$, Davis et al. show that $n^{1/\eta} d_n$ converges in distribution to the unique minimizer of $Z(\delta)$. Therefore, stronger results are available for some of the parameter configurations allowed for in part (d). Their results (Theorem 4.1 and its sequel) also deliver an exact $O_p(n^{1/\eta})$ convergence rate accompanied by a limiting distribution when $\eta < 1$ and $\gamma < 1/\eta$, with the latter condition ensuring that $Z(\delta)$ has a unique minimizer. This represents a refinement of part (e) of Proposition 4 obtained at the cost of imposing a condition on γ. Pollard (1991) has obtained an $O_p(n)$ convergence rate for $\eta = 1$ (and $\gamma = 1$), a case not covered by Proposition 4.

Some comparisons with the OLS estimator seem worth making. The convergence rate of that estimator is determined entirely by the tails of F. Under

https://doi.org/10.1017/S0266466601174074 Published online by Cambridge University Press

A3 it is $O_p(n^{1/2})$ when $\eta > 2$ (or $\alpha = 2$ under A2). So, under the conditions of part (a) of Proposition 4, the relationship between the two estimators has the same flavor as in the preceding sections: the convergence rate of the OLS estimator exceeds that of the LAD estimator if $\gamma > 1$, with the ranking reversed for $\gamma < 1$; the case $\gamma = 1$ is resolved by the behavior of L (in A1) near zero. The conditions of part (b) imply $\gamma > 1$, so the more rapid convergence rate of the OLS estimator relative to the LAD estimator is perhaps not surprising. When $\eta < 2$ in A3 (or $\alpha < 2$ in A2), the convergence rate of the OLS estimator is $O_p([n/\ln(n)]^{1/\eta})$. (The factor $\ln(n)$ arises here because of the presence of the products $u_i u_{i-1}$ in the formula for this estimator: in $\sum_{i=2}^{n} u_i u_{i-1} / \sum_{i=2}^{n} u_{i-1}^2$, the numerator and denominators have nondegenerate limiting distributions once they have been normalized by, respectively, $[n \ln(n)]^{1/\eta}$ and $n^{2/\eta}$. See Phillips, 1990, Appendix A.) But the relative performance of the two estimators in terms of their rates of convergence differs markedly from the case where the u_i have finite variance: in part (c) the LAD estimator has a faster convergence rate than the OLS estimator, as might be expected in view of Propositions 1–3 because the conditions of part (c) imply $\gamma < 1$. Under the conditions of part (d) the two estimators have similar convergence rates regardless of whether or not $\gamma > 1$, and the slight superiority of the LAD estimator under the subset of conditions mentioned earlier (i.e., $2 > 1 + \gamma > \eta > 1$ or $\eta < \min\{1, 1/\gamma\}$), and due to Davis et al. (1992), is noteworthy. The same slight advantage of the LAD estimator is present under the conditions of part (e) but now with the only restriction on F near zero that it has zero as a median.

6. 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 that seems intuitively attractive, in addition to being convenient from the point of view of derivations. The results for some models extend those of Smirnov (1952) for the location model in ways that provide reasonably straightforward parallels with these early findings, but others are rather more involved.

REFERENCES

Chan N.H. & L.T. Tran (1989) On the first order autoregressive process with infinite variance. *Econometric Theory* 5, 354–362.

Davis, R.A., K. Knight, & J. Liu (1992) *M*-estimation for autoregressions with infinite variances. *Stochastic Processes and Their Applications* 40, 145–180.

Feller, W.J. (1966) *An Introduction to Probability Theory and Its Applications*, vol II. New York: Wiley.

Geyer, C.J. (1996) On the Asymptotics of Convex Stochastic Optimisation. Manuscript, University of Minnesota.

https://doi.org/10.1017/S0266466601174074 Published online by Cambridge University Press

Gnedenko, B.V. (1989) *The Theory of Probability and the Elements of Statistics*, 4th ed. New York: Chelsea.
Herce, M.A. (1996) Asymptotic theory of LAD estimation in a unit root process with finite variance errors. *Econometric Theory* 12, 129–153.
Ibragimov, I.A. & Y.V. Linnik (1971) *Independent and Stationary Sequences of Random Variables*. Groningen: Wolters-Noordorf.
Knight, K. (1989) Limit theory for autoregressive parameters in an infinite variance random walk. *Canadian Journal of Statistics* 17, 261–278.
Knight, K. (1991) Limit theory for *M*-estimates in an integrated infinite variance process. *Econometric Theory* 7, 200–212.
Knight, K. (1993) Estimation in dynamic linear regression models with infinite variance errors. *Econometric Theory* 9, 570–588.
Knight, K. (1998a) Limiting distributions for L_1 regression estimators under general conditions. *Annals of Statistics* 26, 755–770.
Knight, K. (1998b) Some limit theory for L_1-estimators in autoregressive models under general conditions. In Y. Dodge (ed.), L_1 *Statistical Procedures and Related Topics*. IMS Monograph Series.
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 deviations estimators. *Econometric Theory* 7, 186–199.
Seneta, E. (1976) *Regularly Varying Functions*. New York: Springer-Verlag.
Serfling, R.J. (1980) *Approximation Theorems of Mathematical Statistics*. New York: Wiley.
Smirnov, N.V. (1952) *Limit Distributions for the Terms of a Variational Series*. American Mathematical Society Translation 67.

APPENDIX A: PROOFS OF LEMMAS 1 AND 2

Proof of Lemma 1. The proof proceeds by showing that, for $z \geq 0$,

(a) (i) implies (ii) with $\psi(z) = z^{\gamma}$, $(\gamma > 0)$;
(b) (ii) implies G is regularly varying at zero (from above) with exponent $\gamma > 0$;
(c) (i) implies (iii) with $\Psi(z) = z^{1+\gamma}/(1+\gamma)$ and $\rho_{n+1}/\rho_n \to 1$ as $n \to \infty$;
(d) (iii) implies that if we write $\Psi_n(z) = \int_0^z n^q G(s/\rho_n)ds = n^q \rho_n \int_0^{z/\rho_n} G(s)ds$, or $\Psi_n(z) = n^q \rho_n R(z/\rho_n)$ where $R(t) \equiv \int_0^t G(s)ds$, then R is regularly varying at zero (from above) with exponent equal to $1+\gamma$ and G is regularly varying at zero (from above) with exponent equal to γ, with $\gamma > 0$.

(a) We note that (i) means

$$n^q G(z/\rho_n) = n^q L(z/\rho_n)(z/\rho_n)^{\gamma}$$
$$= (n^q/\rho_n^{\gamma}) L(1/\rho_n)\{L(z/\rho_n)/L(1/\rho_n)\} z^{\gamma}.$$

Now, for fixed $q > 0$, choose $\{\rho_n\}$ so that $\rho_n^{\gamma} \sim n^q L(1/\rho_n)$ as $n \to \infty$; then $\rho_n \to \infty$ as $n \to \infty$, from $\gamma > 0$, $q > 0$ and the slowly varying property of L; and this last property also yields $L(z/\rho_n)/L(1/\rho_n) \to 1$ as $n \to \infty$, for each $z \in (0,\infty)$. And because $G(0) = 0$

https://doi.org/10.1017/S0266466601174074 Published online by Cambridge University Press

(because $G(z) = L(z)z^{\gamma} \searrow 0$ as $z \searrow 0$ by the slow variation of L at zero), we can conclude that $n^q G(z/\rho_n) \to z^{\gamma}$, as $n \to \infty$, for each $z \in [0,\infty)$, as required.

(b) Because, in (ii), z^o is a continuity point of ψ with $\psi(z^o) \in (0,\infty)$, (ii) implies that, as $n \to \infty$, $G(z/\rho_n)/G(z^o/\rho_n) \to \psi(z)/\psi(z^o) \equiv \phi(z) \in (0,\infty)$ for all z sufficiently close to z^o. Now define $\rho_n^o = \rho_n/z^o$ so $G(z/(z^o\rho_n^o))/G(1/\rho_n^o) \to \phi(z) \in (0,\infty)$ for all z sufficiently close to z^o. We can then apply a straightforward modification of Theorem 1.9 of Seneta (1976) to see that G is regularly varying at zero from above with non-negative exponent if it is also the case that

$$G(1/\rho_n^o)/G(1/\rho_{n+1}^o) \to 1 \quad \text{as } n \to \infty;$$

but this is so because $G(1/\rho_n^o)/G(1/\rho_{n+1}^o) = \{n^q G(z^o/\rho_n)\}/\{n^q G(z^o/\rho_{n+1})\}$ has, from (ii), the same limit (as $n \to \infty$) as $\{n^q G(z^o/\rho_n)\}/\{(n+1)^q G(z^o/\rho_{n+1})\}$, namely, $\psi(z^o)/\psi(z^o) = 1$. Finally, to check that the exponent of regular variation is positive, note that if not, G must, because it is nondecreasing, be slowly varying, i.e., $G(z^o/\rho_n)/G(\lambda z^o/\rho_n) \to 1$ as $n \to \infty$ for all $\lambda > 0$, which implies that $\psi(\lambda z^o) = \psi(z^o)$ for all such λ, so that ψ cannot be continuous at zero with $\psi(0) = 0$, which contradicts (ii).

(c) From (i), $\Psi_n(z) = \int_0^z n^q G(s/\rho_n)ds = \int_0^z n^q L(s/\rho_n)(s/\rho_n)^{\gamma}ds$, and choosing $\rho_n^{\gamma} \sim n^q L(1/\rho_n)$ as $n \to \infty$, as in (a), we have $\Psi_n(z) \sim \int_0^z [L(s/\rho_n)/L(1/\rho_n)]s^{\gamma}ds$ as $n \to \infty$. Now for each $z > 0$,

$$\int_0^z L(s/\rho_n)s^{\gamma}ds \sim L(1/\rho_n)\int_0^z s^{\gamma}ds \quad \text{as } n \to \infty$$

using a modification of Seneta's (1976) Theorem 2.6 to deal with regular variation at zero rather than at ∞. (The condition $\gamma > 0$ is used here.) So

$$\Psi_n(z) \sim \int_0^z [L(s/\rho_n)/L(1/\rho_n)]s^{\gamma}ds \sim z^{\gamma+1}/(1+\gamma) \quad \text{as } n \to \infty.$$

Note that $\rho_n/\rho_{n+1} \to 1$ as $n \to \infty$ because the definition of $\{\rho_n\}$ allows us to write $\rho_n \sim n^{q/\gamma}\ell(1/n)$, where ℓ is a function that varies slowly at zero; this also implies that $\rho_{\lfloor gn \rfloor}/\rho_n \to g^{q/\gamma}$ as $n \to \infty$ for any $g \in (0,\infty)$.

(d) The strict convexity and finiteness of the limit function Ψ on (z_1, z_2) ensure the existence of a continuity point $z^o \in (z_1, z_2)$ of Ψ for which $\Psi(z^o) \in (0,\infty)$. Also, from the definition of $\{\Psi_n\}$, $\Psi(0) = 0$, and Ψ is everywhere convex by the monotonicity of G. Now, $\Psi_{n+1}(z^o)/\Psi_n(z^o) = \int_0^{z^o}(n+1)^q G(s/\rho_{n+1})ds / \int_0^{z^o} n^q G(s/\rho_n)ds \to 1$ as $n \to \infty$ can be shown to imply $\rho_{n+1}/\rho_n \to 1$ as $n \to \infty$, by making use of the strict convexity of Ψ on a neighborhood of z^o, along with $\Psi(0) = 0$. Next, recalling that $R(z/\rho_n) = \Psi_n(z)/(n^q\rho_n)$, we can apply a similar argument to that used in (b) to see that R is regularly varying at zero (from above) with exponent no less than unity, noting that $\rho_{n+1}/\rho_n \to 1$ ensures that $R(1/\rho_n^o)/R(1/\rho_{n+1}^o) \to 1$ as $n \to \infty$, where $\rho_n^o = \rho_n/z^o$. To show that the exponent of regular variation exceeds one, suppose to the contrary that R is regularly varying at zero with exponent equal to one. Then $\{\rho_n n^q R(\lambda z/\rho_n)\}/\{\rho_n n^q R(z/\rho_n)\} \to \lambda$ as $n \to \infty$ for $\lambda \in (0,\infty)$, implying in particular that $\Psi(\lambda z)/\Psi(z) = \lambda$ for all $\lambda \neq 1$, which contradicts the strict convexity of Ψ on (z_1, z_2). Next, to see that G is regularly varying also, take $a, b > 0$, with $b > a$ and write, for $t > 0$,

https://doi.org/10.1017/S0266466601174074 Published online by Cambridge University Press

$$\frac{t(b-a)G(at)}{R(t)} \le \frac{\left(\int_{at}^{bt} G(s)ds\right)}{R(t)} \le \frac{t(b-a)G(bt)}{R(t)}$$

because G is nondecreasing. The integral in the middle is $[R(bt) - R(at)]/R(t)$, which has limit as $t \searrow 0$ equal to $b^\kappa - a^\kappa$ by the regularly varying property of R; here $\kappa > 1$ denotes the exponent of this regular variation. We see therefore that

$$\liminf_{t\searrow 0}\{tG(bt)/R(t)\} \ge (b^\kappa - a^\kappa)/(b-a),$$

which on letting $a \to b$ (for fixed b) becomes $\liminf_{t\searrow 0}\{tG(bt)/R(t)\} \ge \kappa b^{\kappa-1}$; similarly, $\limsup_{t\searrow 0}\{tG(at)/R(t)\} \le \kappa a^{\kappa-1}$. Here a, b are arbitrary, so we obtain

$$\lim_{t\searrow 0}\{tG(ct)/R(t)\} = \kappa c^{\kappa-1}$$

for each $c > 0$. (The idea here is the same aside from minor modifications as the proof of Theorem 2.4 in Seneta, 1976, but we note that his U is defined differently from R.) The regularly varying property of G follows immediately:

$$\frac{G(ct)}{G(t)} = \frac{tG(ct)}{R(t)} \times \frac{R(t)}{tG(t)} \to c^{\kappa-1} \quad \text{as } t \searrow 0.$$

Note that the γ in (i) is equal to $\kappa - 1 > 0$. ■

Proof of Lemma 2. In both (a) and (b) choose $\{\rho_n\}$ so that $n^q G(1/\rho_n) \sim 1$ as $n \to \infty$ and note that this is possible (with $\rho_n \to \infty$ as $n \to \infty$) because $G(z)$ is strictly increasing on $[0, z^*]$ and is continuous at zero with $G(0) = 0$.

(a) In $\Psi_n(z) = n^q \int_0^z G(s/\rho_n)ds$, $\int_0^z G(s/\rho_n)ds \sim G(1/\rho_n)\int_0^z ds$ as $n \to \infty$ because G is slowly varying at zero and nondecreasing (cf. Seneta, 1976, Theorem 2.6). So, from the definition of $\{\rho_n\}$, $\Psi_n(z) \to z$ as $n \to \infty$, as required. And for any $g > 1$, $\rho_{\lfloor gn \rfloor}/\rho_n \to \infty$ as $n \to \infty$ from the definition of $\{\rho_n\}$ and the slowly varying property of G.

(b) From the definition of $\{\rho_n\}$, $n^q G(z/\rho_n) \to 0$ as $n \to \infty$ for all $z \in [0,1)$ because, for $z \in (0,1)$, $\{n^q G(z/\rho_n)\}/\{n^q G(1/\rho_n)\} \to 0$ because $\rho_n \to \infty$ as $n \to \infty$ and G is rapidly varying at zero; similarly $n^q G(z/\rho_n) \to \infty$ as $n \to \infty$ for each $z > 1$. This establishes the stated property of the limit function ψ, and the properties of Ψ also follow easily by making use of the nondecreasing property of G. And, for any $g > 1$, $\rho_{\lfloor gn \rfloor}/\rho_n \to 1$ as $n \to \infty$ from the definition of $\{\rho_n\}$ and the rapidly varying property of G. ■

APPENDIX B: PROOFS OF PROPOSITIONS 1–4

Throughout $Z_n(\delta)$ is as defined by (1) and (2) and $A_n(\delta)$, $B_n(\delta)$ by (3) and (4), respectively. The sequences $\{\nu_n\}$, $\{\phi_n\}$ are defined subsequently for each case.

https://doi.org/10.1017/S0266466601174074 Published online by Cambridge University Press

Proof of Proposition 1. $B_n(\delta) = E[B_n(\delta)] + (B_n(\delta) - E[B_n(\delta)])$, and in the present case

$$E[B_n(\delta)] = 2(\phi_n/v_n)\sum_{i=1}^{n}\int_0^{i^\omega\delta/\phi_n}\{F(s) - F(0)\}ds$$

$$= 2(\phi_n/v_n)\sum_{i=1}^{n}\int_0^{i^\omega\delta/\phi_n} G(s)ds$$

$$= 2v_n^{-1}\sum_{i=1}^{n}\phi_n R(i^\omega\delta/\phi_n),$$

where $R(z) \equiv \int_0^z G(s)ds$ and $G(z) \equiv F(z) - F(0)$. Under A1 and the assumed symmetry of F, we have $|G(z)| = L(z)|z|^\gamma$, or $G(z) = \mathrm{sgn}(z)L(z)|z|^\gamma$, recalling that $L(-z) = L(z) > 0$ for all $z \neq 0$ and L is slowly varying at zero. Here R is regularly varying at zero from above and below with exponent $1 + \gamma > 1$, and in particular $R(z) = L^o(z)|z|^{1+\gamma}$, where L^o, like L, is an slowly varying even function and

$$L^o(z) \sim L(z)/(1+\gamma) \quad \text{as } z \to 0.$$

The connection between R and G can be obtained by using part (c) of the proof of Lemma 1 followed by part (d), although more direct methods are available. (For example, for $z > 0$ we can write $\int_0^z G(s)ds = \int_0^1 zG(zt)dt = z^{1+\gamma}\int_0^1 L(zt)t^\gamma dt$ and then modify Theorem 2.6 of Seneta (1976) to obtain $\int_0^1 L(zt)t^\gamma dt \sim L(z)/(1+\gamma)$ as $z \searrow 0$.) We will be able to exploit the regular variation of R at zero if $\{\phi_n\}$ is chosen so that $\max_{1\le i\le n}\{i^\omega/\phi_n\} \to 0$, as $n \to \infty$. We now consider parts (a) and (b) in turn.

(a) We set

$$v_n = \left(\sum_{i=1}^{n} i^{2\omega}\right)^{1/2},$$

$$\phi_n \sim \left[v_n^{-1}L^o(n^\omega/\phi_n)\sum_{i=1}^{n} i^{\omega(1+\gamma)}\right]^{1/\gamma} \quad \text{as } n \to \infty,$$

noting that $v_n = O(n^{\omega+1/2})$, $\omega > -\frac{1}{2}$, and we can write $\phi_n \sim \ell^o(1/n)n^{\omega+1/2\gamma}$ as $n \to \infty$, where ℓ^o varies slowly at zero. Note that $\phi_n \to \infty$ as $n \to \infty$, because $\omega + \frac{1}{2}\gamma > 0$, and also $n^\omega/\phi_n \to 0$ as $n \to \infty$. Then

$$E[B_n(\delta)] = 2|\delta|^{1+\gamma}v_n^{-1}\sum_{i=1}^{n}\phi_n^{-\gamma}L^o(i^\omega\delta/\phi_n)i^{\omega(1+\gamma)}$$

$$= 2|\delta|^{1+\gamma}\sum_{i=1}^{n}\{L^o(i^\omega\delta/\phi_n)/L^o(n^\omega/\phi_n)\}i^{\omega(1+\gamma)}\Big/\sum_{i=1}^{n} i^{\omega(1+\gamma)}.$$

Now, the slowly varying property of L^o ensures that $L^o(i^\omega\delta/\phi_n)/L^o(n^\omega/\phi_n) \to 1$ as $n \to \infty$ uniformly on $\{i : \lfloor \varepsilon n \rfloor \le i \le n\}$ for any $\varepsilon \in (0,1)$, and so we obtain

$$E[B_n(\delta)] \to 2|\delta|^{1+\gamma} \quad \text{as } n \to \infty$$

https://doi.org/10.1017/S0266466601174074 Published online by Cambridge University Press

provided we can show that

$$\sum_{i=1}^{\lfloor \varepsilon n \rfloor} \{L^o(i^\omega \delta/\phi_n)/L^o(n^\omega/\phi_n)\} i^{\omega(1+\gamma)} \Big/ \sum_{i=1}^{n} i^{\omega(1+\gamma)}$$

has a limit as $n \to \infty$ that can be made arbitrarily close to zero by making ε sufficiently small. We distinguish between the two cases $\omega \geq 0$ and $\omega < 0$. For $\omega \geq 0$, the slow variation of L^o at zero ensures that for $\kappa > 0$, $\sup_{0<t\leq z}\{L^o(t)t^\kappa\} \sim L^o(z)z^\kappa$ as $z \searrow 0$ (see, e.g., Seneta, 1976, p. 68). So for $\omega \geq 0$, we can deduce that for any $\varepsilon > 0$, as $n \to \infty$,

$$\sup_{1<i\leq \lfloor \varepsilon n \rfloor} \{L^o(i^\omega/\phi_n)(i^\omega/\phi_n)^{1+\gamma}\} \sim L^o(\lfloor \varepsilon n \rfloor^\omega/\phi_n)(\lfloor \varepsilon n \rfloor^\omega/\phi_n)^{1+\gamma},$$

from which we obtain the result we seek without difficulty by again making use of the slow variation of L^o. For $\omega < 0$, we first note that $w(1 + \gamma) > -1$ is implied by the conditions $w > -\frac{1}{2}$, $w > -\frac{1}{2}\gamma$, so that $\sum_{i=1}^{n} i^{\omega(1+\gamma)}$ diverges as $n \to \infty$. This leads to the result once it is combined with

$$\sup_{1\leq i\leq \lfloor \varepsilon n \rfloor} \{L^o(i^\omega/\phi_n)\} \sim L^o(\lfloor \varepsilon n \rfloor^\omega/\phi_n),$$

as $n \to \infty$, which holds because $\omega < 0$ and $L(z)$ is (by assumption) nonincreasing on $(-z^*, z^*)$ for sufficiently small $z^* > 0$, and hence so is L^o.

The next step is to dispose of the remainder, $B_n(\delta) - E[B_n(\delta)]$. To do this we can proceed as in Knight (1998a):

$$\begin{aligned}
\operatorname{var}[B_n(\delta)] &= \operatorname{var}\left[2(\phi_n/v_n)\sum_{i=1}^{n}\int_0^{i^\omega \delta/\phi_n}\{I(u_i \leq s) - I(u_i \leq 0)\}ds\right] \\
&\leq 2(\phi_n/v_n)\max_{1\leq i\leq n}\{|i^\omega \delta/\phi_n|\}E[B_n(\delta)] \\
&\to 0 \quad \text{as } n \to \infty
\end{aligned}$$

because $i^\omega/v_n \to 0$ as $n \to \infty$ for all i whenever $\omega > -\frac{1}{2}$. And, under this last restriction on ω, and from the definition of $\{v_n\}$, $A_n(\delta) \to_d -\delta\xi$, $\xi \backsim N(0,1)$, (see, e.g., Gnedenko, 1989, p. 317). These results for $A_n(\delta)$ and $B_n(\delta)$ mean that for each δ, $Z_n(\delta) \to_d Z(\delta)$ as $n \to \infty$ where $Z(\delta) = -\delta\xi + 2|\delta|^{1+\gamma}$, $\xi \backsim N(0,1)$. It is easy to show that the convergence in distribution of $Z_n(\delta)$ to $Z(\delta)$ holds for any finite collection of δ's (see, e.g., Geyer, 1996, Lemma 3.3). And, because $\gamma > 0$, $Z(\delta)$ is strictly convex in δ and consequently has a unique minimizer; the minimizer of $Z_n(\delta)$ has for its limiting distribution the distribution of the minimizer of the random function $Z(\delta)$ as in Geyer (1996); and the result follows because this minimizer satisfies the first order condition $-\xi + 2(1 + \gamma)\operatorname{sgn}(\delta)|\delta|^\gamma = 0$.

(b) For the case where $\omega < -\frac{1}{2}$ and $\omega(1 + \gamma) > -1$ we set $v_n = 1$ for all n and define $\{\phi_n\}$ as in (a). This now means that we can write $\phi_n = \ell^o(1/n)n^{[1+\omega(1+\gamma)]/\gamma}$ where ℓ^o varies slowly at zero. And, $|A_n(\delta)/\delta| = O_p(1)$ for each $\delta \neq 0$, whereas the same argument as that used for part (a) (for the case $\omega < 0$) reveals that $E[B_n(\delta)] \to 2|\delta|^{1+\gamma}$ as $n \to \infty$. Note here that $i^\omega/\phi_n \leq 1/\phi_n$ for all i and that $\phi_n \to \infty$ as $n \to \infty$. And, for any $\varepsilon > 0$,

https://doi.org/10.1017/S0266466601174074 Published online by Cambridge University Press

$$\Pr[|B_n(\delta) - E[B_n(\delta)]| \geq \varepsilon E[B_n(\delta)]] \leq \operatorname{var}[B_n(\delta)]/\{\varepsilon E[B_n(\delta)]\}^2$$
$$\leq 2 \max_{1\leq i\leq n} \{|i^{\omega}\delta|\}/\{\varepsilon^2 E[B_n(\delta)]\}$$
$$\sim 2|\delta|/(\varepsilon^2 2|\delta|^{1+\gamma}) \quad \text{as } n \to \infty$$
$$= 1/(\varepsilon^2|\delta|^{\gamma})$$

using $\omega < 0$ and $v_n = 1$, along with $\operatorname{var}[B_n(\delta)] = \sum_{i=1}^n \operatorname{var}[b_{i,n}(\delta)] \leq \sum_{i=1}^n \max\{b_{i,n}(\delta)\}E[b_{i,n}(\delta)]$. Hence for fixed $\varepsilon \in (0,1)$,

$$\lim_{n\to\infty} \Pr[B_n(\delta) \leq (1-\varepsilon)2|\delta|^{1+\gamma}] \leq 1/(\varepsilon^2|\delta|^{\gamma}),$$

and because $\lim_{n\to\infty} \Pr[|A_n| \leq |\delta|M] \to 1$ as $M \to \infty$, it follows from $\gamma > 0$ that $\lim_{n\to\infty} \Pr[Z_n(\delta) > 0]$ is close to one for sufficiently large $|\delta|$. Therefore, because, for each n, $Z_n(\delta)$ is convex with $Z_n(0) = 0$ for all n, the minimizer of $Z_n(\delta)$ is contained in $[-d, d]$ with probability close to one for all large n, provided that d is sufficiently large. So, $\phi_n d_n = O_p(1)$ as $n \to \infty$, as claimed. ■

Proof of Proposition 2. We set $v_n = n^{1/2}a_n$, where $a_n = n^{1/\alpha}$, with $\alpha = 2$ for part (a), $\alpha \in (0,2)$ for part (b), and $\phi_n = a_n\rho_n$ where $\{\rho_n\}$ is defined subsequently in each case.

(a) Here $a_n = n^{1/2}$ and

$$A_n(\delta) = -\delta\sigma n^{-1/2} \sum_{i=2}^n [y_{i-1}/(\sigma a_n)]\operatorname{sgn}(u_i) \to_d -\sigma\delta \int_0^1 W_1(s)dW_2(s),$$

where W_1, W_2 are dependent Wiener processes, exactly as in Herce (1996). To consider

$$B_n(\delta) = 2(\rho_n/n^{1/2}) \sum_{i=2}^n \int_0^{\delta y_{i-1}/(a_n\rho_n)} [I(u_i \leq s) - I(u_i < 0)]ds$$

we first define

$$b_{i,n}(\delta) = 2(\rho_n/n^{1/2})I(1/m \leq |\delta y_{i-1}/a_n| \leq m) \int_0^{\delta y_{i-1}/(a_n\rho_n)} [I(u_i \leq s) - I(u_i < 0)]ds,$$

where we have introduced truncations of the $|\delta y_{i-1}/a_n|$ below at $1/m$ and above at $m > 1$. We see subsequently that these are inconsequential asymptotically (as $n \to \infty$) for sufficiently large m. Next let

$$\mu_{i,n}(\delta) = E[b_{i,n}(\delta)|y_{i-1}]$$

so

$$\sum_{i=2}^n \mu_{i,n}(\delta) = 2n^{-1} \sum_{i=2}^n I(1/m \leq |\delta y_{i-1}/a_n| \leq m) \int_0^{\delta y_{i-1}/(a_n\rho_n)} n^{1/2}\rho_n G(s)ds,$$

where $G(z) = F(z) - F(0)$. As in the proof of Proposition 1, we write

$$\int_0^{\delta y_{i-1}/(a_n\rho_n)} n^{1/2}\rho_n G(s)ds = n^{1/2}\rho_n R(\delta y_{i-1}/(a_n\rho_n)),$$

https://doi.org/10.1017/S0266466601174074 Published online by Cambridge University Press

where $R(z) = L^o(z)|z|^{1+\gamma}$ is an even function, regularly varying at zero with exponent $1 + \gamma$; L^o is slowly varying at 0. Next choose $\{\rho_n\}$ to satisfy

$$n^{1/2}\rho_n^{-\gamma} \sim 1/L^o(1/\rho_n) \quad \text{as } n \to \infty$$

so that $\rho_n \to \infty$ as $n \to \infty$. Now, as $n \to \infty$,

$$[R(\delta y_{i-1}/(a_n\rho_n)) - L^o(1/\rho_n)|\delta y_{i-1}/(a_n\rho_n)|^{1+\gamma}]I(1/m \le |\delta y_{i-1}/a_n| \le m) \to 0$$

uniformly for all i because $L^o(\lambda z)/L^o(z) \to 1$ as $z \to 0$ uniformly for all $\lambda \in [1/m, m]$ by the slowly varying property of L^o. So, for the preceding choice of $\{\rho_n\}$, as $n \to \infty$,

$$\sum_{i=2}^{n} \mu_{i,n}(\delta) - 2n^{-1}\sum_{i=2}^{n} |\delta y_{i-1}/a_n|^{1+\gamma} I(1/m \le |\delta y_{i-1}/a_n| \le m) \to_p 0.$$

Next note that for $r \in [0,1]$, $y_{\lfloor rn \rfloor}/(\sigma a_n) \Rightarrow W_1(r)$ in $D[0,1]$ (the space of real valued right continuous functions on $[0,1]$ with left hand limits), where $W_1(r)$ is the Wiener process arising in the limiting distribution of $A_n(\delta)$. (Here $\Rightarrow$ denotes weak convergence as $n \to \infty$ of a sequence of probability measures.) So $|\delta y_{\lfloor rn \rfloor}/(\sigma a_n)|^{1+\gamma} \Rightarrow |\delta W_1(r)|^{1+\gamma}$ by the continuity of $|\cdot|^{1+\gamma}$, whereas, for each δ,

$$|y_{\lfloor rn \rfloor}/a_n|^{1+\gamma} I(1/m \le |\delta y_{\lfloor rn \rfloor}/a_n| \le m) \Rightarrow |\sigma\delta W_1(r)|^{1+\gamma} I(1/m \le |\sigma\delta W_1(r)| \le m)$$

because the set of discontinuity points of this last function of $W_1(r)$ has zero Wiener measure. And, by the continuous mapping theorem, for each δ, as $n \to \infty$,

$$\sum_{i=2}^{n} \mu_{i,n}(\delta) \to_d \int_0^1 |\sigma\delta W_1(r)|^{1+\gamma} I(1/m \le |\sigma\delta W_1(r)| \le m)\, dr.$$

We next want to show that $\sum_{i=2}^{n}[b_{i,n}(\delta) - \mu_{i,n}(\delta)] \to_p 0$ as $n \to \infty$. But the summands here are martingale differences, so we need only demonstrate (as in Knight, 1989) that $\sum_{i=2}^{n} E[b_{i,n}(\delta)^2 | u_1, \ldots, u_{i-1}] \to_p 0$, which is so because, by the definition of the $\mu_{i,n}(\delta)$,

$$\sum_{i=2}^{n} E[b_{i,n}(\delta)^2 | u_1, \ldots, u_{i-1}] \le \max_{2 \le i \le n} \{b_{i,n}(\delta)\} \times \sum_{i=2}^{n} \mu_{i,n}(\delta)$$

$$= o_p(1) \quad \text{as } n \to \infty,$$

because $\sum_{i=2}^{n} \mu_{i,n}(\delta) = O_p(1)$ (as we have just seen), and, uniformly for all i,

$$0 \le b_{i,n}(\delta) \le 2(\rho_n/n^{1/2})(|\delta y_{i-1}|/(a_n\rho_n))I(1/m \le |\delta y_{i-1}/a_n| \le m) \le 2m/n^{1/2}$$

from the definition of the $b_{i,n}(\delta)$. Hence, for each m, $\sum_{i=2}^{n} b_{i,n}(\delta) - \sum_{i=2}^{n} \mu_{i,n}(\delta) \to_p 0$ as $n \to \infty$.

The final step involves showing that the two truncations used previously do not matter asymptotically for large enough m. For the truncation at m, we again follow Knight (1989):

https://doi.org/10.1017/S0266466601174074 Published online by Cambridge University Press

$$\Pr\left[2(\rho_n/n^{1/2})\sum_{i=2}^{n} I(|\delta y_{i-1}/a_n| > m)\int_0^{\delta y_{i-1}/(a_n\rho_n)} [I(u_i \le s) - I(u_i < 0)]ds > \varepsilon\right]$$

$$\le \Pr\left[\max_{2\le i\le n}\{|\delta y_{i-1}/a_n|\} > m\right],$$

which has a limit as $n \to \infty$, with this limit tending to zero as $m \to \infty$; the limit in question is $\Pr[\max_{r\in[0,1]}\{|\sigma\delta W_1(r)|\} > m]$. For the truncation at $1/m$, we use a different argument. Note that

$$2(\rho_n/n^{1/2})\sum_{i=2}^{n} I(|\delta y_{i-1}/a_n| < 1/m)\int_0^{\delta y_{i-1}/(a_n\rho_n)} [I(u_i \le s) - I(u_i < 0)]ds$$

is no greater than

$$\sum_{i=2}^{n} 2(\rho_n/n^{1/2})(m\rho_n)^{-1} I(|u_i| \le 1/(m\rho_n)),$$

and recalling that $G(z) = \mathrm{sgn}(z)L(z)|z|^{\gamma}$, the expectation of this last sum is, by the assumed symmetry of G at zero,

$$2n^{1/2}(1/m)\Pr[|u_i| \le 1/(m\rho_n)] = 4n^{1/2}G(1/(m\rho_n))/m = 4n^{1/2}L(1/(m\rho_n))|m\rho_n|^{-\gamma}/m,$$

and, from the definition of $\{\rho_n\}$, we have $n^{1/2}\rho_n^{-\gamma}L(1/(m\rho_n)) \sim L(1/(m\rho_n))/L^o(1/\rho_n)$, as $n \to \infty$, and $L(1/(m\rho_n)) \sim (1+\gamma)L^o(1/(m\rho_n))$ as $n \to \infty$ for each $m > 0$, using the connection between G and R noted in the proof of Proposition 1. Also, by the slow variation of L^o at zero, $L^o(1/m\rho_n) \sim L^o(1/\rho_n)$ as $n \to \infty$, for each m. Therefore

$$n^{1/2}L(1/(m\rho_n))|m\rho_n|^{-\gamma}/m \sim (1+\gamma)/m^{1+\gamma} \quad \text{as } n \to \infty,$$

so for all large n the expectation in question is close to zero when m is large. Hence, for any $\varepsilon > 0$, the limit as $n \to \infty$ of

$$\Pr\left[2(\rho_n/n^{1/2})\sum_{i=2}^{n} I(|\delta y_{i-1}/a_n| < 1/m)\int_0^{\delta y_{i-1}/(a_n\rho_n)} [I(u_i \le s) - I(u_i < 0)]ds > \varepsilon\right]$$

is close to zero for large m. Putting the arguments for the two truncations together, we can therefore deduce that, for any $\varepsilon > 0$,

$$\lim_{n\to\infty} \Pr\left[\left|B_n(\delta) - \sum_{i=2}^{n} b_{i,n}(\delta)\right| > \varepsilon\right]$$

can be made as close to zero as desired by making m sufficiently large; from this and the argument given previously it follows that the same is true of

$$\lim_{n\to\infty} \Pr\left[\left|B_n(\delta) - \sum_{i=2}^{n} \mu_{i,n}(\delta)\right| > \varepsilon\right].$$

https://doi.org/10.1017/S0266466601174074 Published online by Cambridge University Press

Next note that, for each δ,

$$\int_0^1 |\delta W_1(r)|^{1+\gamma} I(|\delta W_1(r)| < 1/m) dr \leq 1/m^{1+\gamma}$$

and, for any $\varepsilon > 0$,

$$\Pr\left[\int_0^1 |\delta W_1(r)|^{1+\gamma} I(|\delta W_1(r)| > m) dr > \varepsilon\right] \leq \Pr\left[\max_{r\in[0,1]} \{|\delta W_1(r)|\} > m\right]$$
$$\to 0 \quad \text{as } m \to \infty.$$

Hence $B_n(\delta) \to_d 2\int_0^1 |\sigma\delta W_1(r)|^{1+\gamma} dr$, and combining this result with that given earlier for $A_n(\delta)$ reveals that for each δ,

$$\{A_n(\delta), B_n(\delta)\} \to_d \left\{-\sigma\delta \int_0^1 W_1(s) dW_2(s),\ 2|\delta|^{1+\gamma} \int_0^1 |\sigma W_1(r)|^{1+\gamma} dr\right\}$$

as $n \to \infty$, so $Z_n(\delta) \to_d Z(\delta)$ as $n \to \infty$ for each δ, where

$$Z(\delta) \equiv -\sigma\delta \int_0^1 W_1(s) dW_2(s) + 2|\delta|^{1+\gamma} \int_0^1 |\sigma W_1(r)|^{1+\gamma} dr.$$

Moreover, the convergence in distribution of $Z_n(\delta)$ to $Z(\delta)$ as $n \to \infty$ holds for any finite collection, $(\delta_1, \ldots, \delta_h)$, of δ's, i.e.,

$$\{Z_n(\delta_1), \ldots, Z_n(\delta_h)\} \to_d \{Z(\delta_1), \ldots, Z(\delta_h)\}$$

as $n \to \infty$. Now, recalling that for each n, $Z_n(\delta)$ is a random convex function, we can appeal to Geyer (1996) to obtain the result in the statement of Proposition 2(a) by making use of the strict convexity of $Z(\delta)$ (a consequence of $\gamma > 0$), which ensures that $Z(\delta)$ has a unique minimizer with probability one. The convergence rate of d_n is given by the divisor of δ in $Z_n(\delta)$, i.e., $\phi_n = a_n\rho_n$, for $a_n = n^{1/2}$, and from the definition of $\{\rho_n\}$ we can write $\rho_n = n^{1/2\gamma}\ell^o(1/n)$ where ℓ^o is slowly varying at zero.

(b) In this case $a_n = n^{1/\alpha}$ and

$$A_n(\delta) = -\delta n^{-1/2} \sum_{i=2}^{n} (y_{i-1}/a_n)\mathrm{sgn}(u_i) \to_d -\delta \int_0^1 S_\alpha(s) dW(s),$$

where S_α is a stable process with index $\alpha \in (0,2)$ and W is a Wiener process that is independent of S_α, exactly as in Knight (1989). Now we can proceed as in part (a) to obtain $B_n(\delta) \to_d \int_0^1 |\delta S_\alpha(r)|^{1+\gamma} dr$ where S_α is the stable process arising above in the limiting distribution of $A_n(\delta)$ and $\{\rho_n\}$ again satisfies $n^{1/2}\rho_n^{-\gamma} \sim 1/L^o(1/\rho_n)$ as $n \to \infty$, so $\rho_n = n^{1/2\gamma}\ell^o(1/n)$, where ℓ^o is slowly varying at zero; the convergence rate of d_n is given by $a_n\rho_n$, $a_n = n^{1/\alpha}$. ■

Proof of Proposition 3. For parts (a) and (b) we set $v_n = n^{1/2}a_n$, where $a_n = n^{1/\alpha}$ (with $\alpha = 2$ in part (a), $\alpha \in (0,2)$ in part (b)) and $\phi_n = a_n\rho_n$, with $\{\rho_n\}$ chosen to satisfy $n^{1/2}G(1/\rho_n) \sim 1$ as $n \to \infty$.

https://doi.org/10.1017/S0266466601174074 Published online by Cambridge University Press

(a) $A_n(\delta)$ behaves exactly as in the proof of Proposition 2(a):

$$A_n(\delta) \to_d -\sigma\delta \int_0^1 W_1(s)dW_2(s),$$

where W_1, W_2 are dependent Wiener processes. To examine the limiting behavior of $B_n(\delta)$ we let

$$b_{i,n}(\delta) = 2(\rho_n/n^{1/2})I(|\delta y_{i-1}/a_n| \le m) \int_0^{\delta y_{i-1}/(a_n\rho_n)} [I(u_i \le s) - I(u_i < 0)]ds$$

for $m > 1$, where the truncations of the $|\delta y_{i-1}/a_n|$ below at $1/m$ used in the proof of Proposition 2(a) are omitted. Letting $\mu_{i,n}(\delta) = E[b_{i,n}(\delta)|y_{i-1}]$,

$$\sum_{i=2}^n \mu_{i,n}(\delta) = 2n^{-1} \sum_{i=2}^n I(|\delta y_{i-1}/a_n| \le m) \int_0^{\delta y_{i-1}/a_n} n^{1/2}G(s/\rho_n)ds,$$

where $G(z) = F(z) - F(0)$. Now from the definition of $\{\rho_n\}$, and Lemma 2(b) and its proof, we know that $\int_0^z n^{1/2}G(s/\rho_n)ds \to 0$ for $|z| \le 1$ as $n \to \infty$. Therefore, for any $m > 1$ and large n, $\sum_{i=2}^n \mu_{i,n}(\delta)$ is close to zero with probability close to one when $|\delta y_{i-1}/a_n| \le 1$ for all i, i.e., when $|\delta| \le 1/V_n$ where

$$V_n \equiv \max_{2\le i\le n} \{|y_{i-1}/a_n|\}.$$

We next want to show that $\sum_{i=2}^n \mu_{i,n}(\delta)$ is large with probability close to one when n is large and this condition on δ is not satisfied, i.e., when $|\delta| \ge (1+\varepsilon)/V_n$ for any $\varepsilon > 0$. We note first that

$$\sum_{i=2}^n \mu_{i,n}(\delta) \ge 2n^{-1} \sum_{i=2}^n I(|\delta|V_n - \varepsilon < |\delta y_{i-1}/a_n| \le m) \int_0^{\delta y_{i-1}/a_n} n^{1/2}G(s/\rho_n)ds$$

and because, from the definition of $\{\rho_n\}$ and Lemma 2(b), $\int_0^z n^{1/2}G(s/\rho_n)ds \to \infty$ as $n \to \infty$, whenever $|z| > 1$, it will be enough to show that

$$\lim_{n\to\infty} \Pr\left[n^{-1} \sum_{i=2}^n I(|\delta|V_n - \varepsilon < |\delta y_{i-1}/a_n| \le m) \ge \varepsilon^o \right]$$

tends to one as $m \to \infty$ for $\varepsilon^o > 0$ sufficiently small. Recalling that W_1 is the standard Wiener process arising above in the limiting distribution of $A_n(\delta)$, and letting $U_1 \equiv \max_{r\in[0,1]}\{|W_1(r)|\}$,

$$\{|y_{\lfloor rn\rfloor}/(\sigma a_n)|, V_n/\sigma\} \Rightarrow \{|W_1(r)|, U_1\}$$

as $n \to \infty$, the limiting probability we seek is given by

$$\Pr\left[\int_0^1 I(|\delta|\sigma U_1 - \varepsilon < |\sigma\delta W_1(r)| \le m)dr \ge \varepsilon^o \right],$$

which for each $\delta \ne 0$, and fixed $\varepsilon > 0$, has limit equal to one, as $m \to \infty$, for sufficiently small $\varepsilon^o > 0$, as required.

https://doi.org/10.1017/S0266466601174074 Published online by Cambridge University Press

The arguments used in the proof of Proposition 2(a) can be now be applied with little modification to see that $B_n(\delta) - \sum_{i=2}^{n} \mu_{i,n}(\delta)$ is close to zero with high probability for $|\delta| \leq 1/V_n$, when n is large, in which case $Z_n(\delta) - A_n(\delta)$ is also close to zero; whereas if $|\delta| > (1+\varepsilon)/V_n$ (for any $\varepsilon > 0$) and n is sufficiently large, an argument used in the proof of Proposition 1(b) can be adapted to show that $B_n(\delta) > M$ with high probability for arbitrarily large M. So, for each δ, $Z_n(\delta) \to_d Z(\delta)$, where

$$Z(\delta) = \begin{cases} -\sigma\delta \int_0^1 W_1(s)dW_2(s) & \text{if } |\delta| \leq 1/\max_{r\in[0,1]}\{\sigma|W_1(r)|\} \\ \infty & \text{otherwise} \end{cases},$$

and the convergence of the finite dimensional distributions also holds, as in the proof of Proposition 2(a). Because $Z(\delta)$ has a unique minimizer with probability one, we can appeal to Geyer (1996) to obtain the result, by making use of the preceding definition of $\{\rho_n\}$ and the fact that, from Lemma 2(b) $\rho_n/\rho_{\lfloor ng \rfloor} \to 1$ as $n \to \infty$ for any $g > 1$. The convergence rate of the estimator is given by $\phi_n = a_n\rho_n = n^{1/2}\rho_n$.

(b) This modifies the proof of Proposition 2(b) in essentially the same way as the proof of part (a) modifies the proof of Proposition 2(a). Note that attention needs to be paid to the sample path properties of the stable process S_α that occurs here in place of the Wiener process W_1 of part (a). (For some illustrations, see Phillips, 1990). The convergence rate of the estimator is given by $\phi_n = a_n\rho_n = n^{1/\alpha}\rho_n$.

(c) In the definition of $Z_n(\delta)$ we set $v_n = a_n n^{1/v}$, for $v < 2$, and $a_n = n^{1/\alpha}$, $\phi_n = a_n\rho_n$ with $\{\rho_n\}$ being defined subsequently. Then $A_n(\delta) \to_p 0$ as $n \to \infty$. We let

$$b_{i,n}(\delta) = 2(\rho_n/n^{1/v})I(1/m \leq |\delta y_{i-1}/a_n| \leq m)\int_0^{\delta y_{i-1}/(a_n\rho_n)} [I(u_i \leq s) - I(u_i < 0)]ds,$$

and if $\mu_{i,n}(\delta) = E[b_{i,n}(\delta)|y_{i-1}]$, we can write

$$\sum_{i=2}^{n} \mu_{i,n}(\delta) = 2n^{-1}\sum_{i=2}^{n} I(1/m \leq |\delta y_{i-1}/a_n| \leq m)\int_0^{\delta y_{i-1}/(a_n\rho_n)} n^{1-1/v}\rho_n G(s)ds,$$

where $G(z) = F(z) - F(0)$. Setting $R(z) = \int_0^z G(s)ds$, R is readily seen to be regularly varying at zero with exponent equal to one. (For example, we can write $\int_0^z G(s)ds = \int_0^1 zG(zt)dt$ as in the proof of Proposition 1(a) and then use the monotonicity of G to obtain, for arbitrarily small $\varepsilon > 0$, $zG(\varepsilon z)\int_\varepsilon^1 dt \leq \int_0^1 zG(zt)dt \leq zG(z)\int_0^1 dt$, which yields $\int_0^z G(s)ds \sim zG(z)$ as $z \to 0$ from the slowly varying property of G.) Specifically, $R(z) \sim G(z)z$ as $z \to 0$, or $R(z) = L^o(z)|z|$ where $\mathrm{sgn}(z)L^o(z)/G(z) \to 1$ as $z \to 0$, and L^o is slowly varying at zero. So, we write

$$\int_0^{\delta y_{i-1}/(a_n\rho_n)} n^{1-1/v}\rho_n G(s)ds = n^{1-1/v}\rho_n R(\delta y_{i-1}/(a_n\rho_n))$$

$$= n^{1-1/v}L^o(\delta y_{i-1}/(a_n\rho_n))|\delta y_{i-1}/a_n|,$$

and, if we choose $\{\rho_n\}$ to satisfy $n^{1-1/v}G(1/\rho_n) \sim 1$ as $n \to \infty$ (or $n^{1-1/v}L^o(1/\rho_n) \sim 1$ as $n \to \infty$), we can proceed along the lines of the proof of Proposition 2 to obtain $B_n(\delta) - 2n^{-1}\sum_{i=2}^{n}|\delta y_{i-1}/a_n| \to_p 0$. When $\alpha = 2$, $B_n(\delta) \to_d 2\int_0^1 |\sigma\delta W_1(r)|dr$ as $n \to \infty$, and it follows from the properties of a standard Wiener process that

https://doi.org/10.1017/S0266466601174074 Published online by Cambridge University Press

$\lim_{n\to\infty}\Pr[B_n(\delta) > \varepsilon]$ approaches one as $\varepsilon \searrow 0$ for any fixed $\delta \neq 0$. When $\alpha \in (0,2)$, $B_n(\delta) \to_d 2\int_0^1 |\delta S_\alpha(r)|\,dr$, and the same conclusion is obtained. Because $A_n(\delta) \to_p 0$, we can again appeal to Geyer (1996) to see that the minimizer of $Z_n(\delta)$ converges in probability to zero as $n \to \infty$, and the result follows from the fact that $v < 2$ can be made arbitrarily close to 2, the definition of $\{\rho_n\}$ just given, and the fact that the divisor of δ in the definition of $Z_n(\delta)$ is $\phi_n = a_n\rho_n = n^{1/\alpha}\rho_n$.

Proof of Proposition 4. Throughout we let $B_n(\delta) = \sum_{i=2}^n b_{i,n}(\delta)$ where, from (4),

$$b_{i,n}(\delta) = 2(\phi_n/v_n)\int_0^{\delta y_{i-1}/\phi_n} [I(u_i \le s) - I(u_i < 0)]ds,$$

noting that $0 \le b_{i,n}(\delta) \le 2|\delta y_{i-1}|/v_n$, and define

$$\mu_{i,n}(\delta) = E[b_{i,n}(\delta)|u_1,\ldots,u_{i-1}|] = 2n^{-1}\int_0^{\delta y_{i-1}} (n/v_n)[F(s/\phi_n) - F(0)]ds.$$

We let $G(z) = F(z) - F(0)$ and, for parts (a)–(d), $R(z) = \int_0^z G(s)ds = L^o(z)|z|^{1+\gamma}$ where L^o is an even function that is slowly varying at zero, as in the proofs of Propositions 1 and 2, and we also let

$$\mu^o_{i,n}(\delta) = t_{i,n}\mu_{i,n}(\delta), \qquad b^o_{i,n}(\delta) = t_{i,n}b_{i,n}(\delta),$$

where $t_{i,n}$ are indicator functions depending on $m > 1$ that are defined subsequently for each case.

(a) We set $v_n = n^{1/2}$. Then, using $\beta = 0$, $A_n(\delta) = -\delta n^{-1/2}\sum_{i=2}^n y_{i-1}\operatorname{sgn}(u_i) \to_d -\delta\sigma\xi$, where $\xi \backsim N(0,1)$ and $E[u_i^2] = \sigma^2 < \infty$. Next let $t_{i,n} = t_i = I(1/m \le |\delta y_{i-1}| \le m)$; now, $n/v_n = n^{1/2}$ and

$$\sum_{i=2}^n \mu^o_{i,n}(\delta) = 2n^{-1}\sum_{i=2}^n t_i\int_0^{\delta y_{i-1}} n^{1/2}G(s/\phi_n)ds$$

$$= 2n^{-1}\sum_{i=2}^n t_i\, n^{1/2}\phi_n R(\delta y_{i-1}/\phi_n).$$

We choose $\{\phi_n\}$ so that $n^{1/2}\phi_n^{-\gamma} \sim 1/L^o(1/\phi_n)$ as $n \to \infty$. Then from the definition of R and the slow variation of L^o, for each $\delta \neq 0$,

$$\sum_{i=2}^n \mu^o_{i,n}(\delta) - 2n^{-1}\sum_{i=2}^n |\delta y_{i-1}|^{1+\gamma}t_i \to 0 \quad \text{as } n \to \infty,$$

and so

$$\sum_{i=2}^n \mu^o_{i,n}(\delta) \to_p E[|\delta y_{i-1}|^{1+\gamma}t_i] \quad \text{as } n \to \infty.$$

Next, an argument used by Knight (1998b) applies to yield $\sum_{i=2}^n [b^o_{i,n}(\delta) - \mu^o_{i,n}(\delta)] \to_p 0$ as $n \to \infty$. By making m sufficiently large, and using the fact that $E[|\delta y_i|^{1+\gamma}t_i] \to E[|\delta u_i|^{1+\gamma}]$ as $m \to \infty$ (because $\beta = 0$), for each δ by the finiteness of $E[|u_i|^{1+\gamma}]$, we obtain $B_n(\delta) \to_p 2|\delta|^{1+\gamma}E[|u_i|^{1+\gamma}]$ as $n \to \infty$. Combining this with the limiting behavior of $A_n(\delta)$ leads to the result, as in the proof of Proposition 1(a); the

https://doi.org/10.1017/S0266466601174074 Published online by Cambridge University Press

rate of convergence here is given by ϕ_n, and we can write $\phi_n = n^{1/2\gamma}\ell^o(1/n)$ where ℓ^o varies slowly at zero. Note that in the present case, if A3 is satisfied for some $\eta > 0$, then $\eta > \max\{2, 1+\gamma\}$.

(b) Here $1+\gamma \geq \eta > 2$. Let $\upsilon_n = n^{1/2}$, so $A_n(\delta)$ behaves as in (a) and $n/\upsilon_n = n^{1/2}$. We set $t_{i,n} = I(s_n/m \leq |\delta y_{i-1}/\phi_n| \leq s_n m)$ where $\{s_n\}$ is a sequence for which $s_n \to 0$ and $s_n\phi_n \to \infty$ as $n \to \infty$; $\{\phi_n\}$ is defined subsequently. Then, for $\delta \neq 0$,

$$\sum_{i=1}^{n} \mu^o_{i,n}(\delta) = 2n^{-1}\sum_{i=1}^{n} n^{1/2}\phi_n^{-\gamma}L^o(\delta y_{i-1}/\phi_n)|\delta y_{i-1}|^{1+\gamma}t_{i,n}$$

$$\sim 2n^{-1}\sum_{i=1}^{n} n^{1/2}\phi_n^{-\gamma}L^o(s_n)|\delta y_{i-1}|^{1+\gamma}t_{i,n} \quad \text{as } n \to \infty$$

by the slowly varying property of L^o and because, for each δ, $|\delta y_{i-1}/\phi_n| \to 0$ as $n \to \infty$, uniformly for all i such that $s_n/m \leq |\delta y_{i-1}/\phi_n| \leq s_n m$. Now, A3 along with $1+\gamma \geq \eta$ ensures that $E[|\delta y_i|^{1+\gamma}t_{i,n}] \sim |\delta|^{1+\gamma}c(m)(s_n\phi_n)^{1+\gamma-\eta}$ as $n \to \infty$ where $c(m) \to \infty$ as $m \to \infty$. Then, as $n \to \infty$,

$$E\left[\sum_{i=1}^{n} \mu^o_{i,n}(\delta)\right] \sim 2n^{1/2}\phi_n^{-\gamma}L^o(s_n)|\delta|^{1+\gamma}c(m)(s_n\phi_n)^{1+\gamma-\eta}$$

$$= 2|\delta|^{1+\gamma}c(m)n^{1/2}\phi_n^{1-\eta}s_n^{1+\gamma-\eta}L^o(s_n),$$

and we choose $\{\phi_n\}$ so that $n^{1/2}\phi_n^{1-\eta}s_n^{1+\gamma-\eta}L^o(s_n) \sim 1$ as $n \to \infty$. Observe that for each (large) n and this choice of $\{\phi_n\}$, $E[\sum_{i=1}^n \mu^o_{i,n}(\delta)] \to \infty$ as $m \to \infty$ for each $\delta \neq 0$. And, as in the proof of Proposition 1(b), for $\varepsilon > 0$,

$$\Pr\left[\sum_{i=1}^{n} b^o_{i,n}(\delta) - E\left[\sum_{i=1}^{n} \mu^o_{i,n}(\delta)\right] > \varepsilon E\left[\sum_{i=1}^{n} \mu^o_{i,n}(\delta)\right]\right]$$

$$\leq \operatorname{var}\left[\sum_{i=1}^{n} b^o_{i,n}(\delta)\right] \Big/ \left(E\left[\varepsilon\sum_{i=1}^{n} \mu^o_{i,n}(\delta)\right]\right)^2$$

$$\leq \max_{1\leq i\leq n}\{b^o_{i,n}(\delta)\} \Big/ E\left[\varepsilon^2\sum_{i=1}^{n} \mu^o_{i,n}(\delta)\right]$$

$$\leq 2\max_{2\leq i\leq n}\{n^{-1/2}|\delta y_{i-1}|t_{i,n}\} \Big/ E\left[\varepsilon^2\sum_{i=1}^{n} \mu^o_{i,n}(\delta)\right],$$

and because $\max_i\{n^{-1/2}|y_{i-1}|\} = o_p(1)$ given $\eta > 2$, it follows that for fixed $\delta \neq 0$, and any $M > 0$, $\Pr[B_n(\delta) > M] \to 1$ as $n \to \infty$ for $\{\phi_n\}$ as defined previously, in view of $B_n(\delta) = \sum_{i=1}^n b_{i,n}(\delta) \geq \sum_{i=1}^n b^o_{i,n}(\delta)$ for any $m > 1$. The result follows from the convexity of $Z_n(\delta)$, the definition of $\{\phi_n\}$, and the fact that we can choose $\{s_n\}$ so that $s_n \to 0$ arbitrarily slowly as $n \to \infty$.

(c) Here $\eta \leq 2$ and $1+\gamma < \eta$. Let $\upsilon_n = n^{1/\upsilon}$, $1 < \upsilon < \eta$. Then $A_n(\delta) \to_p 0$ as $n \to \infty$ and $n/\upsilon_n = n^{1-1/\upsilon}$. Choose $t_{i,n} = t_i = I(1/m \leq |\delta y_{i-1}| \leq m)$, as in the proof of part (a). Now,

$$\sum_{i=2}^{n} \mu_{i,n}(\delta) = 2n^{-1}\sum_{i=2}^{n} n^{1-1/\upsilon}\phi_n R(|\delta y_{i-1}/\phi_n|),$$

https://doi.org/10.1017/S0266466601174074 Published online by Cambridge University Press

and we choose $\{\phi_n\}$ so that $n^{1-1/v}\phi_n^{-\gamma} \sim L^o(1/\phi_n)$ as $n \to \infty$, or $\phi_n = \ell^o(1/n)n^{(v-1)/(v\gamma)}$, where ℓ^o is slowly varying at zero. We can then use the slowly varying property of L^o to obtain, as in the proof of part (a),

$$\sum_{i=2}^{n} \mu_{i,n}^o(\delta) \to_p E[|\delta y_{i-1}|^{1+\gamma} t_i] \quad \text{as } n \to \infty,$$

whereas, $\sum_{i=2}^n [b_{i,n}^o(\delta) - \mu_{i,n}^o(\delta)] \to_p 0$ because

$$\begin{aligned}\sum_{i=2}^{n} E[b_{i,n}^o(\delta)^2 | u_1, \ldots, u_{i-1}] &\le \max_{1\le i\le n} \{b_{i,n}^o(\delta)\} \times \sum_{i=2}^{n} \mu_{i,n}^o(\delta) \\ &\le \max_{1\le i\le n} \{b_{i,n}(\delta)\} \times \sum_{i=2}^{n} \mu_{i,n}^o(\delta) \\ &\le \max_{2\le i\le n} \{|\delta y_{i-1}/v_n|\} \times \sum_{i=2}^{n} \mu_{i,n}^o(\delta) \\ &= o_p(1) \quad \text{as } n \to \infty,\end{aligned}$$

using the definition of the $b_{i,n}(\delta)$, $v_n = n^{-1/v}$, and the fact that $n^{-1/\eta}\max_i\{|y_{i-1}|\} = O_p(1)$ as $n \to \infty$, given $\eta \le 2$. By making m sufficiently large, we see therefore that $B_n(\delta) \to_p 2|\delta|^{1+\gamma}E[|u_i|^{1+\gamma}]$. Combining this with $A_n(\delta) \to_p 0$ yields $Z_n(\delta) \to_p 2|\delta|^{1+\gamma}E[|u_i|^{1+\gamma}]$, and so, by the convexity of $Z_n(\delta)$, $\phi_n d_n \to_p 0$ for $\{\phi_n\}$ as defined earlier, for any $v \in (1,\eta)$; the slow variation of ℓ^o ensures that we can assert that $n^\kappa d_n \to_p 0$ for any $\kappa < (\eta - 1)/\eta\gamma$.

(d) Here $1 < \eta \le 2$ and $1 + \gamma \ge \eta$. Let $v_n = n^{1/v}$, $1 < v < \eta$, so $A_n(\delta) \to_p 0$, as for part (c). We can use the same sort of argument as in part (b) to see that, for fixed $\delta \ne 0$ and any $M > 0$, $\Pr[B_n(\delta) > M] \to 1$ as $n \to \infty$ for $\{\phi_n\}$ defined by $n^{1-1/v}\phi_n^{1-\eta}s_n^{1+\gamma-\eta}L^o(s_n) \sim 1$ as $n \to \infty$, where again $s_n \to 0$ as $n \to \infty$, arbitrarily slowly. Here $\max_i\{n^{-1/v}|y_{i-1}|\} = o_p(1)$ because $v < \eta$. Noting that $\sup_{0<v<\eta}\{(-1 + 1/v)/(1 - \eta)\} = 1/\eta$, it follows that $n^\kappa d_n \to_p 0$ as $n \to \infty$, for any $\kappa < 1/\eta$.

(e) Here $\eta < 1$. Choose $v_n = \phi_n = n^{1/\eta}$ and define

$$a_{i,n}(\delta) = -(\delta y_{i-1}/v_n)\mathrm{sgn}(u_i)$$

so that $A_n(\delta) = \sum_{i=1}^n a_{i,n}(\delta)$. Note that for $\delta \ne 0$, $A_n(\delta)/\delta$ has a nondegenerate limiting distribution, namely, a symmetric stable distribution with index η, whereas $B_n(\delta)/|\delta| \le \sum_{i=2}^n |y_{i-1}|/\phi_n = O_p(1)$ in view of $\eta < 1$. And, from the definition of the $b_{i,n}(\delta)$ leading to (4), if we let $S_{i,n}(\delta) = I(\mathrm{sgn}(\delta y_{i-1})\mathrm{sgn}(u_i) > 0)$,

$$\begin{aligned}b_{i,n}(\delta) &= 2(\phi_n/v_n)(-u_i + \delta y_{i-1}/\phi_n)\{I(\delta y_{i-1}/\phi_n > u_i \ge 0) - I(\delta y_{i-1}/\phi_n < u_i < 0)\} \\ &= 2|-u_i + \delta y_{i-1}/\phi_n| I(|u_i| < |\delta y_{i-1}/\phi_n|) S_{i,n}(\delta) \\ &\ge 2|-u_i + \delta y_{i-1}/\phi_n| I(|u_i| < \theta|\delta y_{i-1}/\phi_n|) S_{i,n}(\delta) \\ &\ge 2(1-\theta)|\delta y_{i-1}/\phi_n| I(|u_i| < \theta|\delta y_{i-1}/\phi_n|) S_{i,n}(\delta)\end{aligned}$$

for $\theta \in (0,1)$, which we hold fixed in the following. We let

$$a_{i,n}(\delta) = a_{i,n}^o(\delta) + a_{i,n}^*(\delta) + r_{i,n}(\delta),$$

https://doi.org/10.1017/S0266466601174074 Published online by Cambridge University Press

where

$$a^o_{i,n}(\delta) = -(\delta y_{i-1}/v_n)\mathrm{sgn}(u_i)I(|u_i| < \theta|\delta y_{i-1}/\phi_n|)S_{i,n}(\delta)$$

$$a^*_{i,n}(\delta) = -(\delta y_{i-1}/v_n)\mathrm{sgn}(u_i)[1 - S_{i,n}(\delta)]$$

$$r_{i,n}(\delta) = -(\delta y_{i-1}/v_n)\mathrm{sgn}(u_i)I(|u_i| \geq \theta|\delta y_{i-1}/\phi_n|)S_{i,n}(\delta).$$

Here, for all i and n, $a^o_{i,n}(\delta) \leq 0$, $r_{i,n}(\delta) \leq 0$, and $a^*_{i,n}(\delta) \geq 0$. First note that because $\eta < 1$, for $\delta \neq 0$, $\sum_{i=1}^n a^*_{i,n}(\delta)/\delta$ has, from A3, a stable limiting distribution with index η and support $\Re^+$, so that $\lim_{n\to\infty} \Pr[|\sum_{i=1}^n a^*_{i,n}(\delta)|/|\delta| > \varepsilon]$ is close to one when $\varepsilon > 0$ is small. Second, combining A3 with $\beta = 0$ yields, for $c > 0$,

$$\Pr[|r_{i,n}(\delta)|/|\delta| > c] \leq 2\Pr[y_{i-1} > cv_n]\Pr[u_i > \theta|\delta|c]$$

$$\sim 2c_o^2 c^{-2\eta}(\theta|\delta|)^{-\eta}/n \quad \text{as } n \to \infty,$$

implying that

$$\lim_{n\to\infty} \Pr\left[\left|\sum_{i=1}^n r_{i,n}(\delta)\right|/|\delta| > \varepsilon\right]$$

can be made as close to zero as desired by making $|\delta|$ sufficiently large. Third, from the definition of the $a^o_{i,n}(\delta)$ and the bounds on the $b_{i,n}(\delta)$ introduced previously,

$$\sum_{i=1}^n [b_{i,n}(\delta) + a^o_{i,n}(\delta)] \geq [2(1-\theta) - 1]\sum_{i=1}^n |a^o_{i,n}(\delta)| \geq 0$$

for all i and n and $\theta \in (0, \frac{1}{2})$. It follows from these three considerations that

$$\lim_{n\to\infty} \Pr\left[\sum_{i=1}^n b_{i,n}(\delta)/|\delta| + \sum_{i=1}^n a_{i,n}(\delta)/|\delta| > 0\right]$$

is close to one for $|\delta|$ sufficiently large. The same conclusion applies to $\lim_{n\to\infty}\Pr[A_n(\delta) + B_n(\delta) > 0]$. Because $Z_n(0) = 0$ and $Z_n(\delta)$ is convex it follows that with probability approaching one as $n \to \infty$, the minimizer of $Z_n(\delta)$ lies in $[-d, d]$ for sufficiently large d, and the result follows. The idea behind this last step is similar to one employed by Pollard (1991) in the proof of his Theorem 3 (where $\eta = 1$). ■

https://doi.org/10.1017/S0266466601174074 Published online by Cambridge University Press