

**COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
AT YALE UNIVERSITY**

**Box 2125, Yale University
New Haven, Connecticut 06520**

COWLES FOUNDATION DISCUSSION PAPER NO. 897R

NOTE: Cowles Foundation Discussion Papers are preliminary materials circulated to stimulate discussion and critical comment. Requests for single copies of a Paper will be filled by the Cowles Foundation within the limits of the supply. References in publications to Discussion Papers (other than acknowledgment that a writer had access to such unpublished material) should be cleared with the author to protect the tentative character of these papers.

**TIME SERIES REGRESSION WITH UNIT ROOT
AND INFINITE VARIANCE ERRORS**

by

P. C. B. Phillips

August 1989

TIME SERIES REGRESSION WITH A UNIT ROOT AND INFINITE VARIANCE ERRORS

by

P. C. B. Phillips*

*Cowles Foundation for Research in Economics
Yale University*

Proposed Running Head: Regression with infinite variance errors

first revision: April 1989

second revision: August 1989

*I am grateful to D. Andrews, N. Chan and C. Z. Wei for helpful comments on the first version of this paper. My thanks also go to Mico Loretan for preparing the figures, to Glenna Ames for her skill and effort in keyboarding the manuscript of this paper and to the NSF for research support under Grant No. SES 8519595.

ABSTRACT

In [4] Chan and Tran give the limit theory for the least squares coefficient in a random walk with iid errors that are in the domain of attraction of a stable law. This note discusses their results and provides generalizations to the case of $I(1)$ processes with weakly dependent errors whose distributions are in the domain of attraction of a stable law. General unit root tests are also studied. It is shown that the semiparametric corrections suggested by the author in other work [22] for the finite variance case continue to work when the errors have infinite variance. Surprisingly, no modifications to the formulae given in [22] are required. The limit laws are expressed in terms of ratios of quadratic functionals of a stable process rather than Brownian motion. The correction terms that eliminate nuisance parameter dependencies are random in the limit and involve multiple stochastic integrals that may be written in terms of the quadratic variation of the limiting stable process. Some extensions of these results to models with drifts and time trends are also indicated.

1. INTRODUCTION

Suppose $\{y_t\}$ is generated by

$$(1) \quad y_t = \beta y_{t-1} + u_t; \quad t = 1, \dots, n$$

with

$$\beta = 1$$

from an initialization at $t = 0$ in which y_0 is any random variable. Interest centers on the least squares estimate

$$\hat{\beta} = \left[\sum_1^n y_{t-1}^2 \right]^{-1} \left(\sum_1^n y_t y_{t-1} \right)$$

of β in (1). Chan and Tran [4] investigate the asymptotic behavior of $\hat{\beta}$ as $n \rightarrow \infty$ and show that for a certain family of iid errors u_t with infinite variance the limit distribution of $n(\hat{\beta}-1)$ can be characterized in terms of a functional of a Lévy process. They assume that $y_0 = 0$, that $u_1 \in \mathcal{D}(\alpha)$, i.e. u_1 is in the domain of attraction of a stable law with index α ($0 < \alpha < 2$), and that the limit law satisfies the scaling condition

$$(2) \quad n^{-1/\alpha}(u_1 + \dots + u_n) \equiv u_1$$

where " \equiv " signifies equality in distribution (both here and elsewhere in the paper).

Chan and Tran show that under these conditions

$$(3) \quad n(\hat{\beta}-1) \Rightarrow \int_0^1 U^- dU / \int_0^1 U^2$$

where U^- denotes the left limit of U , $U(r)$ is a Lévy process on the space of CADLAG functions $D[0,1]$ and " \Rightarrow " signifies weak convergence of measures. Their proof uses the weak convergence (from Resnick [27]) of the component processes

$$(4) \quad \left[a_n^{-1} \Sigma_1^{[nr]} u_j, a_n^{-2} \Sigma_1^{[nr]} u_j^2 \right] \Rightarrow (U(r), V(r))$$

where $(U(r), V(r))$ is a Lévy process in $D[0,1]^2$ and the normalization is

$$(5) \quad a_n = n^{1/\alpha} \ell(n)$$

for some slowly varying function $\ell(n)$. They also show that

$$(6) \quad (1/2)(U^2(1) - V(1)) \equiv \int_0^1 U^- dU$$

and then (4) and (5) together with the continuous mapping theorem give the final result (3).

We shall start with some remarks on these results.

(i) Suppose u_1 belongs to what is known as the normal domain of attraction of a stable law with index α (see [12] Ch. 2 for a discussion of normal domains of attraction).

We shall denote this by writing

$$(7) \quad u_1 \in \mathcal{ND}(\alpha).$$

The tail behavior of u_1 when $0 < \alpha < 2$ is then of the Pareto-Lévy form

$$(8) \quad P(u_1 < u) = \frac{c_1 a^\alpha}{|u|^\alpha} (1 + o(1)), \quad u < 0$$

$$(9) \quad P(u_1 > u) = \frac{c_2 a^\alpha}{u^\alpha} (1 + o(1)), \quad u > 0$$

as $|u| \rightarrow \infty$ ([12], p. 92). Here c_1 and c_2 are constants with $c_1, c_2 \geq 0$ and $c_1 + c_2 = 1$ (by suitable selection of a). We shall call a the scale parameter. In this case the norming sequence in (5) is of the simple form $a_n = an^{1/a}$ so that $\ell(n) = a$ in (5).

(ii) Second, when (7) applies we have $u_1^2 \in \mathcal{ND}(\alpha/2)$. Moreover, (4) can be replaced with an explicit limit law given as follows in terms of a stable process U_α :

$$(10) \quad \left[a_n^{-1} \Sigma_1^{[nr]} u_j, a_n^{-2} \Sigma_1^{[nr]} u_j^2 \right] \Rightarrow \left[U_\alpha(r), \int_0^r (dU_\alpha)^2 \right]$$

where $a_n = an^{1/\alpha}$. Here $U_\alpha(r)$ is a standard stable process with index α and unit scale coefficient. When $u_t \equiv -u_t$ (so that the distribution of u_t is symmetric), $U_\alpha(r)$ is a symmetric stable process and the characteristic function of $U_\alpha(1)$ has the form $e^{-c|s|^\alpha}$ where

$$c = \begin{cases} \Gamma(1-\alpha)\cos(\pi\alpha/2), & \alpha \neq 1 \\ \pi/2 & , \alpha = 1 \end{cases}$$

[12, pp. 44–45]. Moreover, $U_\alpha(r) \equiv r^{1/\alpha}U_\alpha(1)$. These and other properties of stable processes are derived in Ito [13, pp. 157–162]. Figures 1–4 in the Appendix display typical trajectories of stable processes for various values of α . These may be contrasted with that of a typical Wiener process (Figure 5). We shall henceforth write

$$U_\alpha(r) \equiv \text{SP}(\alpha)$$

to signify that U_α is a standard stable process with index α . Note that the class $\{\text{SP}(\alpha) : 0 < \alpha \leq 2\}$ is a subclass of the Lévy processes and that each member of $\text{SP}(\alpha)$ has no Wiener component in its Ito representation [27, p. 72] when $\alpha < 2$. (Thus, $S = 0$ and $a = 0$ in equation (10) of [4].)

In the representation (8) above, $\int_0^r (dU_\alpha)^2$ is a multiple stochastic integral which represents the usual quadratic variation (or square bracketed) process. This is sometimes represented in the notation $[U]_r$ (e.g. [19, p. 175]). But we prefer the integral notation in the present context because it helps to make the weak convergence that is given in (10) more intuitive and easily understood. In fact, Resnick proves (10) in [27, p. 94] but uses a notation for the limit in terms of point processes rather than stochastic integrals.

(iii) In place of the distributional equivalence of (6) (Theorem 2(ii) of [4]) we have indeed the direct equation

$$(11) \quad V(1) = \int_0^1 (dU_\alpha)^2 = U_\alpha^2(1) - 2 \int_0^1 U_\alpha^- dU_\alpha.$$

This follows from the Ito calculus for semimartingales (see, e.g., Kopp's second formula on page 160 of [15]).

It is most easily understood by noting that the stochastic differential dU_α^2 can be broken down as follows:

$$\begin{aligned} dU_\alpha^2 &= (U_\alpha^- + dU_\alpha)^2 - (U_\alpha^-)^2 \\ &= 2U_\alpha^- dU_\alpha + (dU_\alpha)^2. \end{aligned}$$

Integration then yields formula (11) directly.

(iv) Finally, we note that when $\alpha = 2$

$$U(r) \equiv W(r) \equiv \text{BM}(1)$$

or standard Brownian motion (BM). In this case

$$(12) \quad (dU_\alpha)^2 = (dW)^2 = dr$$

and (11) reduces to the usual formula

$$\int_0^1 W dW = (1/2)(W^2(1) - 1)$$

for the Brownian motion stochastic integral. This is, in fact, the only case for which $(dU_\alpha)^2$ is nonrandom. Note also that since $u_1 \in \mathcal{M}(\alpha)$ we necessarily have a finite variance $\sigma^2 = E(u_1^2) < \infty$ when $\alpha = 2$ (see [8], p. 92) and the scale factor is $a_n = \sigma n^{1/2}$. The limit distribution given in (3) is then the ratio of Brownian functionals $\int_0^1 W dW / \int_0^1 W^2$.

2. GENERAL I(1) MODELS WITH INFINITE VARIANCE ERRORS

In econometrics there has recently been a good deal of interest in models such as (1) where allowance is made for some weak dependence in the errors u_t . The resulting time series are known as I(1) or integrated processes. My review paper [23] and Park and Phillips [20, 21] provide a general discussion of models where these processes occur. In an earlier paper on scalar time series [22] I showed how to deal with such general error processes in constructing tests for the presence of a unit root. This involved a semiparametric correction to eliminate the bias in estimation of the regression coefficient that is due to the serial correlation in u_t . It is interesting to explore how this procedure needs to be modified when u_t has infinite variance.

2.1. Models with MA(1) Errors

Let us start by considering the simple case of MA(1) errors

$$(13) \quad u_t = \epsilon_t + \theta\epsilon_{t-1}$$

where $\epsilon_t \in \mathcal{ND}(\alpha)$, $0 < \alpha \leq 2$, $|\theta| < 1$ and ϵ_t is iid with $\epsilon_t \equiv -\epsilon_t$ (i.e. ϵ_t is symmetrically distributed). When $0 < \alpha < 2$ the tails of ϵ_t are of the Pareto-Lévy form with scale parameter a and $c_1 = c_2 = 1/2$. When $\alpha = 2$, $a^2 = E(\epsilon_t^2) < \infty$. Define $P_t = \sum_1^t \epsilon_j$ and then

$$\begin{aligned} \Sigma_1^n y_{t-1} u_t &= \Sigma_1^n P_{t-1} \epsilon_t + \theta^2 \Sigma_1^n P_{t-2} \epsilon_{t-1} + \theta \Sigma_1^n P_{t-1} \epsilon_{t-1} + \theta \Sigma_1^n P_{t-2} \epsilon_t \\ &= \Sigma_1^n P_{t-1} \epsilon_t + \theta^2 \Sigma_1^n P_{t-2} \epsilon_{t-1} + \theta \Sigma_1^n \epsilon_{t-1}^2 + \theta \Sigma_1^n P_{t-2} (\epsilon_t + \epsilon_{t-1}) \\ (14) \quad &= \Sigma_1^n P_{t-1} \epsilon_t + \theta^2 \Sigma_1^n P_{t-2} \epsilon_{t-1} + \theta (\Sigma_1^n P_{t-1} \epsilon_t + \Sigma_1^n P_{t-2} \epsilon_{t-1}) \\ &\quad + \theta \Sigma_1^n \epsilon_{t-1}^2 - \theta \Sigma_1^n \epsilon_{t-1} \epsilon_t. \end{aligned}$$

$$(15) \quad \Sigma_1^n y_{t-1}^2 = \Sigma_1^n P_{t-1}^2 + \theta^2 \Sigma_1^n P_{t-2}^2 + 2\theta \Sigma_1^n P_{t-1} P_{t-2}.$$

In the above we use the initialization $y_0 = 0$ to simplify formulae but this involves no loss of generality for the subsequent argument. Next observe that

$$(16) \quad a^{-2} n^{-2/\alpha-1} \sum_1^n P_{t-1}^2 \Rightarrow \int_0^1 U_\alpha^2,$$

$$(17) \quad a^{-2} n^{-2/\alpha} \sum_1^n P_{t-1} \epsilon_t \Rightarrow (1/2) \{U_\alpha(1)^2 - \int_0^1 (dU_\alpha)^2\} = \int_0^1 U_\alpha^- dU_\alpha$$

and

$$(18) \quad a^{-2} n^{-2/\alpha} \sum_1^n \epsilon_{t-1}^2 \Rightarrow \int_0^1 (dU_\alpha)^2$$

where $U_\alpha(r) \equiv SP(\alpha)$, whereas

$$(19) \quad a^{-2} n^{-2/\alpha} \sum_1^n \epsilon_t \epsilon_{t-1} \xrightarrow{p} 0.$$

The latter follows because although the product $\epsilon_t^2 \in \mathcal{ND}(\alpha/2)$ the cross product $\epsilon_t \epsilon_{t-1}$ does not lie in $\mathcal{ND}(\alpha/2)$. In fact, $\epsilon_t \epsilon_{t-1} \in \mathcal{D}(\alpha)$ as shown by Cline [5] and Breiman [2] and (19) then follows directly. Note that the norming sequence for sums of these cross products is $b_n = b(n \ln n)^{1/\alpha}$ with $b = a^2$ as shown in the Appendix.

Combining (14)–(19) we obtain the following limit result for the least squares estimator $\hat{\beta}$

$$(20) \quad \begin{aligned} n(\hat{\beta}-1) &\Rightarrow \frac{(1+\theta)^2 \int_0^1 U_\alpha^- dU_\alpha + \theta \int_0^1 (dU_\alpha)^2}{(1+\theta)^2 \int_0^1 U_\alpha^2} \\ &= \frac{\int_0^1 U_\alpha^- dU_\alpha + \theta(1+\theta)^{-2} \int_0^1 (dU_\alpha)^2}{\int_0^1 U_\alpha^2} \end{aligned}$$

This formula generalizes (3) to the case of models with a unit root and MA(1) errors. Note that the second term in the numerator of (20) is random when $\alpha < 2$. When $\alpha = 2$ it is simply the constant $\theta/(1+\theta)^2$. In that case (i.e. $\alpha = 2$) the expression was given in my earlier paper [22, p. 283]. The effect of serially correlated errors in the unit

root model (1) is therefore to induce a second order random bias term in the limit distribution of the least squares estimator. When the errors on (1) have finite variance this bias term is nonrandom and, as shown in [22], it depends on the serial correlation properties of the errors. The latter is still true in the infinite variance case but the bias term also has a random factor which depends on the quadratic variation $\int_0^1 (dU_\alpha)^2$.

The simplest way of dealing with the bias that is induced by MA(1) errors is to use instrumental variables (IV) estimation with y_{t-2} acting as an instrument for y_{t-1} in (1). Call the resulting estimator $\tilde{\beta} = \Sigma_1^n y_t y_{t-2} / \Sigma_1^n y_{t-1} y_{t-2}$. It is easy to see that

$$n(\tilde{\beta}-1) \Rightarrow \int_0^1 U_\alpha^- dU_\alpha / \int_0^1 U_\alpha^2$$

as in (3). IV estimators of this type have been suggested in the finite variance case by Hall [11] and Phillips and Hansen [26]. It is interesting to see that they continue to work as a direct method of eliminating the second order bias in the infinite variance case also.

2.2. Models with Weakly Dependent Errors

Let u_t be generated by the linear process

$$(21) \quad u_t = d(L)\epsilon_t = \sum_{j=0}^{\infty} d_j \epsilon_{t-j}; \quad d_0 = 1, \quad d(1) \neq 0$$

where ϵ_t has the same properties as in (11). As shown in Brockwell and Davis [3, p. 480] (see also [14]) the series defining u_t converges almost surely if the coefficients d_j are δ -summable i.e.

$$(22) \quad \sum_0^{\infty} |d_j|^\delta < \infty, \quad \text{with } 0 < \delta < \alpha \wedge 1.$$

If u_t is generated by a stable ARMA process then its moving average representation (21) has coefficients which decline geometrically so that (22) is certainly satisfied in this case. We shall assume that this is so and that there exists some $K > 0$ and γ ($0 < \gamma < 1$) such that

$$(23) \quad |d_j| < K\gamma^j, \text{ all } j.$$

This includes a wide class of processes like ARMA models although it can certainly be weakened substantially—but that is not our object here.

Define the partial sum processes

$$X_n(r) = a_n^{-1} \sum_1^{[nr]} u_t, \quad Y_n(r) = a_n^{-2} \sum_1^{[nr]} u_t^2.$$

The following theorem gives us a time series extension of the limit law (8).

THEOREM 2.1

$$(24) \quad (X_n(r), Y_n(r)) \Rightarrow (X(r), Y(r)) = \left[\omega U_\alpha(r), \sigma^2 \int_0^r (dU_\alpha)^2 \right]$$

where

$$(25) \quad \omega = \sum_0^\infty d_j = d(1), \quad \sigma^2 = \sum_0^\infty d_j^2.$$

PROOF. Define the approximating time series

$$u_{t\ell} = \sum_0^\ell d_j \epsilon_{t-j}$$

and the random elements

$$X_{n\ell}(r) = a_n^{-1} \sum_1^{[nr]} u_{t\ell}, \quad Y_{n\ell}(r) = a_n^{-2} \sum_1^{[nr]} u_{t\ell}^2$$

with $a_n = an^{1/\alpha}$. For fixed ℓ we find just as in the MA(1) case that

$$(X_{n\ell}(r), Y_{n\ell}(r)) \Rightarrow (\bar{X}_\ell(r), \bar{Y}_\ell(r)) = \left[(\sum_0^\ell d_j) U_\alpha(r), (\sum_0^\ell d_j^2) \int_0^r (dU_\alpha)^2 \right]$$

as $n \rightarrow \infty$. Moreover it is clear that

$$(\bar{X}_\ell(r), \bar{Y}_\ell(r)) \Rightarrow (X(r), Y(r))$$

as $\ell \rightarrow \infty$. To prove (24) we then need only verify that for any $\delta > 0$ we have

$$(26) \quad \lim_{\ell \rightarrow \infty} \limsup_{n \rightarrow \infty} P(d(Z_{n\ell}, Z_n) \geq \delta) = 0$$

where $Z_{n\ell} = (X_{n\ell}, Y_{n\ell})$ and $Z_n = (X_n, Y_n)$ —see Billingsley [1, Theorem 4.2, p. 25]. In (26) $d(\cdot, \cdot)$ denotes the Skorohod metric on the product space $D[0,1]^2$. Since

$$d(Z_{n\ell}, Z_n) \leq \max\{\sup_r |X_{n\ell} - X_n|, \sup_r |Y_{n\ell} - Y_n|\}$$

we have the inequality

$$(27) \quad P(d(Z_{n\ell}, Z_n) \geq \delta) \leq P(\sup_r |X_{n\ell} - X_n| \geq \delta) + P(\sup_r |Y_{n\ell} - Y_n| \geq \delta).$$

It is therefore sufficient to show that the limits of the two members on the right side of (27) are zero.

Write

$$X_n(r) - X_{n\ell}(r) = a_n^{-1} \Sigma_1^{nr} (u_t - u_{t\ell}) = a_n^{-1} \Sigma_1^{nr} \Sigma_{\ell+1}^{\infty} d_j \epsilon_{t-j}$$

and

$$(28) \quad \begin{aligned} & P(\sup_r |X_n(r) - X_{n\ell}(r)| \geq \delta) \\ & \leq P\left[\max_{1 \leq k \leq n} |a_n^{-1} \Sigma_1^k \Sigma_{\ell+1}^{\infty} d_j \epsilon_{t-j} 1(|\epsilon_{t-j}| \leq a_n)| > \delta/2 \right] \\ & \quad + P\left[\sup_r |a_n^{-1} \Sigma_1^{nr} \Sigma_{\ell+1}^{\infty} d_j \epsilon_{t-j} 1(|\epsilon_{t-j}| > a_n)| > \delta/2 \right]. \end{aligned}$$

Let $\xi_{tn} = \Sigma_{\ell+1}^{\infty} d_j \epsilon_{t-j} 1(|\epsilon_{t-j}| \leq a_n)$, $\sigma_n^2 = E(\epsilon_t 1(|\epsilon_t| \leq a_n))^2$ and $\mathcal{F}^t = \sigma(\epsilon_s : s \leq t)$ i.e. the natural filtration associated with $\{\epsilon_t\}$. Then

$$E(\xi_{tn} | \mathcal{F}^{t-m}) = \Sigma_{(\ell+1)\vee m}^{\infty} d_j \epsilon_{t-j} 1(|\epsilon_{t-j}| \leq a_n)$$

and

$$E\left[E(\xi_{tn} | \mathcal{F}^{t-m})\right]^2 = \left[\Sigma_{(\ell+1)\vee m}^{\infty} d_j^2\right] \sigma_n^2 = \psi_m \sigma_n^2$$

so that $\{\xi_{tn}\}$ constitutes a mixingale array [18; 10, p. 25] with $\psi_m = O(\gamma^{m\vee \ell})$ because

of (23). Setting $S_{nk} = \sum_{t=1}^k \xi_{tn}$ we have by McLeish's inequality [18; 10, p. 26]

$$E\left[\max_{k \leq n} S_{nk}^2\right] \leq K_\ell n \sigma_n^2$$

with $K_\ell = O(\gamma^\ell)$. Thus

$$\begin{aligned} P\left[\max_{1 \leq k \leq n} |a_n^{-1} S_{nk}| > \delta/2\right] &\leq 4\delta^{-2} E\left[\max_{1 \leq k \leq n} a_n^{-1} |S_{nk}|\right]^2 \\ &= 4\delta^{-2} E\left[\max_{1 \leq k \leq n} a_n^{-2} S_{nk}^2\right] \leq 4\delta^{-2} a_n^{-2} n \sigma_n^2 K_\ell. \end{aligned}$$

Observe that

$$\sigma_n^2 = E(\epsilon_t^2 1(|\epsilon_t| \leq a_n)) = \int_{-a_n}^{a_n} \epsilon^2 f(\epsilon) d\epsilon \sim \frac{\alpha a^\alpha}{(2-\alpha) a_n^{\alpha-2}}$$

and therefore

$$\lim_{n \rightarrow \infty} n \sigma_n^2 / a_n^2 = \alpha / (2-\alpha).$$

It follows that

$$\lim_{\ell \rightarrow \infty} \lim_{n \rightarrow \infty} \sup P\left[\max_{1 \leq k \leq n} |a_n^{-1} S_{nk}| > \delta/2\right] \leq \lim_{\ell \rightarrow \infty} 4\delta^{-2} \alpha (2-\alpha)^{-1} K_\ell = 0.$$

This deals with the first term on the right side of the inequality (28).

Turning to the second term and setting $\eta_{tn} = \sum_{\ell+1}^{\infty} d_j \epsilon_{t-j} 1(|\epsilon_{t-j}| > a_n)$ we note that for any $0 < f < \alpha \wedge 1$

$$\begin{aligned} P(\sup_r |a_n^{-1} \sum_1^{[nr]} \eta_{tn}| > \delta/2) &= P(\sup_r |a_n^{-1} \sum_1^{[nr]} \eta_{tn}|^f > (\delta/2)^f) \\ &\leq P(a_n^{-f} \sum_1^n |\eta_{tn}|^f > (\delta/2)^f) \\ &\leq 2^f \delta^{-f} a_n^{-f} \sum_1^n \sum_{\ell+1}^{\infty} |d_j|^f E(|\epsilon_{t-j}|^f 1(|\epsilon_{t-j}| > a_n)) \\ &= 2^f \delta^{-f} a_n^{-f} \sum_{\ell+1}^{\infty} |d_j|^f E(|\epsilon_t|^f 1(|\epsilon_t| > a_n)). \end{aligned}$$

But

$$E(|\epsilon_t|^f \mathbf{1}(|\epsilon_t| > a_n)) = \alpha a^{(\alpha-f)-1} a_n^{-(\alpha-f)}$$

so that

$$\lim_{n \rightarrow \infty} n a_n^{-f} a_n^{f-\alpha} = a^{-\alpha}.$$

Since $\sum_{\ell+1}^{\infty} |d_j|^f = O(\gamma^{\ell f}) \rightarrow 0$ as $\ell \rightarrow \infty$ we deduce that

$$\lim_{\ell \rightarrow \infty} \limsup_{n \rightarrow \infty} P(\sup_r |a_n^{-1} \Sigma_1^{[nr]} \eta_{tn}| > \delta/2) = 0.$$

It follows from (28) that

$$\lim_{\ell \rightarrow \infty} \limsup_{n \rightarrow \infty} P(\sup_r |X_n(r) - X_{n\ell}(r)| \geq \delta) = 0.$$

Treating the second term of (27) in the same way we deduce that (26) holds and the theorem is proved. \square

We may now proceed as in the MA(1) error model to obtain the limit distribution of the least squares estimator $\hat{\beta}$. Note that

$$\begin{aligned} a_n^{-2} \Sigma_1^n y_{t-1} u_t &= (1/2) \left\{ (a_n^{-1} \Sigma_1^n u_t)^2 - a_n^{-2} \Sigma_1^n u_t^2 \right\} + o_p(1) \\ &\Rightarrow (1/2) \left\{ \omega^2 U_\alpha^2(1) - \sigma^2 \int_0^1 (dU_\alpha)^2 \right\} \\ (29) \quad &= \omega^2 \int_0^1 U_\alpha^- dU_\alpha + (1/2)(\omega^2 - \sigma^2) \int_0^1 (dU_\alpha)^2. \end{aligned}$$

Also

$$(30) \quad n^{-1} a_n^{-2} \Sigma_1^n y_{t-1}^2 \Rightarrow \omega^2 \int_0^1 U_\alpha^2.$$

Since joint weak convergence applies we deduce that

$$(31) \quad n(\hat{\beta}-1) \Rightarrow \left\{ \int_0^1 U_\alpha^2 \right\}^{-1} \left\{ \int_0^1 U_\alpha^- dU_\alpha + (1/2)(1 - \sigma^2/\omega^2) \int_0^1 (dU_\alpha)^2 \right\}$$

generalizing both (3) and (20). Note that when $\alpha = 2$, (31) reduces to the expression derived earlier in [22, Theorem 3.1]. In that case the second term of (29) is simply the constant $(1/2)(\omega^2 - \sigma^2)$ and the second order bias induced in the numerator of (31) is nonrandom.

Similar results apply for other functionals like the t-ratio

$$t_\beta = (\hat{\beta}-1)/s_{\hat{\beta}}, \quad s_{\hat{\beta}}^2 = n^{-1} \Sigma_1^n (y_t - \hat{\beta}y_{t-1})^2 \left[\Sigma_1^n y_{t-1}^2 \right]^{-1}.$$

Here we have

$$(32) \quad t_\beta \Rightarrow (\omega/\sigma) \left\{ \int_0^1 (dU_\alpha)^2 \int_0^1 U_\alpha^2 \right\}^{-1/2} \left\{ \int_0^1 U_\alpha^- dU_\alpha + (1/2)(1 - \sigma^2/\omega^2) \int_0^1 (dU_\alpha)^2 \right\}.$$

Again this reduces to the formula given in [22, p. 282] for the finite variance case.

2.3. The Effect of Semiparametric Corrections

When $\{\epsilon_t\}$ is iid(0,1) (i.e. $\alpha = 2$ and $\text{var}(\epsilon_t) = 1$) and u_t is defined by (21) we see that

$$\sigma^2 = \text{var}(u_t), \quad \omega^2 = 2\pi f_u(0)$$

where $f_u(\cdot)$ is the spectral density of u_t . When $\alpha < 2$, the variance and the spectrum of u_t are not finite because $E(\epsilon_t^2) = \infty$. The quantities σ^2 and ω^2 do exist in this case at least as they are defined in (25). We shall call them pseudo-variances because they represent what would be the contribution to these variances in the usual formulae after the variance of ϵ_t is scaled out. These contributions remain finite even in the infinite variance case.

Similar remarks apply to the spectrum. When $\alpha = 2$ the autocorrelogram sequence for u_t in (21) is given by

$$(33) \quad \rho(h) = E(u_t u_{t+h}) / E(u_t^2) = \sum_1^{\infty} d_j d_{j+h} / \sum_1^{\infty} d_j^2, \quad h = 1, 2, \dots$$

Again the effects of $E(\epsilon_t^2)$ are scaled out and $\rho(h)$ is well defined as the final member of (33) when $\alpha < 2$ (see also [4] and [5] on this point). The ρ -spectrum of u_t may then be defined as the Fourier transform of (33) i.e.

$$f_u^{(\rho)}(\lambda) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \rho(h) e^{i\lambda h}.$$

Davis and Resnick [7, 8] show that the sample autocorrelations are consistent for $\rho(h)$ when $\alpha < 2$ i.e.

$$r(h) = \sum_1^{n-h} u_t u_{t+h} / \sum_1^n u_t^2 \xrightarrow{p} \rho(h)$$

and that when ϵ_t has Pareto-Lévy tails

$$(34) \quad r(h) - \rho(h) = O_p((\ln n/n)^{1/\alpha}).$$

In consequence, it is simple to construct consistent estimators of $f_u^{(\rho)}(\lambda)$ using conventional spectral estimates based on the sample correlogram $\{r(h) : h = 1, 2, \dots\}$.

In [22] I suggested some semiparametric corrections to $n(\hat{\beta}-1)$ that asymptotically eliminate the nuisance parameter dependencies in the finite variance case. The statistic based on the coefficient estimate $\hat{\beta}$ has the form

$$(35) \quad Z(\hat{\beta}) = n(\hat{\beta}-1) - (1/2) \left\{ n^{-2} \sum_1^n y_{t-1}^2 \right\}^{-1} (\hat{\omega}^2 - \hat{\sigma}^2)$$

where $\hat{\omega}^2$ and $\hat{\sigma}^2$ are consistent estimators of $\omega^2 = 2\pi f_u(0)$ and $\sigma^2 = E(u_t^2)$, respectively. Noting that in this case

$$\omega^2 = \sigma^2 + 2\lambda \quad \text{with} \quad \lambda = \sum_{k=1}^{\infty} E(u_0 u_k)$$

we may write $Z(\hat{\beta})$ in the alternate form

$$(36) \quad Z(\hat{\beta}) = \left[n^{-2} \sum_1^n y_{t-1}^2 \right]^{-1} (n^{-1} \sum_1^n y_{t-1} u_t - \hat{\lambda})$$

where $\hat{\lambda}$ is consistent for λ . As shown in [22] in the finite variance model

$$(37) \quad Z(\hat{\beta}) \Rightarrow \left[\int_0^1 W^2 \right]^{-1} \int_0^1 W dW$$

whose distribution is free of nuisance parameters. Thus, $Z(\hat{\beta})$ forms the basis of a test for the presence of a unit root which is asymptotically similar for a wide class of weakly dependent errors with finite variance.

In the infinite variance ($\alpha < 2$) case we have already seen that

$$a_n^{-2} \sum_1^n u_t^2 \Rightarrow \sigma^2 \int_0^1 (dU_\alpha)^2, \text{ where } \sigma^2 = \sum_0^\infty d_j^2.$$

Since $\hat{\beta}$ is consistent the same result holds if we replace u_t by the residuals $\hat{u}_t = y_t - \hat{\beta} y_{t-1}$. Thus we may write

$$(38) \quad na_n^{-2} \hat{\sigma}^2 \Rightarrow \sigma^2 \int_0^1 (dU_\alpha)^2.$$

Turning to ω^2 , we note that this parameter is usually estimated by a kernel procedure that leads to an expression of the general form

$$(39) \quad \hat{\omega}^2 = 2\pi \hat{f}_u(0) = \sum_{j=-M}^M k(j/M) c(j),$$

where

$$(40) \quad c(j) = n^{-1} \sum_1^n u_t u_{t+j}, \quad 1 \leq t+j \leq n$$

and the lag window $k(\cdot)$ is a bounded even function defined on the interval $[-1, 1]$ with $k(0) = 1$. M is a bandwidth parameter in (39) and it satisfies $M \rightarrow \infty$ and $M/n \rightarrow 0$ as $n \rightarrow \infty$. For example, when $k(j/M) = 1 - |j|/M$, $\hat{\omega}^2$ is the Bartlett estimator of what would be the long run variance of u_t if $\alpha = 2$ (see [16] for further discussion).

Observe that for fixed j

$$na_n^{-2}c(j) \Rightarrow (\sum_0^{\infty} d_i d_{i+j}) \int_0^1 (dU_{\alpha})^2.$$

The same result also applies when u_t is replaced by the residual \hat{u}_t in (40). I shall not give a complete derivation here but using the same approach as that in the proof of Theorem 3.1 of my paper [24] it can be shown that, if $M = o(n^{1/2})$ as $n \rightarrow \infty$, then

$$(41) \quad na_n^{-2}\hat{\omega}^2 \Rightarrow \omega^2 \int_0^1 (dU_{\alpha})^2.$$

Now note that we may write

$$Z(\hat{\beta}) = n(\hat{\beta}-1) - (1/2) \left\{ n^{-1} a_n^{-2} \sum_1^n y_{t-1}^2 \right\}^{-1} \{ na_n^{-2} (\hat{\omega}^2 - \hat{\sigma}^2) \}.$$

From (31), (38) and (41) we deduce that

$$(42) \quad Z(\hat{\beta}) \Rightarrow \left[\int_0^1 U_{\alpha}^2 \right]^{-1} \left(\int_0^1 U_{\alpha}^{-} dU_{\alpha} \right).$$

This result generalizes (37) to the infinite variance case and of course also includes (37) since $U_{\alpha}(r) \equiv W(r)$ when $\alpha = 2$.

The t-ratio statistic may be analyzed in the same way. The statistic I suggested in [22] is based on t_{β} and has the form

$$Z(t) = (\hat{\sigma}/\hat{\omega})t_{\beta} - (1/2)(\hat{\omega}^2 - \sigma^2) \left\{ \hat{\omega} (n^{-2} \sum_1^n y_{t-1}^2)^{1/2} \right\}^{-1}.$$

As $n \rightarrow \infty$ we find that

$$(43) \quad Z(t) \Rightarrow \left\{ \int_0^1 (dU_{\alpha})^2 \int_0^1 U_{\alpha}^2 \right\}^{-1/2} \int_0^1 U_{\alpha}^{-} dU_{\alpha}$$

When $\alpha = 2$ the limit distribution becomes

$$\left[\int_0^1 W^2 \right]^{-1/2} \int_0^1 W dW$$

as given in [22].

These results show that the semiparametric corrections suggested in [22] continue to work when the errors have infinite variance even though they were designed specifically to eliminate nuisance parameters in the finite variance case. The reason for this is that in the infinite variance case there are still parametric dependencies in the limit distributions of the coefficient estimator and its t -statistic (as shown in (31) and (32)). These dependencies involve the parameters (σ^2, ω^2) that we have described as pseudo-variances. They represent what would be the (short and long run) variances of u_t if α were equal to 2 and $E(\epsilon_t^2) < \infty$. The semiparametric corrections eliminate these pseudo-variances from the limit distributions in the infinite variance case just as they do when the actual variances are finite.

3. ADDITIONAL REMARKS

(i) The final results (42) and (43) apply under somewhat weaker assumptions than those given here. We may, for example, replace the requirement that $\epsilon_t \in \mathcal{M}(\alpha)$ with $\epsilon_t \in \mathcal{D}(\alpha)$. This affects the norming sequence a_n but we still find for a suitable choice of a_n that

$$(44) \quad \left[a_n^{-1} \sum_1^{[nr]} \epsilon_t, a_n^{-2} \sum_1^{[nr]} \epsilon_t^2 \right] \Rightarrow \left[U_\alpha(r), \int_0^r (dU_\alpha)^2 \right]$$

and

$$a_n^{-2} \sum_1^n \epsilon_t \epsilon_{t+j} \xrightarrow{p} 0; \quad 1 \leq t+j \leq n, \quad j \neq 0.$$

These limits ensure that (38) and (41) hold, giving the final results (42) and (43) as stated.

(ii) We may also relax the symmetry condition $\epsilon_t \equiv -\epsilon_t$, although many of the arguments given in Section 2 will then need modification. When $\alpha < 1$ no further requirement beyond $\epsilon_t \in \mathcal{D}(\alpha)$ seems to be needed. When $\alpha > 1$ we require $E(\epsilon_t) = 0$, as in [4], so that sums involving ϵ_t do not need to be centered. When $\alpha = 1$ an

additional condition such as

$$(45) \quad b_n = E\{\epsilon_t 1(|\epsilon_t| \leq a_n)\} = 0, \text{ for all } n$$

ensures that centering of the sums involving ϵ_t is unnecessary, although this is hardly weaker than the symmetry condition $\epsilon_t \equiv -\epsilon_t$. An alternative proof of Theorem 2.1 under such conditions will be reported elsewhere.

(iii) Our analysis and results extend easily to models with drifts and time trends (or other deterministic functions) in place of (1). In such cases the time series may be regarded as filtered prior to their use in regressions such as (1). The effects can then be determined by treating the filtered series as regression residuals. For instance, when there is polynomial detrending we construct the residual process y_t from the least squares regression

$$y_t = \hat{\mu}_0 + \hat{\mu}_1 t + \dots + \hat{\mu}_p t^p + \underline{y}_t.$$

Then in place of (29) and (30) we have

$$(29)' \quad a_n^{-1} \sum_{t=1}^n \underline{y}_{t-1} u_t \Rightarrow \omega^2 \int_0^1 \underline{U}_\alpha^- dU_\alpha + (1/2)(\omega^2 - \sigma^2) \int_0^1 (dU_\alpha)^2$$

and

$$(30)' \quad n^{-1} a_n^{-2} \sum_{t=1}^n \underline{y}_{t-1}^2 \Rightarrow \omega^2 \int_0^1 \underline{U}_\alpha^2$$

where $\underline{U}_\alpha = QU_\alpha$ is simply the projection of U_α in $L_2[0,1]$ on the orthogonal complement of the space spanned by the polynomial functions $\{0(r), 1(r), \dots, p(r); j(r) = r^j\}$.

We deduce that if $\hat{\beta}$ is the least squares estimator of β in

$$(1)' \quad y_t = \mu_0 + \mu_1 t + \dots + \mu_p t^p + \beta y_{t-1} + u_t$$

under $\beta = 1$ and $\mu_p = 0$ then

$$(31)' \quad n(\hat{\beta}-1) \Rightarrow \left\{ \int_0^1 U_\alpha^2 \right\}^{-1} \left\{ \int_0^1 U_\alpha^- dU_\alpha + (1/2)(1 - \sigma^2/\omega^2) \int_0^1 (dU_\alpha)^2 \right\}.$$

Semiparametric corrections to eliminate the nuisance parameters in (31)' may be made as in Section 3.2. The situation here is entirely analogous to that explored in the finite variance case by Park and Phillips [20, 21].

(iv) The limit theory given here also has applications in the context of limit theorems for self normalized sums. These have been considered elsewhere recently by several authors [1, 17, 25, 27]. In [25] the bimodality of t -ratio statistics of the form

$$t_\epsilon = \Sigma_1^n \epsilon_j / \left[\Sigma_1^n \epsilon_j^2 \right]^{1/2}$$

was explored when $\epsilon_j \in \mathcal{M}(\alpha)$ and $0 < \alpha < 2$. The reason for the bimodality in the distribution of t_ϵ , which occurs in both finite and asymptotic samples, is the statistical dependence between the numerator and denominator random variables in t_ϵ . When $\epsilon_j \equiv -\epsilon_j$, $\epsilon_j \in \mathcal{D}(\alpha)$ and ϵ_j is iid it follows immediately from (44) that

$$(46) \quad t_\epsilon \Rightarrow U_\alpha(1) / \left[\int_0^1 (dU_\alpha)^2 \right]^{1/2}$$

as $n \rightarrow \infty$. Using (9) we may now write the limit law in the form

$$(47) \quad U_\alpha(1) / \left\{ U_\alpha(1)^2 - 2 \int_0^1 U_\alpha^- dU_\alpha \right\}^{1/2}.$$

The bimodality (with modes at ± 1) then arises because of the occurrence of $U_\alpha(1)$ in the numerator and denominator elements of (47) when $0 < \alpha < 2$. When $\alpha = 2$ the denominator is nonrandom, of course, and (46) corresponds to the conventional limit theory for the t -ratio in the finite variance case i.e. $t_\epsilon \Rightarrow U_2(1) \equiv N(0,1)$.

APPENDIX: ON THE TAIL BEHAVIOR OF THE PRODUCT

$X = x_1 x_2$ OF INDEPENDENT VARIATES $x_i \in \mathcal{D}(\alpha)$

Suppose $x_i \equiv -x_i$ and then $X \equiv -X$. Let $f_i(x)$ be the density of x_i , which we take to be continuous over $(-\infty, \infty)$. Setting $c_1 = c_2 = 1/2$ in (8) and (9), we have the following tail behavior

$$f_i(x) = (1/2) \alpha a_i^\alpha |x|^{-\alpha-1} (1 + O(1)), \quad |x| > k$$

for some (possibly large) constant $k > 0$. The density of X is now

$$\begin{aligned} f(X) &= 2 \int_0^\infty (1/x) f_1(x) f_2(X/x) dx \\ &= 2 \left\{ \int_0^\infty + \int_k^{X/k} + \int_{X/k}^\infty \right\} (1/x) f_1(x) f_2(X/x) dx. \end{aligned}$$

Suppose $X > k^2$ and let c be a generic constant in what follows. The first integral is less than

$$2c \int_0^k (1/x) f_2(X/x) dx = 2c X^{-1-\alpha} \left\{ \int_0^k \alpha a_2^\alpha x^\alpha dx + o(1) \right\} = O(X^{-1-\alpha}).$$

The third is less than

$$2c \int_{X/k}^\infty (1/x) f_1(x) dx = 2c \alpha a_1^\alpha \int_{X/k}^\infty x^{-\alpha-2} dx = O(X^{-1-\alpha}).$$

The second integral dominates and has the form

$$(1/2) \alpha^2 a_1^\alpha a_2^\alpha X^{-1-\alpha} \int_k^{X/k} (1/x) dx = (1/2) \alpha^2 a_1^\alpha a_2^\alpha X^{-1-\alpha} \ln X (1 + o(1))$$

so that $f(X) = O((\ln X) X^{-1-\alpha})$ as $|X| \rightarrow \infty$. Integrating again we get the tail behavior of the cumulative distribution function

$$\text{cdf}(-X) = (1/2)f \ell n(-X)(-X)^{-\alpha}(1 + o(1))$$

$$1 - \text{cdf}(X) = (1/2)f(\ell n X)X^{-\alpha}(1 + o(1))$$

as $X \rightarrow \infty$ where $f = \alpha a_1^{\alpha} a_2^{\alpha}$. Since $\ell n(\cdot)$ is slowly varying at infinity these tail probabilities ensure that $X \in \mathcal{D}(\alpha)$ [12, Theorem 2.6.1]. Setting $b_n = b(n \ell n n)^{1/\alpha}$ with $b = a_1 a_2$ we have

$$n(\text{cdf}(b_n X)) \rightarrow (1/2)(-X)^{-\alpha}, \quad X < 0$$

$$n[1 - \text{cdf}(b_n X)] \rightarrow (1/2)X^{-\alpha}, \quad X > 0.$$

The norming sequence for sums of iid variates distributed as x is therefore $b_n = a_1 a_2 (n \ell n n)^{1/\alpha}$ by the argument in [12, p. 76]. Explicitly we have $b_n = \inf\{x | P(|X| > x) \leq 1/n\} = a_1 a_2 (n \ell n n)^{1/\alpha}$. Note that $X \in \mathcal{D}(\alpha)$ and not $\mathcal{ND}(\alpha)$ because of the presence of the slowly varying function $\ell n(\cdot)$ in the tail formulae.

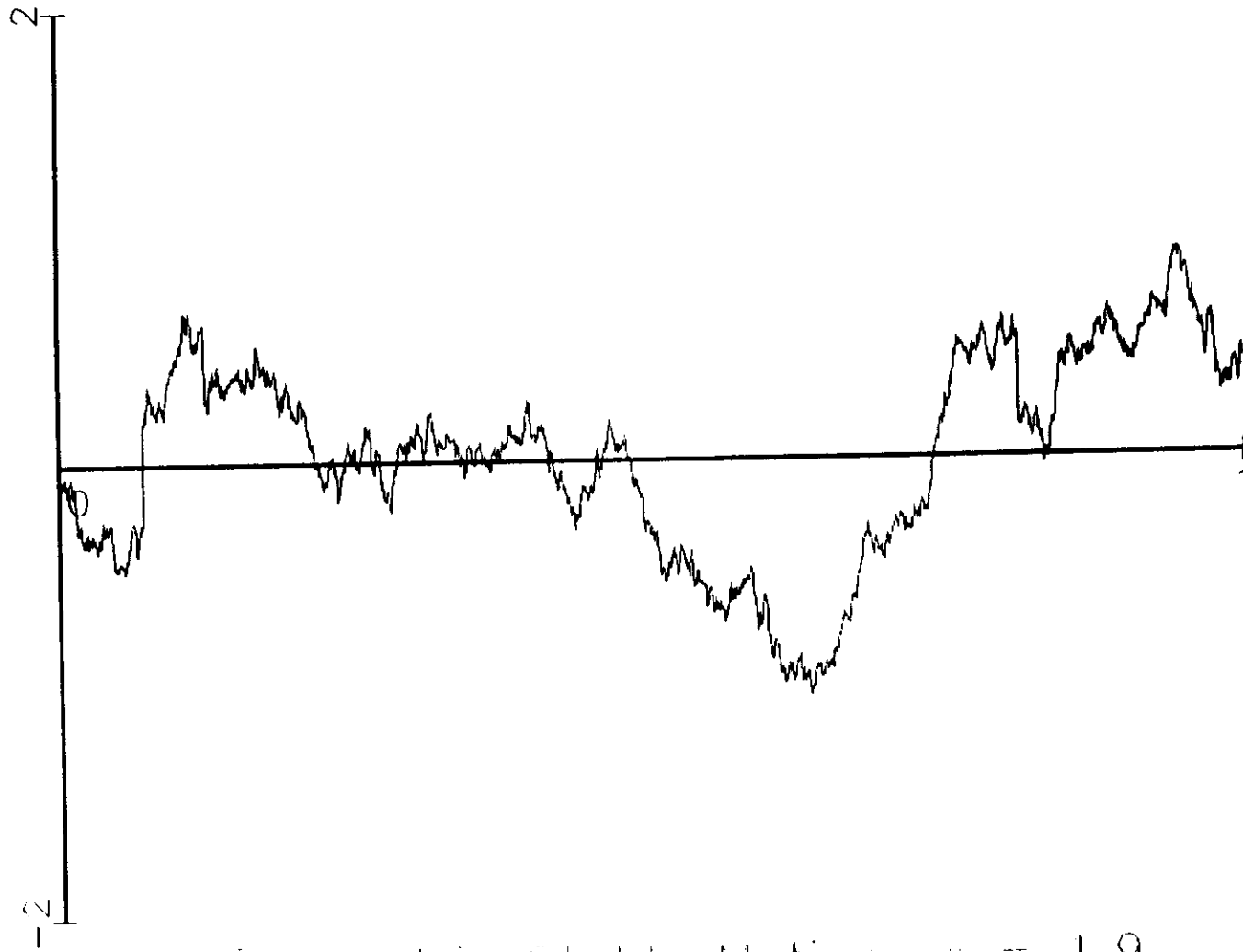
When the variates $x_i \in \mathcal{ND}(\alpha_i)$ $\alpha_1 \neq \alpha_2$ we obtain by a similar argument the result $X \in \mathcal{ND}(\alpha_1 \wedge \alpha_2)$. In this case the product variate X is in the normal domain of attraction of the law of the component variate with the smallest exponent.

REFERENCES

- [1] Billingsley, P. *Convergence of Probability Measures*. New York: Wiley, 1968.
- [2] Breiman, L. "On some limit theorems similar to the arc-sin law," *Theory of Probability and its Applications*, 10 (1965), 323–331.
- [3] Brockwell, P. J. and R. A. Davis. *Time Series: Theory and Methods*. New York: Springer, 1987.
- [4] Chan, N. H. and L. T. Tran. "On the first order autoregressive process with infinite variance," *Econometric Theory* (1989) forthcoming.
- [5] Cline, D. B. H. "Estimation and linear prediction for regression, autoregression and ARMA with infinite variance data," Ph.D dissertation, Colorado State University, 1983.
- [6] Darling, D. A. "A note on a limit theorem," *Annals of Probability*, 3 (1975), 876–888.
- [7] Davis, R. and S. Resnick. "Limit theory for moving averages of random variables with regularly varying tail probabilities," *Annals of Probability*, 13 (1985), 179–195.
- [8] _____. "Limit theory for the sample covariance and correlation functions of moving averages," *Annals of Statistics*, 14 (1986), 533–558.
- [9] Feller, W. *An Introduction to Probability Theory and its Applications*, Vol. 2 (2nd Ed.). New York: Wiley, 1971.
- [10] Gallant, A. R. and H. White. *A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models*. Oxford: Blackwell, 1988.
- [11] Hall, A. "Testing for a unit root in the presence of moving average errors," *Biometrika*, 1989 (forthcoming).
- [12] Ibragimov, I. A. and Y. V. Linnik. *Independent and Stationary Sequences of Random Variables*. Groninger: Wolters-Noordhoff, 1971.
- [13] Ito, K. *Lectures on Stochastic Processes*. Bombay: Tata Institute, 1961.
- [14] Kanter, M. "Linear sample spaces and stable processes," *Journal of Functional Analysis*, 9 (1972), 441–459.
- [15] Kopp, P. E. *Martingales and Stochastic Integrals*. Cambridge: Cambridge University Press, 1984.
- [16] LePage, R., M. Woodroffe and J. Zinn. "Convergence to a stable distribution via order statistics," *Annals of Probability*, 9 (1981), 624–632.
- [17] Logan, B., C. Mallows, S. O. Rice and L. Shepp. "Limit distributions of self normalized sums," *Annals of Probability*, 1 (1973), 788–809.

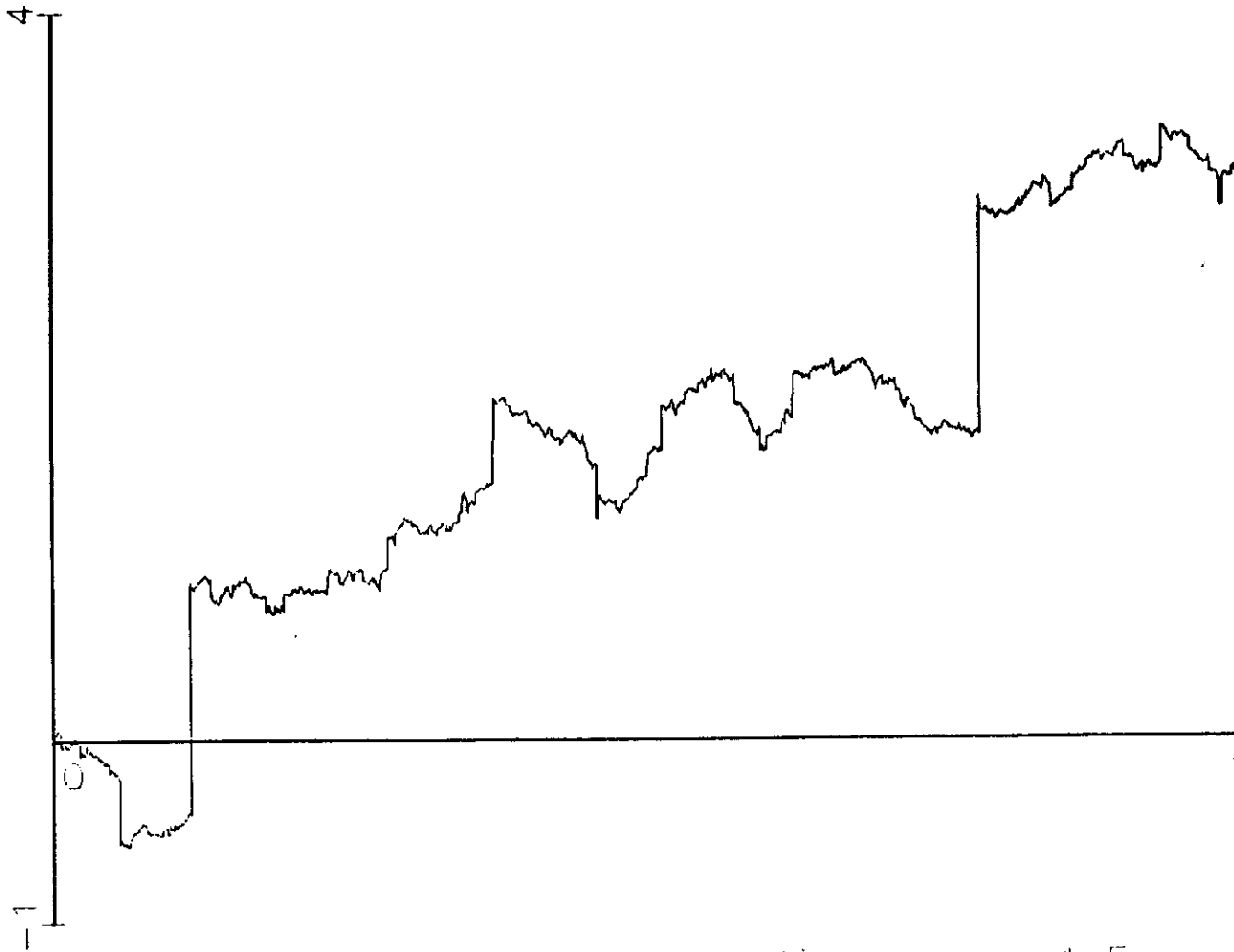
- [18] McLeish, D. L. "A maximal inequality and dependent strong laws," *Annals of Probability*, 3 (1975), 829–839.
- [19] Métivier, M. *Semimartingales*. New York: Walter de Gruyter, 1982.
- [20] Park, J. Y. and P. C. B. Phillips. "Statistical inference in regressions with integrated processes: Part 1," *Econometric Theory*, 4 (1988), 468–498.
- [21] _____. "Statistical inference in regressions with integrated processes: Part 2," *Econometric Theory*, 5 (1989), 95–132.
- [22] Phillips, P. C. B. "Time series regression with a unit root," *Econometrica*, 55 (1987), 277–301.
- [23] _____. "Multiple regression with integrated processes," in N. U. Prabhu (ed.), *Statistical Inference from Stochastic Processes, Contemporary Mathematics, Vol. 80*. Providence: American Mathematical Society, 1988.
- [24] _____. "Spectral regression for cointegrated time series," in W. Barnett (Ed.), *Nonparametric and Semiparametric Methods in Economics and Statistics*, CUP, 1989 (forthcoming).
- [25] Phillips, P. C. B. and V. Hajivassiliou. "Bimodal t-ratios," Cowles Foundation Discussion Paper No. 842, July 1987.
- [26] Phillips, P. C. B. and B. Hansen. "Statistical inference in instrumental variables regression with I(1) processes," Cowles Foundation Discussion Paper No. 869, July 1988.
- [27] Resnick, S. I. "Point processes regular variation and weak convergence," *Advances in Applied Probability*, 18 (1986), 66–138.

Figure 1



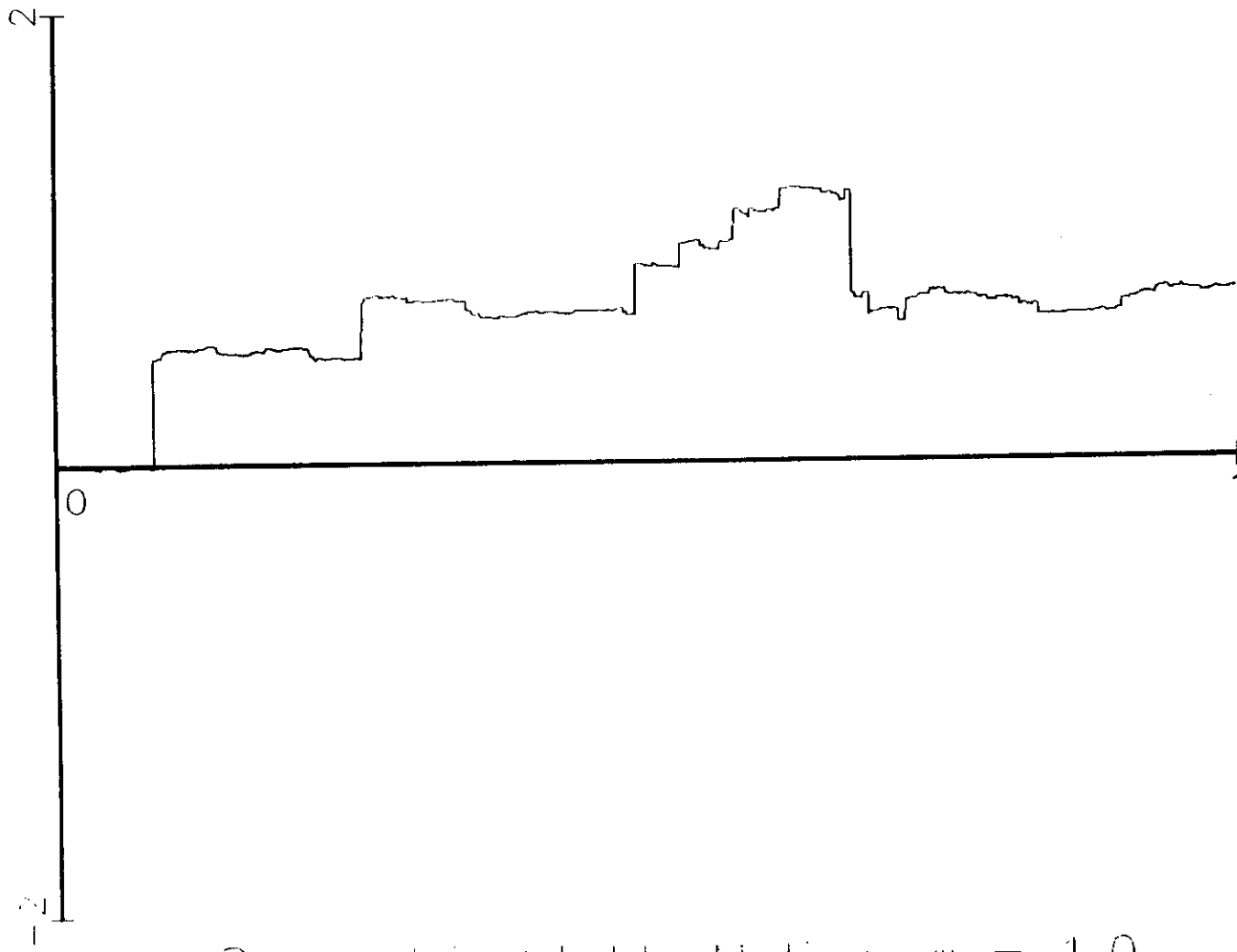
Symmetric Stable Motion, $\alpha = 1.9$

Figure 2



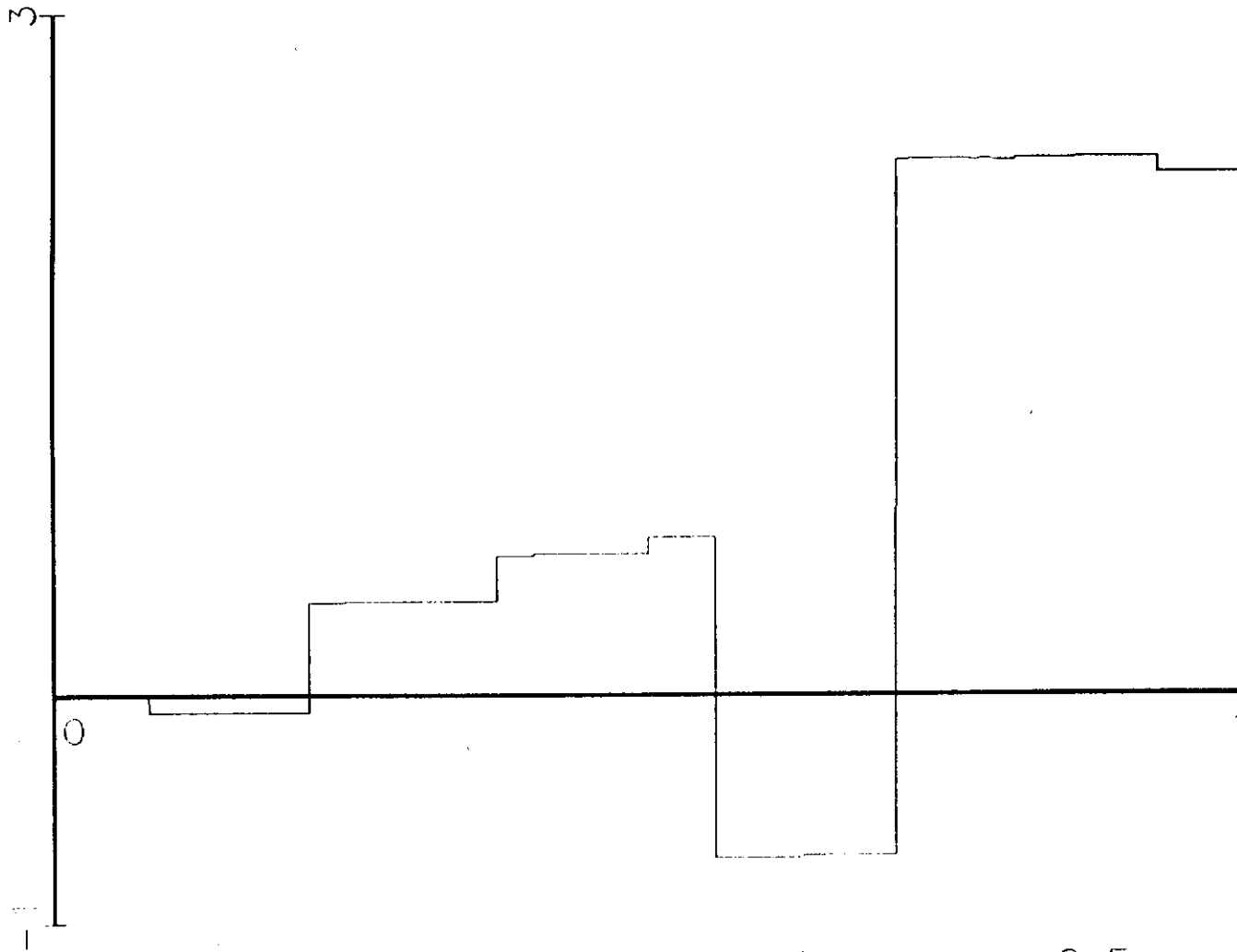
Symmetric Stable Motion, $\alpha = 1.5$

Figure 3



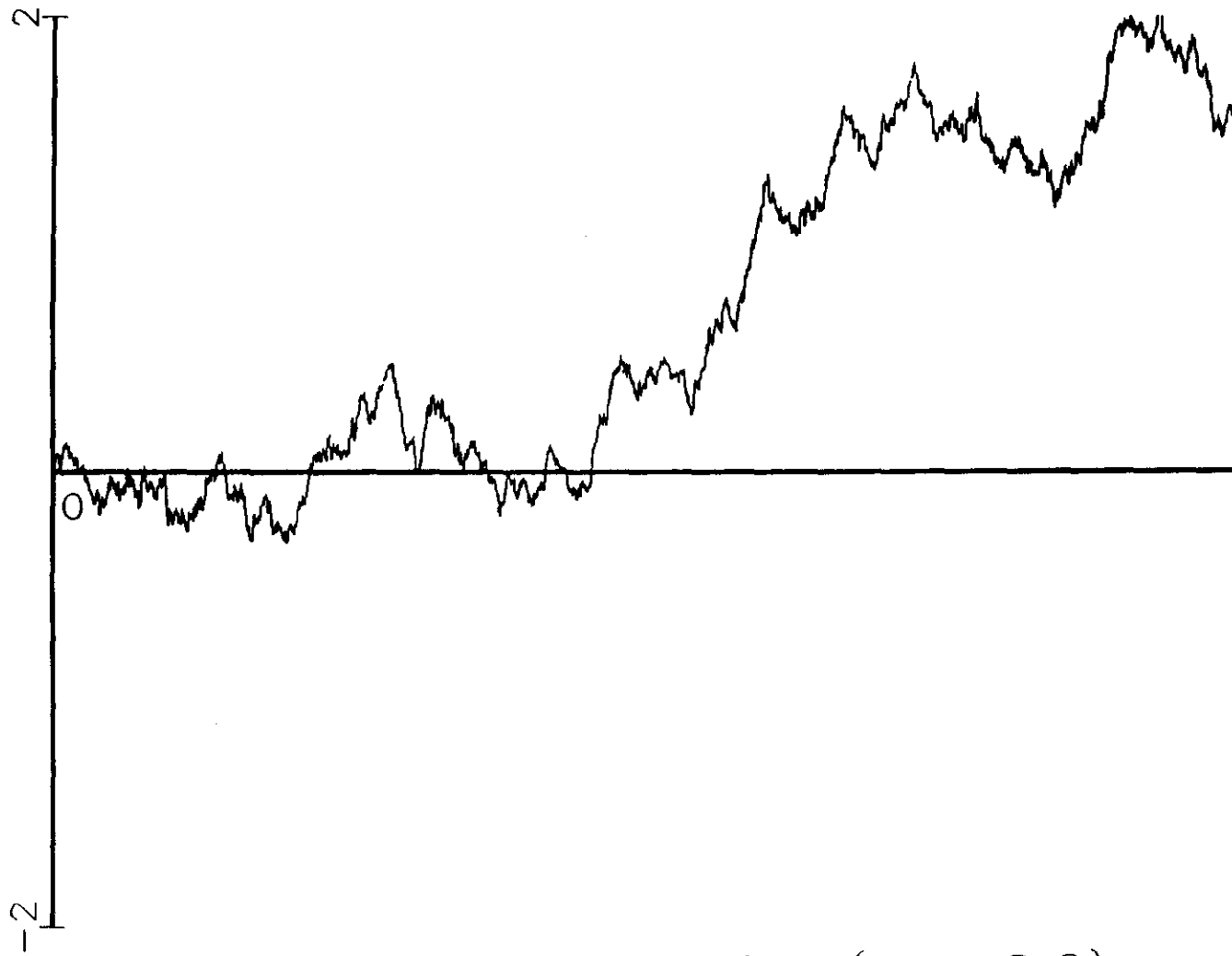
Symmetric Stable Motion, $\alpha = 1.0$

Figure 4



Symmetric Stable Motion, $\alpha = 0.5$

Figure 5



Brownian Motion ($\alpha = 2.0$)