

Yale University

EliScholar – A Digital Platform for Scholarly Publishing at Yale

Cowles Foundation Discussion Papers

Cowles Foundation

1-1-1989

The Durbin-Watson Ratio under Infinite Variance Errors

Peter C.B. Phillips

Mico Loretan

Follow this and additional works at: <https://elischolar.library.yale.edu/cowles-discussion-paper-series>



Part of the [Economics Commons](#)

Recommended Citation

Phillips, Peter C.B. and Loretan, Mico, "The Durbin-Watson Ratio under Infinite Variance Errors" (1989). *Cowles Foundation Discussion Papers*. 1142.

<https://elischolar.library.yale.edu/cowles-discussion-paper-series/1142>

This Discussion Paper is brought to you for free and open access by the Cowles Foundation at EliScholar – A Digital Platform for Scholarly Publishing at Yale. It has been accepted for inclusion in Cowles Foundation Discussion Papers by an authorized administrator of EliScholar – A Digital Platform for Scholarly Publishing at Yale. For more information, please contact elischolar@yale.edu.

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

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

COWLES FOUNDATION DISCUSSION PAPER NO. 898R

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 References in publications to Discussion Papers (other than acknowledgment that a writer had access to such unpublished material) should be cleared with the author to protect the tentative character of these papers.

THE DURBIN-WATSON RATIO UNDER INFINITE VARIANCE ERRORS

by

P. C. B. Phillips and Mico Loretan

August 1989

THE DURBIN-WATSON RATIO UNDER INFINITE VARIANCE ERRORS*

by

Peter C. B. Phillips and Mico Loretan

*Cowles Foundation for Research in Economics
Yale University*

Address for Proofs: Professor P. C. B. Phillips
Cowles Foundation
Box 2125 Yale Station
New Haven, CT 06520-2125
USA

Tel: 203 421-4708
Fax: 203 432-6167

ABSTRACT

This paper studies the properties of the von Neumann ratio for time series with infinite variance. The asymptotic theory is developed using recent results on the weak convergence of partial sums of time series with infinite variance to stable processes and of sample serial correlations to functions of stable variables. Our asymptotics cover the null of iid variates and general moving average (MA) alternatives. Regression residuals are also considered. In the static regression model the Durbin-Watson statistic has the same limit distribution as the von Neumann ratio under general conditions. However, in dynamic models the results are more complex and more interesting. When the regressors have thicker tail probabilities than the errors we find that the Durbin-Watson and von Neumann ratio asymptotics are the same. But when the tail probability of the errors is at least as thick as that of the regressors then different results apply. It is shown that for finite variance models our results specialize to those of the Durbin h -statistic and equivalent LM test asymptotics. Some Monte Carlo results are reported, illustrating the effects of infinite variance errors and regressors in finite samples.

First Draft: January, 1989
Revision: August, 1989

*We thank Maxwell King, two referees and participants of the University of Zurich econometrics seminar for constructive comments and suggestions. Our thanks also go to Glenna Ames for keyboarding the manuscript of this paper and to the NSF for research support under Grant No. SES 8519595 and No. SES 8821180.

1. INTRODUCTION

In recent years there has been a revival of interest in the properties of the Durbin-Watson (DW) statistic for testing for serial correlation in regression residuals. The statistic was originally designed to detect the presence of first order autoregressive (AR(1)) errors in the linear regression model and it is known to deliver good power and to have certain optimal properties in this case. For example, the DW statistic is approximately the locally best invariant (LBI) test against AR(1) errors, as shown in Durbin and Watson (1971). Recent work by Kariya (1988) and King and Evans (1988) shows that the test continues to have good power and retains the LBI behavior against other forms of error behavior whenever the first order serial correlation coefficient is non zero. The reader is referred to King (1987) for an historical review and for further references on the subject.

Although the properties of the DW statistic have been intensively studied by econometricians, there seems to have been little work on the behavior of this test in regression models where the errors have infinite variance. The case of regression errors in the spherically symmetric class, which admits some infinite variance distributions, has been studied analytically; and the robustness properties of the Durbin-Watson test in this case are quite well known—see Kariya (1977, 1980) and King (1980). But the null distribution of the DW statistic is the same for spherically symmetric errors as it is for Gaussian errors, so the assumption of spherical symmetry has no distributional effects. Similar distributional invariance under spherical symmetry applies to other scale invariant statistics like the t -ratio, as has been known since Fisher (1925). In contrast, the effects of statistically independent infinite variance errors can be substantial. Indeed, for the conventional t -ratio it is known that infinite variance, independent draws induce a bimodality in the density—see Logan *et al.* (1973) and Phillips and Hajivassiliou (1987). These major differences in the distribution of the t -ratio between spherically symmetric and statistically independent variates with infinite variance indicate that it is of real interest to explore the

effects of the latter in the case of other test statistics and regression diagnostics. This is a topic on which there has been virtually no work to date, with the exception of some simulation evidence such as that of Bartels and Goodhew (1981).

The aim of the present paper is to show that an asymptotic analysis at least is well within reach. We start by developing a limit theory for the von Neumann ratio in a non regression context. Our asymptotics cover the null of iid variates and a general time series alternative that allows for infinite order MA representations. Next we provide an asymptotic distribution theory for the DW test in static models where the regressors and the errors have distributions within the normal domain of attraction of a stable law with exponent $\alpha < 2$. Our results are extended to dynamic regression models where they are related to known asymptotic theory for Durbin's h-statistic and associated Lagrange multiplier (LM) tests, which are now in popular use for models with lagged dependent variables. A Monte Carlo study is reported, illustrating the behavior of the tests in finite samples under infinite variance errors.

Our theoretical development is made possible by some recent results in the probability literature on weak convergence for time series with infinite variance. In particular, Resnick (1986) provides many fundamental results of this type for sequences of partial sums and Davis and Resnick (1985a, 1985b, 1986) develop a general theory for sample covariances. Our theory is in large part an application of their results to the regression diagnostic context.

2. THE VON NEUMANN RATIO

Let $\{u_t\}$ be iid and suppose u_1 lies in the normal domain of attraction of a stable law of index α . We shall write this for convenience in the form

$$(1) \quad u_1 \in \mathcal{ND}(\alpha).$$

Note that a necessary and sufficient condition for $u_1 \in \mathcal{ND}(\alpha)$ when $0 < \alpha < 2$ is that the

tail behavior of u_1 be of the Pareto–Lévy form, i.e.

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

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

as $|u| \rightarrow \infty$, where a is a scale parameter and $c_1, c_2 \geq 0$ with $c_1 + c_2 = 1$. (See Ibragimov and Linnik (1971, Ch. 2) for this and for further information about domains of attraction, normal domains of attraction and stable variates.)

We shall also assume that $u_1 \equiv -u_1$ where " \equiv " signifies equality in distribution. This condition ensures that the distribution of u_1 is symmetric. It is convenient but not essential to our development (especially when $0 < \alpha < 1$). If u_1 were strictly stable rather than simply in $\mathcal{ND}(\alpha)$ its characteristic function would be $e^{-a^\alpha |s|^\alpha}$.

Our first concern will be with the von Neumann ratio

$$(2) \quad \text{VN} = \frac{\sum_2^n (u_t - u_{t-1})^2}{\sum_1^n u_t^2}.$$

Under (1) $u_t^2 \in \mathcal{ND}(\alpha/2)$ whereas $u_t u_{t-1} \in \mathcal{D}(\alpha)$, the domain of attraction of a stable law with index α . Note that $u_t u_{t-1}$ does not lie in a normal domain of attraction and this affects its norming sequence in partial sums such as those that occur in (2). Indeed, as shown in Phillips (1988), the tails of the variate $X = u_1 u_2$ when $0 < \alpha < 2$ are characterized by

$$(3) \quad \text{pdf}(X) = (1/2) \alpha^2 a^{2\alpha} (\ln |X|) X^{-\alpha-1} (1 + o(1)),$$

as $|X| \rightarrow \infty$. These tails are not of the Pareto–Lévy form, so that X does not lie in $\mathcal{ND}(\alpha)$. However, the function $\ln(\cdot)$ is slowly varying and the form of (3) does ensure that $X \in \mathcal{D}(\alpha)$ (Ibragimov and Linnik (1971), Theorem 2.6.1, p. 76).

Under these conditions suitably normed sample moments and sample serial covariances of u_t converge weakly to limiting stable variates. In particular, it is known (see

Davis and Resnick (1985b), p. 278) that

$$(4) \quad (a_n^{-2} \sum_1^n u_t^2, \tilde{a}_n^{-1} \sum_1^n u_t u_{t-1}) \Rightarrow (S_0, S_1)$$

where S_0 is stable (and positive) with index $\alpha/2$, S_1 is stable (and symmetric) with index α and the limit variates (S_0, S_1) are independent. These limit variates have characteristic functions

$$\begin{aligned} \text{cf}_{S_0}(t) &= E(e^{itS_0}) = \exp\left\{-\Gamma(1-\alpha/2)\cos(\pi\alpha/4)|t|^{\alpha/2}(1 - i\text{sgn}(t)\tan(\pi\alpha/4))\right\} \\ \text{cf}_{S_1}(t) &= E(e^{itS_1}) = \begin{cases} \exp\{-\Gamma(1-\alpha)\cos(\pi\alpha/2)|t|^\alpha\}, & \text{for } \alpha \neq 1 \\ \exp\{-(\pi/2)|t|\} & , \text{ for } \alpha = 1 \end{cases} \end{aligned}$$

(cf. Brockwell and Davis (1987), pp. 482–483; set $C = 1$ in their formulae (12.5.14) and (12.5.15)). Note that in (4) and throughout this paper the symbol " \Rightarrow " signifies weak convergence of the associated probability measures.

Note that the norming sequences in (4) are given by

$$(5) \quad a_n = an^{1/\alpha}, \quad \tilde{a}_n = b(n \ln(n))^{1/\alpha}$$

where $b = a^2$.

Next we write

$$(6) \quad a_n^2 \tilde{a}_n^{-1} (VN-2) = (-2\tilde{a}_n^{-1} \sum_2^n u_t u_{t-1}) / (a_n^{-2} \sum_1^n u_t^2) - \tilde{a}_n^{-1} (u_1^2 + u_n^2) / a_n^{-2} \sum_1^n u_t^2.$$

Then, noting that $a_n^2 \tilde{a}_n^{-1} = (n/\ln(n))^{1/\alpha}$, we deduce from (2), (4) and (6) that

$$(7) \quad (n/\ln(n))^{1/\alpha} (VN-2) \Rightarrow -2S_1/S_0$$

by direct use of the continuous mapping theorem. This gives us the asymptotic distribution of the von Neumann ratio under the null that $\{u_t\}$ is iid and in $\mathcal{ND}(\alpha)$.

We can proceed in a similar way under the alternative hypothesis. Let us suppose that u_t has a general moving average representation of the form

$$(8) \quad u_t = \sum_0^{\infty} d_j \epsilon_{t-j}, \quad d_0 = 1$$

where ϵ_t is iid, symmetric and in $\mathcal{MD}(\alpha)$. It will be convenient, but is not essential, to assume that the coefficients d_j in (8) are majorized by geometrically declining weights i.e. $|d_j| < K\theta^j$ for some $K > 0$ and some $0 < \theta < 1$. Then the series (8) converges a.s. (see e.g. Brockwell and Davis (1987), p. 480). Under the usual AR(1) alternative of

$$(9) \quad u_t = \varphi u_{t-1} + \epsilon_t, \quad |\varphi| < 1$$

we have $d_j = \varphi^j$ directly.

The tail behavior of the density of u_t , $f(u)$, under the alternative (8) is also of the Pareto form. Specifically,

$$(10) \quad f(u) \sim (1/2)\alpha a^\alpha \left[\sum_{j=0}^{\infty} |d_j|^\alpha \right]^{1/\alpha} |u|^{-\alpha-1}, \quad |u| \rightarrow \infty$$

(see, e.g. Brockwell and Davis (1987), p. 481). In (10) the symbol " \sim " is used in the usual sense that the difference between the expressions is $o(1)$. When ϵ_t is symmetric stable we have the distributional equivalence

$$(11) \quad u_t \equiv \left[\sum_{j=0}^{\infty} |d_j|^\alpha \right]^{1/\alpha} \epsilon_1$$

as may be verified by looking at the respective characteristic functions.

Although second moments of u_t in (8) are not finite when $0 < \alpha < 2$ we may, as in Davis and Resnick (1986), define a pseudo-correlogram for u_t by

$$(12) \quad \rho(h) = \sum_{j=0}^{\infty} d_j d_{j+h} / \sum_{j=0}^{\infty} d_j^2, \quad h = 1, 2, \dots$$

When $\alpha = 2$ and $Eu_t^2 < \infty$ this is equivalent to the usual correlogram given by

$$(13) \quad \rho(h) = E(u_t u_{t+h}) / E(u_t^2).$$

In studying the asymptotic behavior of the von Neumann ratio under the general alternative (8) it follows from (6) that the dominant term is given by the sample first correlation

$$\hat{\rho}(1) = \sum_1^{n-1} u_t u_{t+1} / \sum_1^n u_t^2.$$

Indeed, we may rewrite (6) as

$$a_n^2 \tilde{a}_n^{-1} \{VN - 2(1 - \rho(1))\} = -2a_n^2 \tilde{a}_n^{-1} \{\hat{\rho}(1) - \rho(1)\} + o_p(1)$$

where $a_n^2 \tilde{a}_n^{-1} = (n/\ln(n))^{1/\alpha}$ as before.

From the limit theory in Davis and Resnick (1986, pp. 555–556) we have

$$(n/\ln(n))^{1/\alpha} \{\hat{\rho}(1) - \rho(1)\} \Rightarrow \left[\sum_{j=1}^{\infty} |f_j|^\alpha \right]^{1/\alpha} S_1/S_0$$

where

$$(14) \quad f_j = \rho(j+1) + \rho(j-1) - 2\rho(j)\rho(1)$$

and the limit variates (S_0, S_1) are independent and stable with indexes $\alpha/2$ and α , respectively, just as in the case of (4) under the null hypothesis. We deduce that

$$(n/\ln(n))^{1/\alpha} \{VN - 2(1 - \rho(1))\} \Rightarrow -2 \left[\sum_{j=1}^{\infty} |f_j|^\alpha \right]^{1/\alpha} S_1/S_0.$$

Collecting these results together we have:

THEOREM 1. *Under the null hypothesis that u_t is iid symmetric and in $\mathcal{ND}(\alpha)$ with $0 < \alpha < 2$ the limit of the von Neumann statistic is given by*

$$(15) \quad (1/2)(n/\ln(n))^{1/\alpha} (VN - 2) \Rightarrow S_1/S_0.$$

Under the alternative that u_t is generated by (8) we have

$$(16) \quad (1/2)(n/\ell n(n))^{1/\alpha}(VN - 2(1 - \rho(1))) \Rightarrow \left[\sum_{j=1}^{\infty} |f_j|^\alpha \right]^{1/\alpha} S_1/S_0.$$

In both (15) and (16) the limit variates (S_0, S_1) are as in (4) and the constants f_j in (16) are given in (14).

REMARK (a) When $\alpha = 2$ the variance of u_1 is finite since $u_1 \in \mathcal{ND}(2)$. In this case the limit distributions (15) and (16) are both normal but the norming factor is $n^{1/2}$ rather than $(n/\ell n(n))^{1/2}$. In fact, $S_1 \equiv N(0,1)$ and $S_0 = 1$ a.s. The latter follows from the fact that when $\alpha = 2$ the norming sequence is $a_n = an^{1/2}$ and $n^{-1} \sum_1^n u_t^2 \xrightarrow{\text{a.s.}} a^2 = \text{var}(u_1)$. Relating this to the above we may write

$$(17) \quad a_n^{-2} \sum_1^n u_t^2 \Rightarrow S_0 \equiv \int_0^1 (dU_2)^2 = \int_0^1 (dW)^2 = \int_0^1 dr = 1 \text{ a.s.}$$

where $W(r)$ is standard Brownian motion and $U_\alpha(r)$ is a standard stable process with index α (see Phillips (1988) for further discussion of the multiple stochastic integral representation that appears in (17)). Also, when u_t is iid and in $\mathcal{ND}(2)$ we have $u_t u_{t-1} \in \mathcal{ND}(2)$ and

$$(18) \quad \tilde{a}_n^{-1} \sum_1^n u_t u_{t-1} \Rightarrow S_1 \equiv N(0,1)$$

where the norming sequence for the sample covariance is $\tilde{a}_n = a^2 n^{1/2}$ rather than $\tilde{a}_n = a^2 (n/\ell n(n))^{1/2}$ as in (4). Thus, when $\alpha = 2$ and $u_1 \in \mathcal{ND}(2)$ we find in place of (15) the result

$$(15)' \quad (1/2)n^{1/2}(VN-2) \Rightarrow N(0,1)$$

under the null. Similarly, under the alternative we find that $\epsilon_t \in \mathcal{ND}(2)$ implies that

$$(16)' \quad (1/2)n^{1/2}(VN - 2(1 - \rho(1))) \Rightarrow N(0, \sum_1^\infty f_j^2)$$

(cf. Anderson (1971), Theorem 8.4.6, p. 489 and Davis and Resnick (1986), p. 555 on the limit distribution of sample serial correlations in the finite variance case).

REMARK (b) Figure 1 shows the empirical density of the ratio S_1/S_0 for various values of α including the case $\alpha = 2$ when $S_1/S_0 \equiv N(0,1)$, as remarked above. These densities and the quantiles given in Tables 1a and 1b show that tests that are based on the centered and scaled statistic

$$(19) \quad v = (1/2)(n/\ln(n))^{1/\alpha}(VN-2)$$

and that employ critical values for the $\alpha = 2$ case (delivered from the standard normal distribution) are conservative for α close to 2 but liberal as $\alpha \rightarrow 0$.

REMARK (c) When $\rho(1) \neq 0$, (16) shows that tests based on v are consistent. In particular when $\rho(1) > 0$ we see that v diverges to $-\infty$ at the rate $(n/\ln(n))^{1/\alpha}$. Thus, a one sided test of the hypothesis

$$H_0 : \rho(1) = 0 \text{ against } H_1 : \rho(1) > 0$$

that was based on v would reject H_0 when $v < -v_\alpha(0.05)$ where $-v_\alpha(0.05)$ is the 5% nominal asymptotic critical value in the left tail of the asymptotic distribution under the null. Such a test is consistent. Its sampling properties are investigated in the Monte Carlo study reported in Section 4.

REMARK (d) The representations (15) and (16) in the theorem use the fact that $S_1 \equiv -S_1$ so that both S_1 and the ratio S_1/S_0 have symmetric distributions.

REMARK (e) When the alternative hypothesis is the usual AR(1) representation (9) we find that $f_j = \varphi^{j-1}(1 - \varphi^2)$ and

$$\left[\sum_{j=1}^{\infty} |f_j|^\alpha \right]^{1/\alpha} = (1 - \varphi^2) \left[1 - |\varphi|^\alpha \right]^{-1/\alpha}$$

leading to the explicit form

$$(1 - \varphi^2) \left[1 - |\varphi|^\alpha \right]^{-1/\alpha} S_1/S_0$$

of the limit distribution under this alternative. Again, when $\alpha = 2$ we have

$$(1 - \varphi^2)^{1/2} S_1 \equiv N(0, 1 - \varphi^2)$$

which is the well known asymptotic distribution of the first order serial correlation coefficient in the finite variance case.

REMARK (f) The density of the ratio S_1/S_0 has an infinite mode at the origin when $\alpha < 2$. To see this, write $Z = S_1/S_0$ and note that in view of independence the joint density of (S_1, S_0) factors as follows:

$$\text{pdf}(S_1, S_0) = f_1(S_1)f_0(S_0).$$

Then

$$\text{pdf}(Z) = \int_{S_0 > 0} S_0 f_1(S_0 Z) f_0(S_0) dS_0$$

and the density at $Z = 0$ is infinite since the integral that defines $E(S_0)$ is divergent when $\alpha < 2$. The sharp peak in the density is reflected in the graphs of Figure 1, especially for the case where $\alpha = 0.5$. Cases such as this where there is a discontinuity in the density pose interesting problems for kernel density estimation. In particular, to avoid kernel smoothing over of the discontinuity it is necessary to use one sided kernel estimation on either side of the discontinuity. This seems to be a problem that is not discussed in the nonparametric estimation literature.

3. ASYMPTOTICS FOR THE DURBIN-WATSON STATISTIC BASED ON REGRESSION RESIDUALS

3.1. Regressions with independent regressors

We shall work with the linear regression model

$$(20) \quad y_t = c + b'x_t + u_t; \quad t = 1, 2, \dots$$

where under the null hypothesis u_t is iid symmetric with $u_t \in \mathcal{ND}(\alpha)$ when $0 < \alpha < 2$. We shall also assume that the sequence x_t is completely exogenous in the sense either that it is fixed or completely independent of the error sequence $\{u_t\}$.

We shall further assume that the parameters c and b in (20) are consistently estimated by \hat{c} and \hat{b} . When $\alpha > 1$ this may be achieved by least squares methods under rather weak conditions on x_t . Indeed, Kanter and Steiger (1974), Chen, Lai and Wei (1981), Cline (1983) and Andrews (1986) all demonstrate the consistency of the least squares slope coefficient estimator \hat{b} under a variety of conditions which ensure sufficient regressor variability. For example, when x_t is scalar, iid and lies in $\mathcal{ND}(r)$ with $r \leq \alpha$ then a minor variation of Kanter and Steiger's Lemma 4.3 ensures that $\hat{b} \xrightarrow{p} b$; and if $\alpha \geq r > 1$ then $\hat{c} \xrightarrow{p} c$ also, by the weak law of large numbers. Andrews (1986) extends this result for the slope coefficient to the multiple regressor case where each regressor may be distributed with separate characteristic exponents r_i ($i = 1, \dots, k$). He shows *inter alia* that consistency continues to hold even when independence between x_t and u_t is relaxed provided $r = \max_i(r_i) < 2$ when $\alpha > 1$. However, stochastic dependence between x_t and u_t does affect the limit distribution theory (and, hence, that of residual diagnostic tests) even though consistency is unaffected. When $\alpha \leq 1$ consistent estimates of the coefficients in (20) may be obtained by various robust methods such as bounded

influence estimators or the classical Huber M-estimator (Huber (1964)). The latter has been shown to perform well in simulations in the present context (Andrews (1986)).

The residuals from a fitted regression are written as

$$\hat{u}_t = y_t - \hat{c} - \hat{b}'x_t = u_t - (\hat{c}-c) - (\hat{b}-b)'x_t$$

and the DW statistic then has the form

$$(21) \quad DW = \Sigma_2^n(\hat{u}_t - \hat{u}_{t-1})^2 / \Sigma_1^n \hat{u}_t^2.$$

Since $(\hat{c}, \hat{b}) \xrightarrow{p} (c, b)$ and since x_t and u_t are independent we find that the sample variance and covariance of \hat{u}_t behave asymptotically like the corresponding quantities for u_t upon appropriate standardization. In fact as $n \rightarrow \infty$ we have

$$\left[a_n^{-2} \Sigma_1^n \hat{u}_t^2, \hat{a}_n^{-1} \Sigma_2^n \hat{u}_t \hat{u}_{t-1} \right] \sim \left[a_n^{-2} \Sigma_1^n u_t^2, \hat{a}_n^{-1} \Sigma_2^n u_t u_{t-1} \right] \Rightarrow (S_0, S_1).$$

This leads us directly to the conclusion that the DW statistic (21) based on regression residuals is asymptotically equivalent to the von Neumann ratio based on the regression errors from (20). In particular, under the null for u_t we have

$$(1/2)(n/\ln(n))^{1/\alpha} (DW-2) \Rightarrow S_1/S_0.$$

In a similar way under the general time series alternative (8) we obtain

$$(1/2)(n/\ln(n))^{1/\alpha} (DW - 2(1 - \rho(1))) \Rightarrow \left[\sum_{j=1}^{\infty} |f_j|^\alpha \right]^{1/\alpha} S_1/S_0.$$

Thus, the conclusions of Theorem 1 apply equally well to the standardized DW statistic

$$d = (1/2)(n/\ln(n))^{1/\alpha} (DW-2).$$

In addition, when $\alpha = 2$ we find as usual that $d \Rightarrow N(0,1)$.

3.2. Regressions with lagged dependent variables

In models with a finite error variance the asymptotic theory for the DW statistic changes when lagged dependent variables enter the regressor set. Use of the correct asymptotics leads, of course, to the LM and h statistics. The purpose of this section is to examine how the asymptotics are affected in the infinite variance case by the presence of lagged dependent variables. We note that recent arguments in terms of power properties and small-sigma asymptotic behavior by Inder (1984a, b) and King and Wu (1989) have been advanced to justify the use of the DW statistic in dynamic regressions. It seems likely that these arguments could be extended (but with uncertain outcome) to accommodate infinite variance errors. This would be of interest but it is not our intention here.

Let us suppose that our model has the form

$$(22) \quad y_t = \gamma y_{t-1} + \beta' x_t + u_t, \quad |\gamma| < 1$$

where u_t and x_t are as in the previous section. We shall estimate (22) by least squares giving $\hat{\gamma}$, $\hat{\beta}$ and suppose these estimates are consistent. This is a relatively innocuous requirement since least squares is known to deliver strongly consistent estimates in autoregressions with infinite variance errors (see Hannan and Kanter (1978)).

To fix ideas we shall again assume that under the null hypothesis u_t is iid symmetric and in $\mathcal{ND}(\alpha)$ when $0 < \alpha < 2$. The regressor x_t in (22) will be taken to be scalar for convenience (although extensions to multiple regressors are relatively straightforward) and to be generated by a general MA process of the form

$$(23) \quad x_t = \sum_{j=0}^{\infty} g_j e_{t-j}, \quad g_0 = 1.$$

Here the e_t are iid symmetric with $e_t \in \mathcal{ND}(r)$ when $0 < r < 2$. The coefficients g_j in (23) are assumed to be majorized by geometrically declining weights. As discussed in Section 2 this implies the distributional equivalence

$$x_t \equiv \left[\sum_{j=0}^{\infty} |g_j|^r \right]^{1/r} e_1$$

so that the tails of the distribution of x_t are Pareto with index r like those of the innovations e_t when $0 < r < 2$. We also assume that $\{e_t\}$ and $\{u_t\}$ are completely independent and thus x_t is strictly exogenous. Distributional results may be obtained without imposing this exogeneity condition but they will, in general, be different from those given below. We shall write the densities of x_t and u_t as $h(x)$ and $f(u)$, respectively, and their tails as

$$h(x) = (1/2)rb^r|x|^{-r-1}, \quad |x| \rightarrow \infty$$

$$f(u) = (1/2)\alpha a^\alpha|u|^{-\alpha-1}, \quad |u| \rightarrow \infty$$

when $0 < r, \alpha < 2$.

We shall start our development of the theory with a preliminary but useful result on the tail behavior of the product $X = xu$. It is helpful to employ the following index condition on r and α :

$$(C1) \quad \begin{aligned} r &< 2\alpha \text{ when } \alpha < 1 \\ r &< 2 \text{ when } \alpha \geq 1, \end{aligned}$$

Under (C1) it can be shown, as in Andrews (1986), that the least squares estimates $\hat{\gamma}$, $\hat{\beta}$ are consistent. Moreover, from the Appendix of Phillips (1988), we have

$$(24) \quad X \in \mathcal{M}(q)$$

where $q = \alpha \wedge r = \min(\alpha, r)$ when $\alpha \neq r$ and $X \in \mathcal{D}(\alpha)$ when $\alpha = r$. Both cases are therefore included by the general statement that $X \in \mathcal{D}(q)$.

We write the residuals from a least squares regression on (22) as

$$\hat{u}_t = u_t - (\hat{\gamma} - \gamma)y_{t-1} - (\hat{\beta} - \beta)x_t$$

where

$$\hat{\beta} - \beta = \left\{ \mathbf{x}'\mathbf{x} - \mathbf{x}'\mathbf{y}_{-1}(\mathbf{y}'_{-1}\mathbf{y}_{-1})^{-1}\mathbf{y}'_{-1}\mathbf{x} \right\}^{-1} \{ \mathbf{x}'\mathbf{u} - \mathbf{x}'\mathbf{y}_{-1}(\mathbf{y}'_{-1}\mathbf{y}_{-1})^{-1}\mathbf{y}'_{-1}\mathbf{u} \}$$

$$\hat{\gamma} - \gamma = \left\{ \mathbf{y}'_{-1}\mathbf{y}_{-1} - \mathbf{y}'_{-1}\mathbf{x}(\mathbf{x}'\mathbf{x})^{-1}\mathbf{x}'\mathbf{y}_{-1} \right\}^{-1} \{ \mathbf{y}'_{-1}\mathbf{u} - \mathbf{y}'_{-1}\mathbf{x}(\mathbf{x}'\mathbf{x})^{-1}\mathbf{x}'\mathbf{u} \}$$

in usual regression notation. Also

$$(25) \quad \hat{\mathbf{u}}'\hat{\mathbf{u}}_{-1} = \mathbf{u}'\mathbf{u}_{-1} - (\hat{\gamma} - \gamma)\mathbf{y}'_{-1}\mathbf{u}_{-1} - (\hat{\beta} - \beta)\mathbf{x}'\mathbf{u}_{-1}$$

$$- (\hat{\gamma} - \gamma)\mathbf{y}'_{-2}\mathbf{u} - (\hat{\beta} - \beta)\mathbf{x}'_{-1}\mathbf{u} + (\hat{\gamma} - \gamma)^2\mathbf{y}'_{-1}\mathbf{y}_{-2} + (\hat{\beta} - \beta)^2\mathbf{x}'\mathbf{x}_{-1}$$

$$+ (\hat{\beta} - \beta)(\hat{\gamma} - \gamma)(\mathbf{x}'\mathbf{y}_{-2} + \mathbf{x}'_{-1}\mathbf{y}_{-1}).$$

The remainder of the discussion will be carried out in three distinct cases.

Case 1: $r < \alpha$. Here the tail behavior of x_t dominates that of u_t and we have $x_t y_{t-1}$, y_{t-1}^2 , $x_t^2 \in \mathcal{ND}(r/2)$, and $x_t u_t$, $x_t u_{t-1}$, $y_{t-1} u_t \in \mathcal{D}(r)$. This leads to

$$(n/\ln(n))^{1/r}(\hat{\beta} - \beta), (n/\ln(n))^{1/r}(\hat{\gamma} - \gamma) = O_p(1)$$

and

$$(n\ln(n))^{-1/\alpha}(\hat{\beta} - \beta)\mathbf{x}'\mathbf{u}_{-1} = [(n/\ln(n))^{1/r}(\hat{\beta} - \beta)][(n\ln(n))^{-1/r}\mathbf{x}'\mathbf{u}_{-1}]$$

$$\cdot [(\ln(n))^{2/r}(n\ln(n))^{-1/\alpha}] = o_p(1).$$

In a similar way we see that upon standardization by $(n/\ln(n))^{-1/\alpha}$ the fourth to the ninth terms on the right of (25) are all $o_p(1)$ as $n \rightarrow \infty$. Turning to the second term we note that $y_{t-1} u_{t-1} \in \mathcal{ND}(\alpha/2)$ if $\alpha/2 < r$ and $\in \mathcal{D}(r)$ if $r \leq \alpha/2$. Writing $p = r \wedge \alpha/2$ we then have

$$\begin{aligned}
(n \ln(n))^{-1/\alpha} (\hat{\gamma} - \gamma) y'_{-1} u_{-1} &= [(n / \ln(n))^{1/r} (\hat{\gamma} - \gamma)] [(n \ln(n))^{-1/p} y'_{-1} u_{-1}] \\
&\quad \cdot [(\ln(n))^{1/r+1/p} n^{1/p-1/r} (n \ln(n))^{-1/\alpha}] \\
&= o_p(1)
\end{aligned}$$

since $r < \alpha$ by hypothesis in this case. It now follows that

$$a^{-2} (n \ln(n))^{-1/\alpha} \hat{u}' \hat{u}_{-1} = a^{-2} (n \ln(n))^{-1/\alpha} u' u_{-1} + o_p(1) \Rightarrow S_1.$$

Similarly

$$a^{-2} n^{-2/\alpha} \hat{u}' \hat{u} = a^{-2} n^{-2/\alpha} u' u + o_p(1) \Rightarrow S_0.$$

Thus

$$(26) \quad (n / \ln(n))^{1/\alpha} (DW-2) \Rightarrow -2S_1/S_0$$

as in the case of regression with no lagged regressors.

The heuristic explanation behind (26) is straightforward. When $r < \alpha$ the independent regressors x_t dominate in determining the tail behavior of the lagged regressor y_{t-1} in (22). Residuals from this regression then behave like the residuals from a conventional static linear regression, leading to (26).

Case 2: $r = \alpha$. In this case we proceed under the simplifying assumption that x_t is iid, although this can be relaxed at the cost of extra notational complexity. Note that $y_t = \beta Z_t + U_t$ where $Z_t = \sum_{j=0}^{\infty} \gamma^j x_{t-j}$, $U_t = \sum_{j=0}^{\infty} \gamma^j u_{t-j}$. Then

$$\begin{aligned}
\frac{1}{a^2 n^{2/\alpha}} \sum_1^n y_t^2 &= \frac{\beta^2 b^2}{a^2} \frac{1}{b^2 n^{2/\alpha}} \sum_1^n Z_t^2 + \frac{1}{a^2 n^{2/\alpha}} \sum_1^n U_t^2 + o_p(1) \\
&\Rightarrow \frac{\beta^2 b^2}{a^2} (\sum_0^\infty \gamma^{2j}) S_{01} + (\sum_0^\infty \gamma^{2j}) S_{02} \\
&\equiv [1 - \gamma^2]^{-1} \{ |\beta b/a|^2 + 1 \}^{2/\alpha} S_0
\end{aligned}$$

where S_{01} and S_{02} are independent copies of S_0 . In a similar way we find

$$\begin{aligned}
 a^{-2(n\ell n(n))^{-1/\alpha} y'_{-1} u} &= \frac{(\beta b a)}{b a (n\ell n(n))^{1/\alpha}} \Sigma_1^n Z_{t-1} u_t + \frac{1}{a^2 (n\ell n(n))^{1/\alpha}} \Sigma_1^n U_{t-1} u_t \\
 &\Rightarrow (\beta b/a) (\Sigma_0^\alpha \gamma^j S_{1j}) + (\Sigma_0^\alpha \gamma^j) S_{2j} \\
 (27) \quad &\equiv \left[|(\beta b/a)| \alpha \Sigma_0^\alpha |\gamma|^{j\alpha} + \Sigma_0^\alpha |\gamma|^{j\alpha} \right]^{1/\alpha} S_1 \\
 &= \left[1 - |\gamma|^\alpha \right]^{-1/\alpha} \left\{ |\beta b/a|^\alpha + 1 \right\}^{1/\alpha} S_1
 \end{aligned}$$

where $\{S_{1j}, S_{2j} : j = 0, 1, \dots\}$ are independent copies of S_1 .

Next we observe that, since x_t is serially independent, $x_t y_{t-1} \in \mathcal{P}(\alpha)$ while $x_t^2 \in \mathcal{ND}(\alpha/2)$ and thus

$$\begin{aligned}
 (n/\ell n(n))^{1/\alpha} (\hat{\gamma} - \gamma) &\sim \left[a^{-2n} n^{-2/\alpha} y'_{-1} y_{-1} \right]^{-1} \left[a^{-2(n\ell n(n))^{-1/\alpha} y'_{-1} u} \right] \\
 (28) \quad &\Rightarrow (1-\gamma^2) \left[1 - |\gamma|^\alpha \right]^{-1/\alpha} \left\{ |\beta b/a|^\alpha + 1 \right\}^{1/\alpha} \left\{ (\beta^2 b^2/a^2)^{\alpha/2} + 1 \right\}^{-2/\alpha} (S_1/S_0) \\
 &= (1-\gamma^2) \left[1 - |\gamma|^\alpha \right]^{-1/\alpha} \left\{ |\beta b/a|^\alpha + 1 \right\}^{1/\alpha} (S_1/S_0) \\
 &= G, \text{ say.}
 \end{aligned}$$

Further

$$a^{-2n} n^{-2/\alpha} \Sigma_1^n y_{t-1} u_{t-1} \Rightarrow P_0.$$

Here P_0 is a copy of S_0 and is not independent of S_0 , at least as it occurs in the limit (28).

Thus

$$\begin{aligned}
& a^{-2}(n\ln(n))^{-1/\alpha}(\hat{\gamma}-\gamma)y'_{-1}u_{-1} = [(n/\ln(n))^{1/\alpha}(\hat{\gamma}-\gamma)](a^{-2}n^{-2/\alpha}y'_{-1}u_{-1}) \\
& \Rightarrow (1-\gamma^2)\left[1-|\gamma|^\alpha\right]^{-1/\alpha}\left\{|\beta b/a|^\alpha+1\right\}^{-1/\alpha}(S_1/S_0)P_0.
\end{aligned}$$

Also

$$\begin{aligned}
a^{-2}(n\ln(n))^{-1/\alpha}(\hat{\beta}-\beta)x'u_{-1} &= a^{-2}[(n/\ln(n))^{1/\alpha}(\hat{\beta}-\beta)] \\
&\quad \cdot [(n\ln(n))^{-1/\alpha}x'u_{-1}][(\ln(n))^{2/\alpha}n^{-1/\alpha}] \\
&= o_p(1)
\end{aligned}$$

and the remaining terms of (25) are similarly seen to be $o_p(1)$ when standardized by $(n\ln(n))^{-1/\alpha}$.

Finally we observe that

$$a^{-2}(n\ln(n))^{-1/\alpha}u'u_{-1} \Rightarrow P_1$$

and

$$a^{-2}n^{-2/\alpha}u'u \Rightarrow P_0.$$

We deduce that

$$(29) \quad (n/\ln(n))^{1/\alpha}(\text{DW}-2) \Rightarrow -2[P_1/P_0 - G] = H, \text{ say.}$$

Thus when $r = \alpha$ we see that the effect of the lagged regressor in (22) is to alter the limiting distribution of the DW statistic. This corresponds with known theory in the finite variance case (see Durbin (1970) and Phillips and Wickens (1978), solution 7.5, pp. 423–429) where the limit distribution of the DW statistic is normal but has a variance that depends on the variance of the limit distribution of $\hat{\gamma}$, in contrast to the static regression model. We shall now see that this finite variance case is a specialization of the general limit theory when $r = \alpha$.

Subcase: $r = \alpha = 2$. When $r = \alpha = 2$ both u_t and x_t have finite variance (since $u_t, x_t \in \mathcal{ND}(2)$). The scale factor is then $n^{1/2}$ in (28) rather than $(n/\ln(n))^{1/\alpha}$. Next we observe that $P_0 = 1$ and

$$-2P_1/P_0 \equiv N(0,4)$$

G is also normal in (29) when $\alpha = 2$ but it is dependent on P_1 . To see this we note from (28) that G has the following more explicit representation in the general case

$$(30) \quad G = (1-\gamma^2)\{(\beta b/a)(\sum_0^\infty \gamma^j S_{1j}) + (\sum_0^\infty \gamma^j S_{2j})\} \left\{ |\beta b/a|^\alpha + 1 \right\}^{-2/\alpha} / S_0$$

where $S_{20} = P_1$. When $u_t, x_t \in \mathcal{ND}(2)$ we may write the limit variate H of (29) as

$$\begin{aligned} H &= -2P_1 + 2(1-\gamma^2)\{(\beta b/a)(\sum_0^\infty \gamma^j S_{1j}) + (\sum_0^\infty \gamma^j S_{2j})\} / (\beta^2 b^2 / a^2 + 1) \\ &= -2[P_1 - P_2], \text{ say.} \end{aligned}$$

Now both P_1 and P_2 are normal and since $S_{20} = P_1$ we have

$$\begin{aligned} \text{cov}(P_1, P_2) &= \text{cov}(P_1, P_1(1-\gamma^2)/(1+\beta^2 b^2/a^2)) \\ &= a^2(1-\gamma^2)/(a^2 + \beta^2 b^2). \end{aligned}$$

Noting that

$$\text{var}(P_2) = (1-\gamma^2)a^2/(a^2 + \beta^2 b^2)$$

we deduce that

$$\begin{aligned} P_1 - P_2 &\equiv N(0, 1 - 2\text{cov}(P_1, P_2) + \text{var}(P_2)) \\ &= N(0, 1 - a^2(1-\gamma^2)/(a^2 + \beta^2 b^2)). \end{aligned}$$

Since the variances of x_t and u_t are finite we may write

$$a^2 = \text{var}(u_t) = \sigma_u^2$$

$$b^2 = \text{var}(x_t) = \sigma_x^2$$

and since x_t is iid we find from the usual formula (cf. Phillips and Wickens (1978), p. 427) that the variance of the limiting distribution of $\sqrt{n}(\hat{\gamma} - \gamma)$ is

$$(31) \quad V_\gamma = \sigma_u^2(1 - \gamma^2)/(\sigma_u^2 + \beta^2\sigma_x^2).$$

Thus the limit variate H in the finite variance ($\alpha = 2$) case is simply

$$H \equiv N(0, 4(1 - V_\gamma))$$

i.e. we have

$$(32) \quad \sqrt{n}(\text{DW} - 2) \Rightarrow N(0, 4(1 - V_\gamma))$$

leading as usual to the Durbin h statistic

$$h = \sqrt{n} r(1)/(1 - \hat{V}_\gamma)^{1/2}, \quad r(1) = \Sigma \hat{u}_t \hat{u}_{t-1} / \Sigma \hat{u}_t^2$$

and the conventional asymptotics:

$$h \Rightarrow N(0, 1),$$

which are the same as that of the LM version of this test.

Case 3: $r > \alpha$. This case may be handled in just the same way, so we present only a sketch of the results here. Since the tails of u_t rather than x_t dominate the behavior of y_t we find

$$\begin{aligned} (n/\ln(n))^{1/\alpha}(\hat{\gamma} - \gamma) &\sim \left[a^{-2} n^{-2/\alpha} y'_{-1} y_{-1} \right]^{-1} \left[a^{-2} (n \ln(n))^{-1/\alpha} y'_{-1} u \right] \\ &\Rightarrow (1 - \gamma^2) \left[1 - |\gamma|^\alpha \right]^{-1/\alpha} \left\{ |\beta b/a|^\alpha + 1 \right\}^{1/\alpha} (S_1/S_0) \end{aligned}$$

and

$$a^{-2}n^{-2/\alpha}y'_{-1}u_{-1} \Rightarrow P_0$$

$$a^{-2}(n\ln(n))^{-1/\alpha}u'_{-1}u_{-1} \Rightarrow P_1$$

$$a^{-2}n^{-2/\alpha}u'_{-1}u_{-1} \Rightarrow P_0$$

as before. This leads to

$$(33) \quad (n/\ln(n))^{1/\alpha}(DW - 2) \Rightarrow -2[P_1/P_0 - G'] = H'$$

where

$$(34) \quad G' = (1 - \gamma^2) \left\{ (\beta b/a) \sum_0^\infty \gamma^j S_{1j} + \sum_0^\infty \gamma^j S_{2j} \right\} / S_0$$

and, as before, $S_{20} = P_1$ and $S_0 = P_0$.

Observe that the limit distribution (33) is different from that of (29). This is because the tails of u_t dominate the behavior of sample moments of y_t when $r > \alpha$, whereas the tails of both x_t and u_t are influential when $r = \alpha$. Thus we have

$$(35) \quad a^{-2}n^{-2/\alpha}y'_{-1}y_{-1} \Rightarrow \begin{cases} [1 - \gamma^2]^{-1} S_0 & , r > \alpha \\ [1 - \gamma^2]^{-1} \left\{ |\beta b/a|^\alpha + 1 \right\}^{2/\alpha} S_0 & , r = \alpha \end{cases}$$

This implies that the limit variates G and G' in (29) and (33) are different. However, (35) suggests the distributional equivalence

$$G \equiv kG' , \quad k = \left\{ |\beta b/a|^\alpha + 1 \right\}^{-2/\alpha} .$$

Since $0 < k < 1$ we anticipate the distribution of G (and hence H in (29)) to be more

concentrated than that of G' (and hence H' in (33)). The Monte Carlo results of the following section corroborate this feature of the asymptotic distribution.

The results of these three cases may now be summarized as follows.

THEOREM 2. *The Durbin–Watson statistic DW computed from the residuals of a least squares regression on the dynamic model (22) with iid regressors $x_t \in \mathcal{ND}(\tau)$ has the limit distribution*

$$(1/2)(n/\ln(n))^{1/\alpha}(DW-2) \Rightarrow \begin{cases} S_1/S_0 & \tau < \alpha \\ P_1/P_0 - G & \tau = \alpha \\ P_1/P_0 - G' & \tau > \alpha \end{cases}$$

for $0 < \alpha < 2$ and $\tau < 2\alpha$. The limit variates (S_0, S_1) are as in (4), (P_0, P_1) are independent copies of (S_0, S_1) and G and G' are as in (30) and (34) respectively.

When $\alpha = \tau = 2$ the limit theory is

$$(1/2)n^{1/2}(DW-2) \Rightarrow P_1/P_0 - G \equiv N(0, 1 - V_\gamma)$$

where V_γ is the variance of the limit distribution of $\sqrt{n}(\hat{\gamma} - \gamma)$ and is given in (31).

4. SOME SIMULATION EVIDENCE

In Sections 2 and 3, we gave the limiting distributions of the v and d statistics in nonregression, regression and dynamic regression models. We now complement these results by investigating the distributional shapes that the statistics possess in large and small samples.

Since closed form expressions for the distributions are not available, we resorted to simulation. We used the Kanter–Steiger (1974) algorithm to generate symmetric standard stable random numbers of index α , for $\alpha < 2$, $\alpha \neq 1$. For $\alpha = 2$, we drew standard normal random numbers. For $\alpha = 1$, we generated Cauchy random numbers by the

inverse distribution function method; i.e. if x is uniform on $(-\pi/2, \pi/2)$, then $\tan(x)$ is standard Cauchy. We then calculated the v and d statistics and used a nonparametric density estimate with a standard normal kernel (cf. Silverman (1986)) to deliver the pdf's. In all simulations reported below, the number of iterations was 20,000 for the v statistics and 10,000 for the d statistics. We selected $n = 1000$ as the "large" sample size, and $n = 20, 50,$ and 100 as the "small" sample sizes. Values of α chosen were $\alpha = 0.5, 1.0, 1.5$ and 2.0 . All regressions were run without a fitted intercept. As discussed earlier this could be modified in the static case by using robust regression methods and fitting an intercept to obtain consistent estimates of the residuals (especially when $\alpha \leq 1$).

4.1. Large sample distribution of v and d

We first investigate the distribution of the standardized von Neumann ratio

$$v = \begin{cases} 1/2(n/\ln(n))^{1/\alpha}(VN - 2), & \alpha < 2, \\ 1/2\sqrt{n}(VN - 2), & \alpha = 2. \end{cases}$$

Figure 1 graphs the density of v for the four values of α , as well as the pdf of the standard normal distribution for reference. We notice that the distributions are leptokurtic for $\alpha < 2$, and that the kurtosis increases as $\alpha \downarrow 0$. On the other hand, the pdf of v in the $\alpha = 2$ case is very close to the pdf of the $N(0,1)$ distribution. To complement the visual information of figure 1, we give the quantiles of the distributions in Table 1a and the probabilities of v being close to zero in Table 1b. E.g., for $\alpha = 0.5$, Table 1a tells us that the distribution of v is both leptokurtic and has very heavy tails. As α increases towards 2, the kurtosis decreases and the tails of the distributions become thinner. In fact, as we see from Table 1b, the probability that $|v| < 1$ is greater when $\alpha = 1.5$ than when $\alpha = 2$. (This is due to the fact that the scale factor $(n/\ln(n))^{1/\alpha} < \sqrt{n}$ as $\alpha \uparrow 2$.) We also notice from table 1a that the distributions are asymmetric: the left tail is heavier

than the right tail. This contrasts with our result in Section 2 that v has a symmetric distribution asymptotically.

We turn to the large sample distribution of d , where d is based on the residuals from the regression model

$$y_t = x_t + u_t .$$

We are interested in how closely the distributions of d coincides with the corresponding distributions of v . Recall our result from Section 3.1: d and v have the same limit distribution. The pdf's of d for $\alpha = 0.5, 1.0, 1.5$ and 2.0 are graphed in Figure 2, quantiles of the distribution are given in Table 2a, and the central probabilities are supplied in Table 2b. Overall, we see that the distributional properties of d are very close to those of v : the distributions are leptokurtic, slightly asymmetric (the left tail being heavier than the right tail), and heavy-tailed (especially for $\alpha \leq 1$).

4.2. Small sample distributions of d

In studying the small sample behavior of the d statistic, we seek to determine in what respects these distributions differ from the corresponding large sample distributions presented above. We also test how well the DW bounds test of the null hypothesis of iid errors vs. the alternative of first order serially correlated errors performs when the errors are symmetric stable. We continue to work with the regression model

$$y_t = x_t + u_t ,$$

where $\{x_t\}$ and $\{u_t\}$ are independent iid series with the same stable exponent α . Since we do not include a constant term in the regression, we use the d_L and d_U values given in Table 1b of Kramer (1971, p. 351). For convenience, we present the values for d_L and d_U for sample sizes of 20, 50, and 100 at the 1% and 5% level of confidence (against the

one-sided alternative of positive first order serial correlation) in Table 3. Farebrother (1980) provides extensive tables for d_L only.

The distributions of d for $\alpha = 0.5$ are graphed in Figure 3. The main difference with the large sample behavior of d is that the distributions are less leptokurtic and become more asymmetric as n decreases. These differences apply to the cases of $\alpha = 1.0$ (Figure 4) and $\alpha = 1.5$ (Figure 5) as well; in particular, the left tails become fatter whereas the right tails do not change much compared with the corresponding large sample distributions.

In Table 4, panel a, we tabulate the relative frequencies of the statistic DW (not d) being less than d_L and d_U for $\alpha = 0.5$. In general, and allowing for Monte Carlo simulation error, the performance of the bounds tests is poor, and we get size distortions in almost all cases. Actually, for $n = 20$ and a test size of 1%, the probabilities of DW being less than d_L and d_U do "bracket" the 1% level, but with a very large difference in size (0.4% vs. 4.9%). For sample sizes of 50 and 100, the bounds tests give too large a size (in all cases greater than 2%, i.e. the 1% value is not bracketed). If we choose a size of 5%, we see that the actual size of the test is too large for $n = 20$, about correct for $n = 50$, and too small for $n = 100$.

Similar results obtain when we look at the performance of the bounds tests for the cases of $\alpha = 1.0$ (Table 4, panel b). For a size of the bounds test of 1%, the actual size is too large for $n = 50$ and $n = 100$; for $n = 20$ d_L and d_U bracket the correct size. For a size of 5% we get a more favorable picture: d_L and d_U bracket the correct size for $n = 20$ and 50, but the size is too small when $n = 100$.

The bounds tests perform quite well when $\alpha = 1.5$ (panel c of Table 4). For a test size of 5%, d_L and d_U bracket the stated size for all three sample sizes. For a test size of 1%, the bounds tests still yield too large a size for $n = 50$ and 100, but less so than when $\alpha = 0.5$ and 1.0. For $n = 20$, the bounds tests bracket the 1% level.

For completeness, we report the results of our simulations when $\alpha = 2.0$, i.e. when the errors are standard normal (panel d of Table 4). As we would expect, d_L and d_U bracket the true size for all three values of n . But since DW is less leptokurtic for $\alpha = 2$ compared with $\alpha < 2$, the bounds are quite wide. E.g. for $n = 20$ and a test size of 5%, the bounds give probabilities of 3.1% and 8.0%, respectively. This illustrates that one should, whenever possible, compute exact DW critical values when the variances are finite, rather than simply rely on the bounds test alone. Our results indicate that this conclusion holds a fortiori when the error variance is infinite.

4.3. Large sample distribution of d in dynamic regression models

Our final set of simulations investigates the large sample properties of the distribution of d in the dynamic regression model

$$y_t = \gamma y_{t-1} + x_t + u_t$$

where $|\gamma| < 1$, x_t is iid symmetric stable with index r , and u_t is iid symmetric stable with index α . As shown in Section 3.2, the limiting distributions of d and v are the same when $\alpha > r$, but are different when $\alpha \leq r$. In our simulations, we held r fixed at 1.5, while we set α equal to 1.1, 1.5 and 1.9. We considered the following values of $\gamma = \{0.9, 0.5\}$. As seen from the results of Section 3.2 the asymptotic distributions are invariant to the sign of γ , so it is unnecessary to consider negative values of γ .

For $\gamma = 0.9$, we have graphed the distributions of d for $\alpha = 1.5$ and 1.9 (as well as the corresponding v statistics) in Figure 6a. We notice that, for this large sample size, the distributions of d match the corresponding distributions of v closely. For $\alpha = 1.1$ (Figure 6b), on the other hand, the distribution of d is less leptokurtic and has more mass in the region $0.5 \leq |d| \leq 3.0$ than the corresponding distribution of v .

For a choice of γ equal to 0.5 (Figures 7a and 7b), we see that the distributions of d do not differ by much from those of v – even when $\alpha = 1.1$. We conclude that the

presence of a lagged dependent variable does not seem to affect the behavior of the regression diagnostics by much when the autoregressive coefficient is small, say $|\gamma| \leq .5$.

5. CONCLUSION

This paper studies the asymptotic and finite sample distributions of the Von Neumann and Durbin–Watson statistics in regression and dynamic regression models with infinite variance errors. Appropriately standardized, the Von Neumann ratio converges weakly to a simple limit given by a ratio of two stable random variables. The DW statistic has the same limiting distribution in regressions with strictly exogenous regressors. The results are more complex when lagged dependent variables are present. The limiting distribution then depends on the relative "importance" (as measured by their tail behavior) of the regressors and the error term.

Our simulation experiments show that the standardized statistic d is leptokurtic when $\alpha < 2$, and slightly asymmetric even when n is large. In smaller samples, the leptokurtosis is less pronounced, whereas the asymmetry increases. In general, the conventional DW bounds tests perform poorly and suffer major size distortions when $\alpha \leq 1.0$. Inclusion of a lagged dependent variable alters the limiting distributions when the errors have thicker tails than the regressors. However, our simulation evidence suggests that this effect will be important only if the coefficient of the lagged dependent variable is large, i.e. close to 1 in absolute value.

The analysis and simulation of this paper can be extended in many different ways. We have looked at the local power properties of tests based on the DW statistic in static regressions and this can be extended to dynamic regressions. Throughout we have assumed that the characteristic exponent α is known. If α has to be estimated consistently (together with the other parameters of the model), we would expect this to have important finite sample and asymptotic effects. Further, our simulations might be broadened to

include asymmetric stable variates and general infinite variance distributions that lie in the domain of attraction of a stable law, as well as regressions with multiple explanatory variables. Finally, there are many other regression diagnostics whose asymptotic behavior is known for the finite error variance case but is unexplored when the error variance is infinite. Some of these topics are now under investigation by the authors.

REFERENCES

- Anderson, T. W. (1971), *The Statistical Analysis of Time Series*. New York: John Wiley.
- Andrews, D. W. K. (1986), "On the Performance of Least Squares in Linear Regression with Undefined Error Means," Cowles Foundation Discussion Paper No. 798 (May).
- Bartels, R. and J. Goodhew (1981), "The Robustness of the Durbin–Watson Test," *Review of Economics and Statistics*, 63, 136–139.
- Brockwell, P. J. and R. A. Davis (1987), *Time Series: Theory and Methods*. New York: Springer.
- Chen, G. J., T. L. Lai, and C. Z. Wei (1987), "Convergence Systems and Strong Consistency of Least Squares Estimators in Regression Models," *Journal of Multivariate Analysis*, 11, 319–333.
- Cline, D. (1983), Estimation and Linear Prediction for Regression, Autoregression and ARMA with Infinite Variance Data. Ph.D. thesis, Department of Statistics, Colorado State University.
- Davis, R. A. and S. I. Resnick (1985a), "Limit Theory for Moving Averages of Random Variables with Regularly Varying Tail Probabilities," *Annals of Probability*, 13, 179–195.
- Davis, R. A. and S. I. Resnick (1985b), "More Limit Theory for Sample Correlation Functions of Moving Averages," *Stochastic Processes and their Applications*, 20, 257–279.
- Davis, R. A. and S. I. Resnick (1986), "Limit Theory for the Sample Covariance and Correlation Functions of Moving Averages," *Annals of Statistics*, 14, 533–558.
- Durbin, J. and G. S. Watson (1971), "Testing for Serial Correlation in Least Squares Regression III," *Biometrika*, 58, 1–19.
- Farebrother, R. W. (1980), "The Durbin–Watson Test for Serial Correlation Where There Is No Intercept in the Regression," *Econometrica*, 48, 1553–1563.
- Fisher, R. A. (1925), "Applications of Student's Distribution," *Metron*, 5, 90–104.
- Hannan, E. J. and M. Kanter (1978), "Autoregressive Processes with Infinite Variance," *Journal of Applied Probability*, 14, 411–415.
- Huber, P. J. (1964), "Robust Estimation of a Location Parameter," *Annals of Mathematical Statistics*, 35, 73–101.
- Ibragimov, I. A. and Y. V. Linnik (1971), *Independent and Stationary Sequences of Random Variables*. Groningen: Wolters–Noordhof.
- Inder, B. A. (1985a), "Finite Sample Power of Tests for Autocorrelation in Models Containing Lagged Dependent Variables," *Economics Letters*, 14, 179–185.
- Inder, B. A. (1985b), "Erratum," *Economics Letters*, 16, 401–402.

- Kanter, M. and W. L. Steiger (1974), "Regression and Autoregression with Infinite Variance," *Advances in Applied Probability*, 6, 768–783.
- Kariya T. (1977), "A Robustness Property of the Tests for Serial Correlation," *Annals of Statistics*, 5, 1212–1220.
- Kariya, T. (1980), "Locally Robust Tests for Serial Correlation in Least Squares Regressions," *Annals of Statistics*, 8, 1065–1070.
- Kariya, T. (1988), "The Class of Models for which the Durbin–Watson Test is Locally Optimal," *International Economic Review*, 29, 167–176.
- King, M. L. (1980), "Robust Tests for Spherical Symmetry and Their Application to Least Squares Regression," *Annals of Statistics*, 1265–1271.
- King, M. L. (1987), "Testing for Autocorrelation in Linear Regression Models: A Survey." In: M. L. King and D. E. A. Evans (eds.), *Specification Analysis in the Linear Model: Essays in Honour of Donald Cochrane*. New York: Routledge, Chapman & Hall.
- King, M. L. and M. A. Evans (1988), "Locally optimal properties of the Durbin–Watson Test," *Econometric Theory*, 4, 509–516.
- King, M. L. and G. H. Hillier (1985), "Locally Best Invariant Tests of the Error Covariance Matrix of the Linear Regression Model," *Journal of the Royal Statistical Society*, B47, 98–102.
- King, M. L. and P. X. Wu (1989), "Small-disturbance Asymptotics and the Durbin–Watson and Related Tests in the Dynamic Regression Model," mimeo, Monash University.
- Kramer, G. (1971), "On the Durbin–Watson Bounds Test in the Case of Regression through the Origin," *Jahrbücher für Nationalökonomie und Statistik*, 185, 345–358.
- Logan, B. F., C. L. Mallows, S. O. Rice and L. A. Shepp (1973), "Limit Distributions of Self-normalized Sums," *Annals of Probability*, 1, 788–809.
- Phillips, P. C. B. (1988), "Time Series Regression with a Unit Root and Infinite Variance Errors," Cowles Foundation Discussion Paper No. (December).
- Phillips, P. C. B. and V. A. Hajivassiliou (1987), "Bimodal t-Ratios," Cowles Foundation Discussion Paper No. 842, July.
- Phillips, P. C. B. and M. R. Wickens (1978), *Exercises in Econometrics*. Oxford: Philip Allan.
- Resnick, S. I. (1986), "Point Processes Regular Variation and Weak Convergence," *Advances in Applied Probability*, 18, 66–138.
- Silverman, B. W. (1986), *Density Estimation*. London: Chapman Hall.

TABLE 1
Large Sample Distribution of v

(a) Quantiles

	α					N(0,1)
	0.5	1.0	1.5	2.0		
1%	-503.8	10.9	-2.23	-2.35		-2.326
5%	-35.8	-3.23	-1.18	-1.69		-1.645
10%	8.04	-1.66	-0.836	-1.33		-1.282
40%	-0.0024	-0.039	-0.062	-0.27		-0.253
60%	0.0009	0.019	0.041	0.25		0.253
90%	7.09	1.50	0.792	1.24		1.282
95%	2.90	2.82	1.14	1.61		1.645
99%	341.8	8.69	1.99	2.33		2.326

(b) Central Probabilities: $P(|v| \leq x)$

	α					N(0,1)
	0.5	1.0	1.5	2.0		
0.25	0.448	0.390	0.382	0.195		0.197
0.5	0.553	0.542	0.626	0.387		0.383
1.0	0.599	0.703	0.863	0.684		0.683

TABLE 2
Large Sample Distribution of d

(a) Quantiles

	α					N(0,1)
	0.5	1.0	1.5	2.0		
1%	-438.4	-10.95	-2.20	-2.32		-2.326
5%	-34.0	-3.24	-1.17	-1.70		-1.645
10%	-7.74	-1.70	-0.84	-1.34		-1.282
40%	-0.0183	-0.0899	-0.131	-0.29		-0.253
60%	0.0103	0.0658	0.102	0.21		0.253
90%	7.30	1.50	0.79	1.25		1.282
95%	27.7	2.79	1.12	1.61		1.645
99%	314.0	8.15	2.00	2.29		2.326

(b) Central Probabilities: $P(|d| \leq x)$

	α					N(0,1)
	0.5	1.0	1.5	2.0		
0.25	0.441	0.385	0.376	0.199		0.197
0.5	0.519	0.534	0.622	0.391		0.383
1.0	0.594	0.700	0.867	0.685		0.683

TABLE 3
 Lower Bound (d_L) and Upper Bound (d_U)
 Significance Points of the Durbin-Watson Bounds Test,
 One Regressor, No Intercept Term

n	1%		5%	
	d_L	d_U	d_L	d_U
20	0.867	1.041	1.108	1.300
50	1.287	1.363	1.464	1.544
100	1.503	1.542	1.634	1.674

Source: Kramer (1971), Table 1b, p. 351.

TABLE 4
 Frequencies of $DW < d_L$ and $DW < d_U$

(a) $\alpha = 0.5$

n	1%		5%	
	$P(DW < d_L)$	$P(DW < d_U)$	$P(DW < d_L)$	$P(DW < d_U)$
20	.0044	.0493	.0637	.0949
50	.0346	.0385	.0442	.0507
100	.0238	.0258	.0296	.0314

(b) $\alpha = 1.0$

n	1%		5%	
	$P(DW < d_L)$	$P(DW < d_U)$	$P(DW < d_L)$	$P(DW < d_U)$
20	.0037	.0245	.0381	.0802
50	.0229	.0314	.0451	.0590
100	.0224	.0251	.0368	.0446

(c) $\alpha = 1.5$

n	1%		5%	
	$P(DW < d_L)$	$P(DW < d_U)$	$P(DW < d_L)$	$P(DW < d_U)$
20	.0039	.0214	.0323	.0780
50	.0137	.0213	.0381	.0620
100	.0135	.0177	.0376	.0523

(d) $\alpha = 2.0$

n	1%		5%	
	$P(DW < d_L)$	$P(DW < d_U)$	$P(DW < d_L)$	$P(DW < d_U)$
20	.0051	.0197	.0310	.0801
50	.0074	.0147	.0351	.0658
100	.0064	.0122	.0396	.0585

Fig. 1: Large-sample distribution of v

$$v = \frac{1}{2}(n/\ln(n))^{1/\alpha}(VN-2), \quad \alpha < 2$$

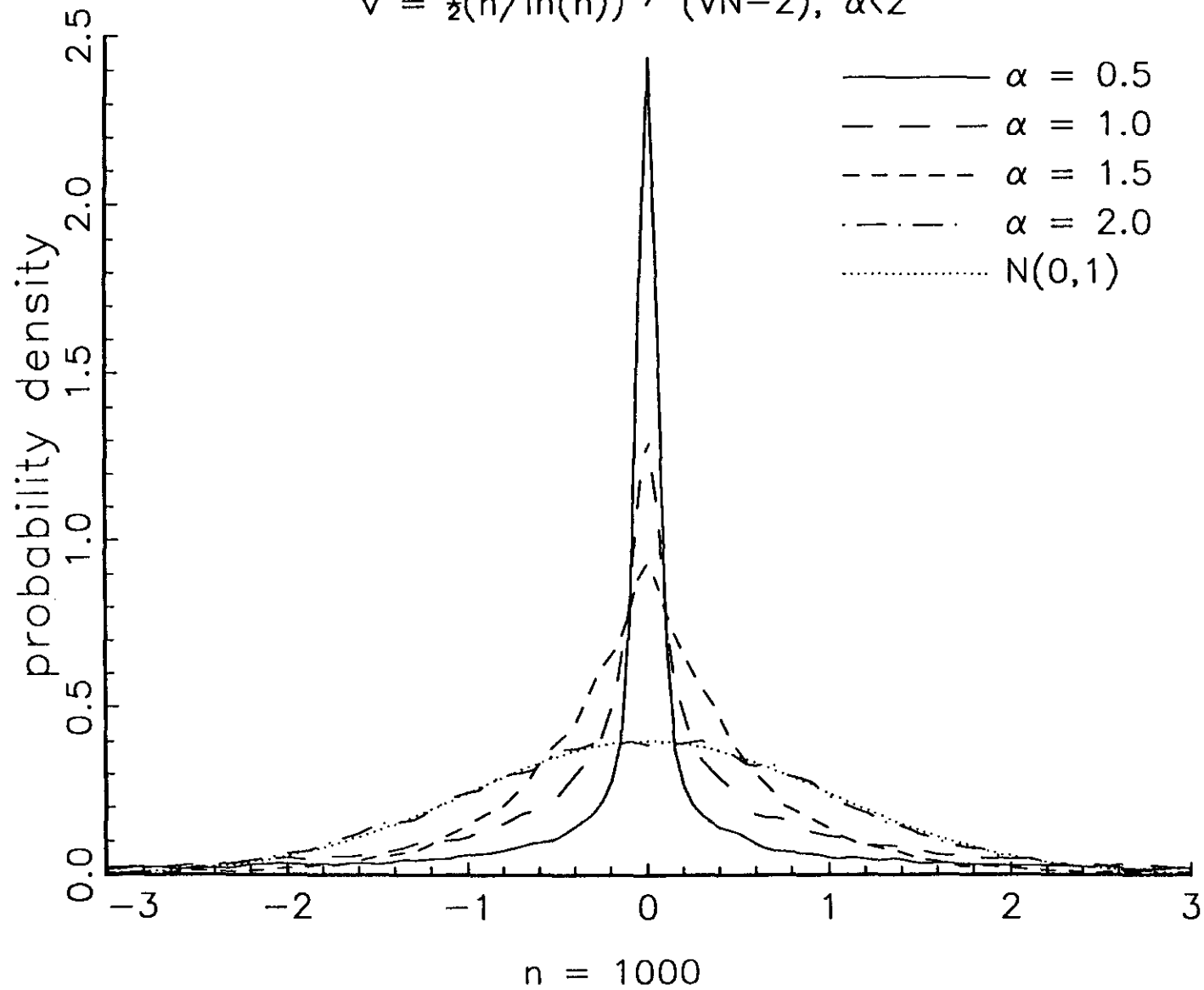


Fig. 2: Large-sample distribution of d

$$d = \frac{1}{2}(n/\ln(n))^{1/\alpha}(DW-2), \quad \alpha < 2$$

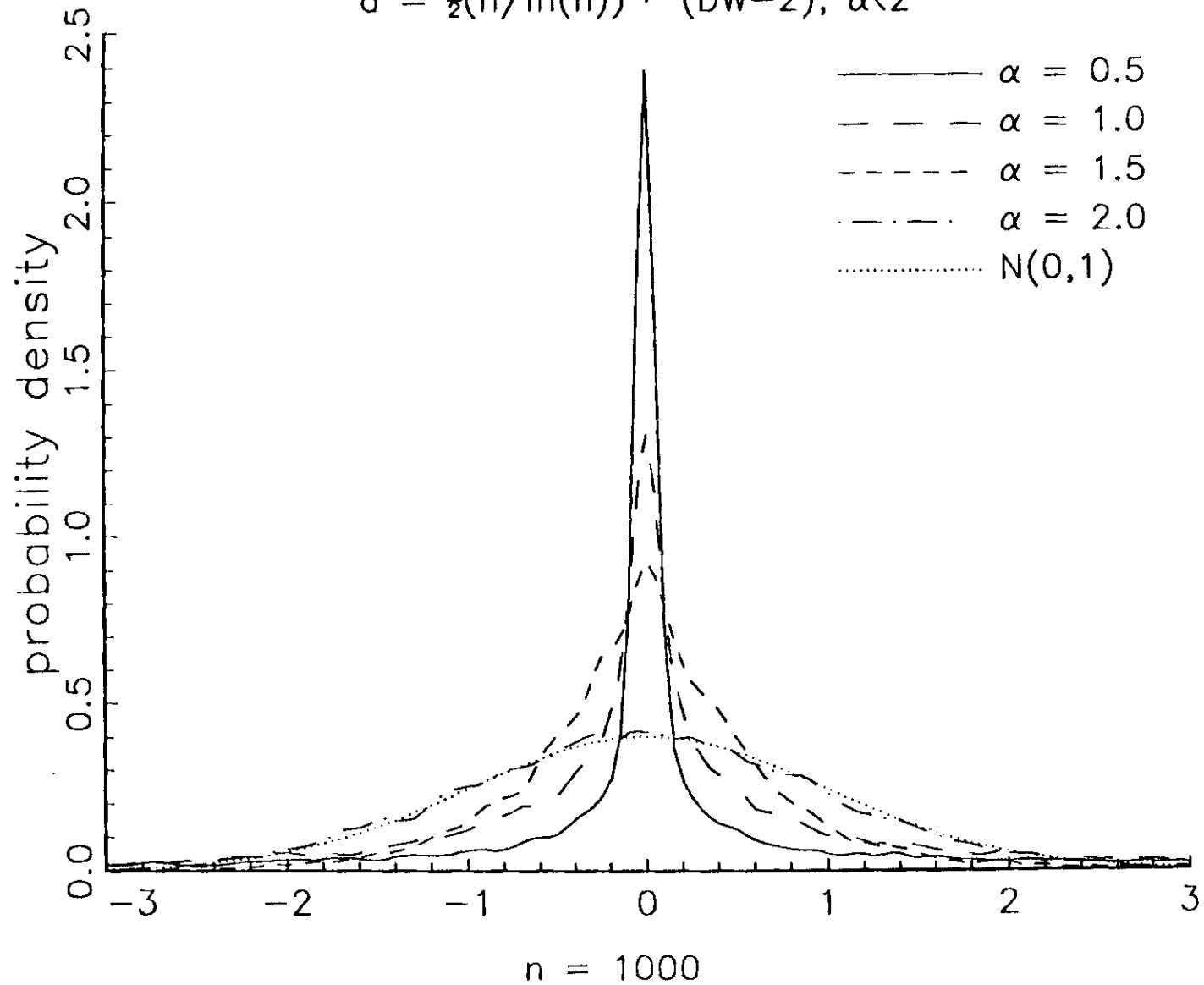


Fig. 3: Small-sample distribution of d
[$\alpha=0.5$]

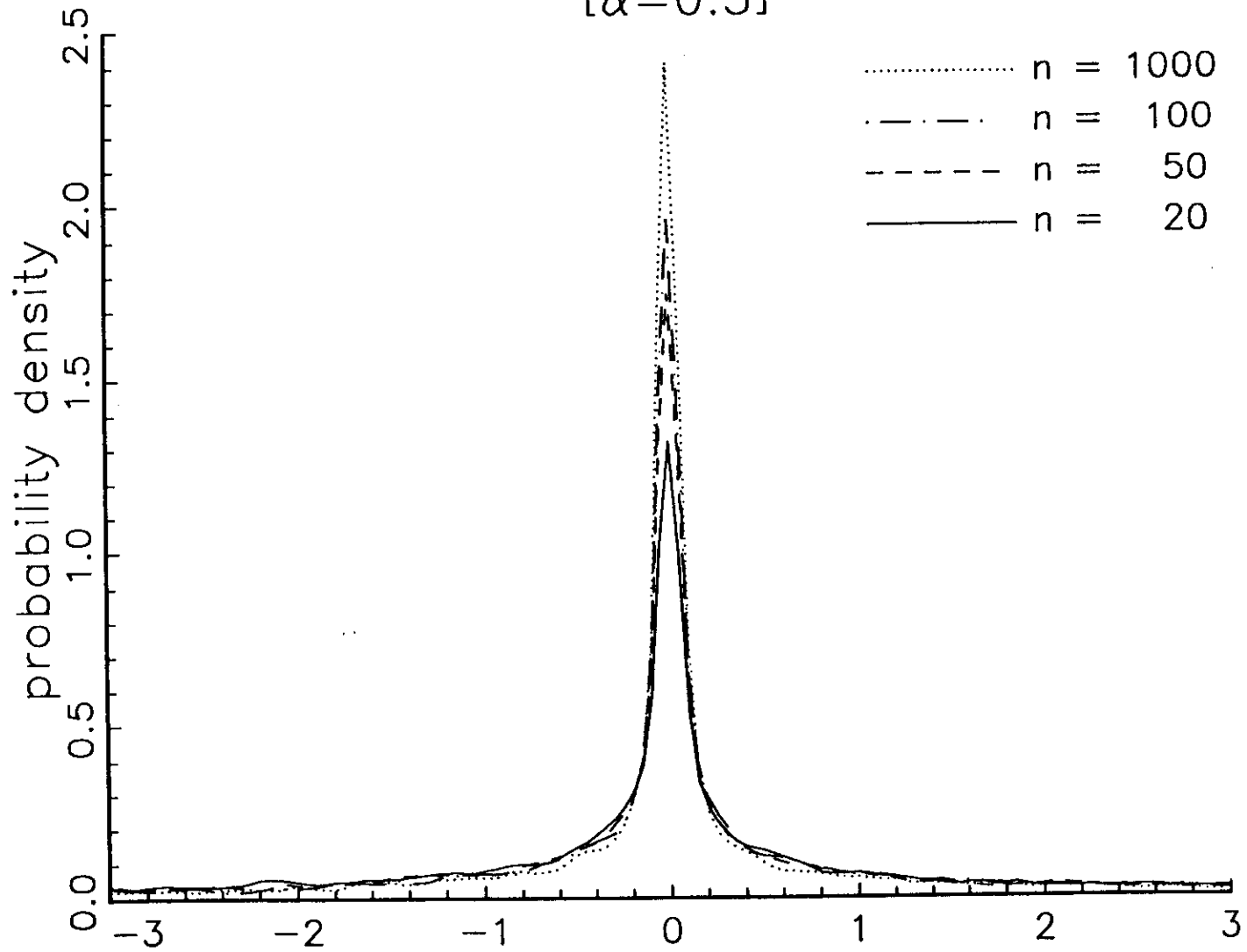


Fig. 3: Small-sample distribution of d
[$\alpha=0.5$]

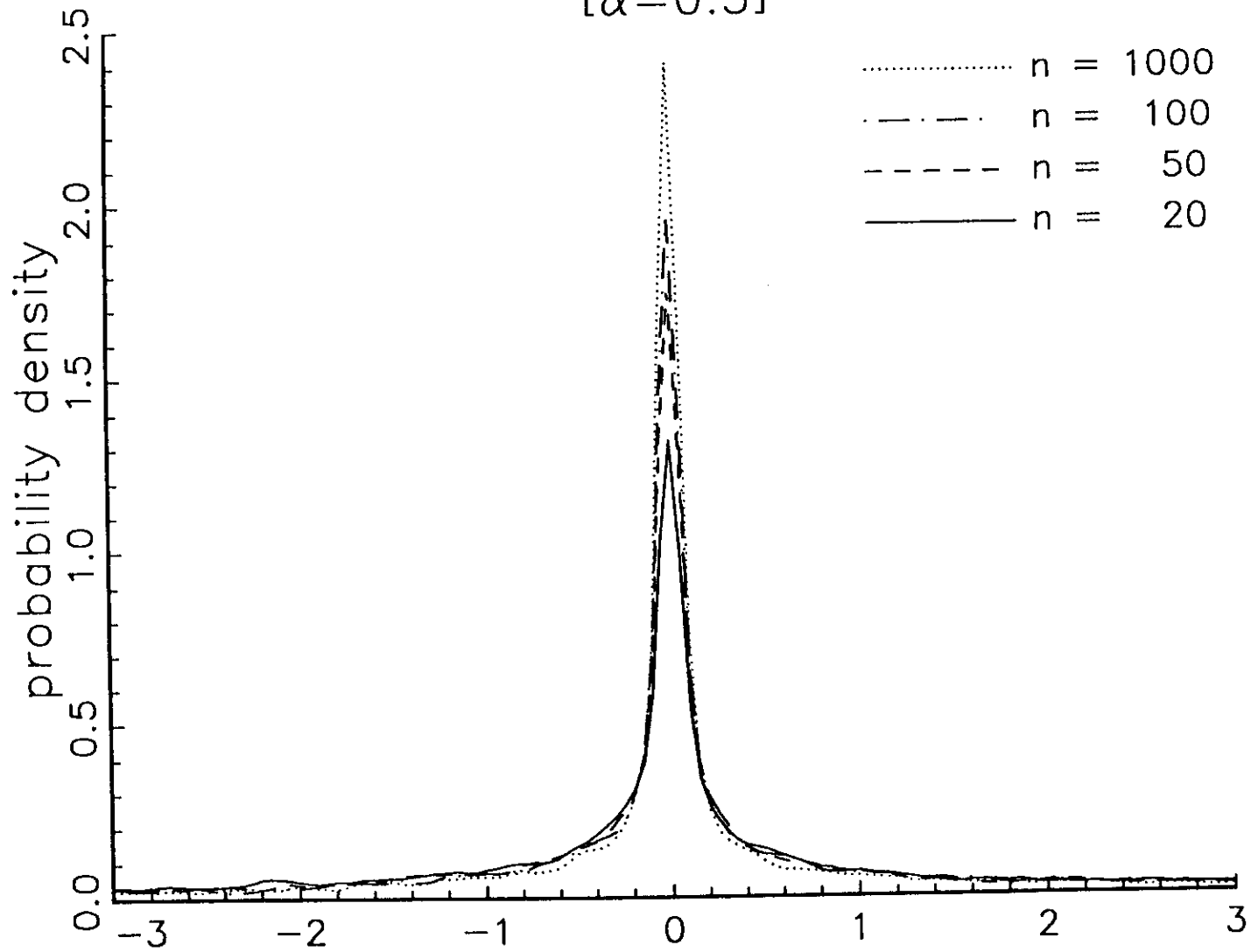


Fig. 4: Small sample distribution of d
[$\alpha=1.0$]

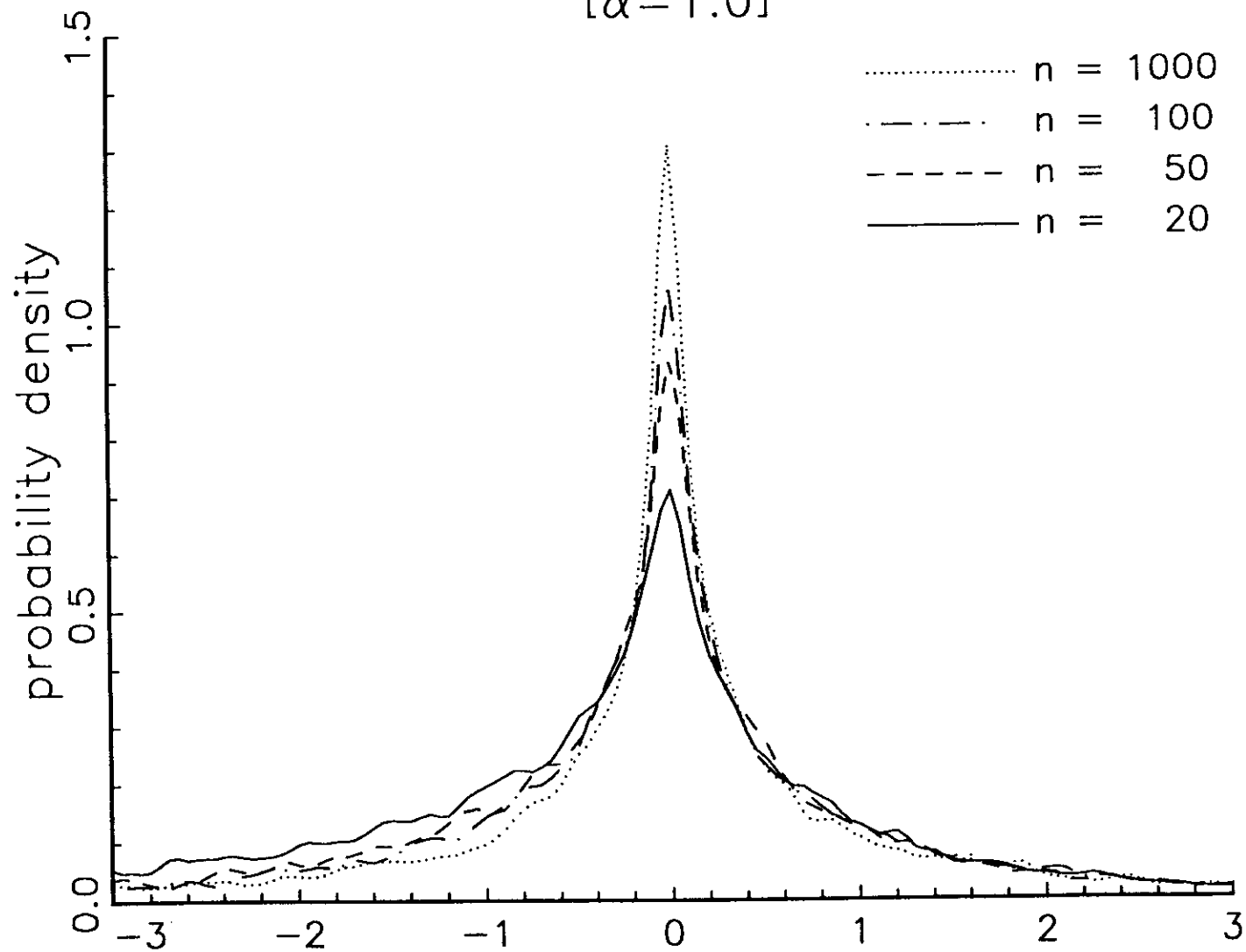


Fig. 5: Small-sample distribution of d
[$\alpha=1.5$]

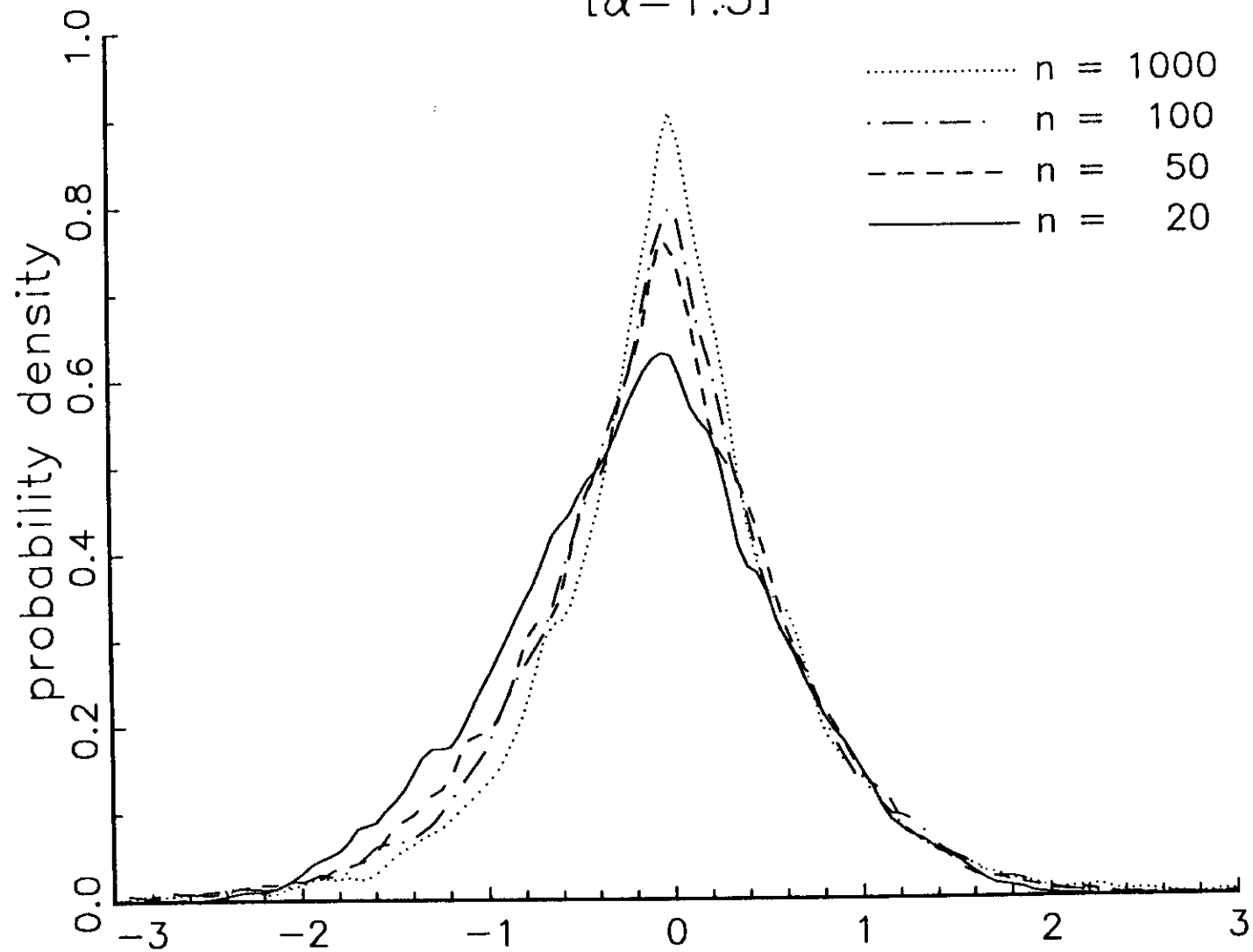


Fig. 6a: Large-sample distribution of d

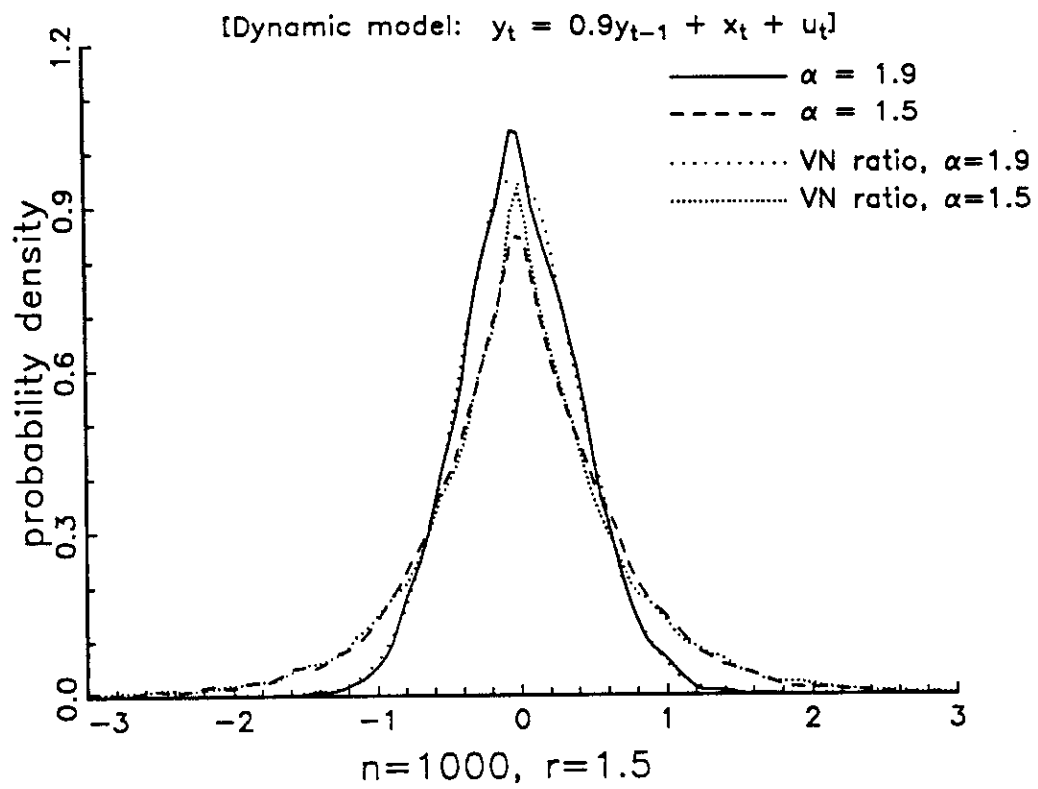


Fig. 6b: Large-sample distribution of d

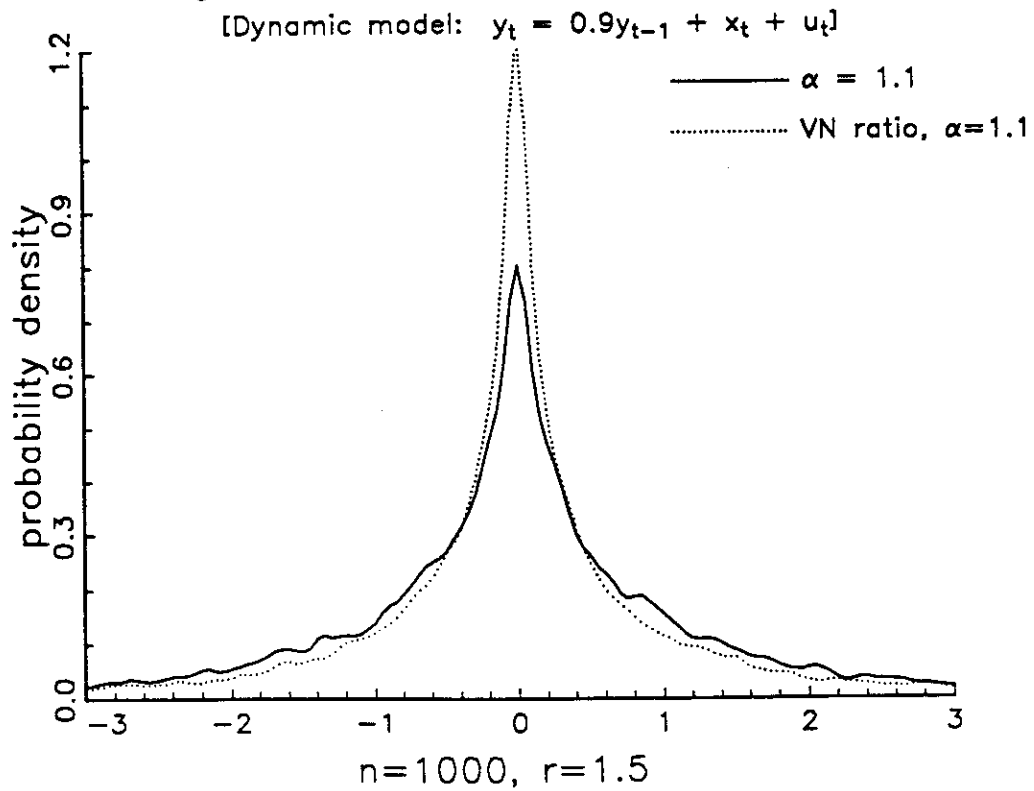


Fig. 7a: Large-sample distribution of d

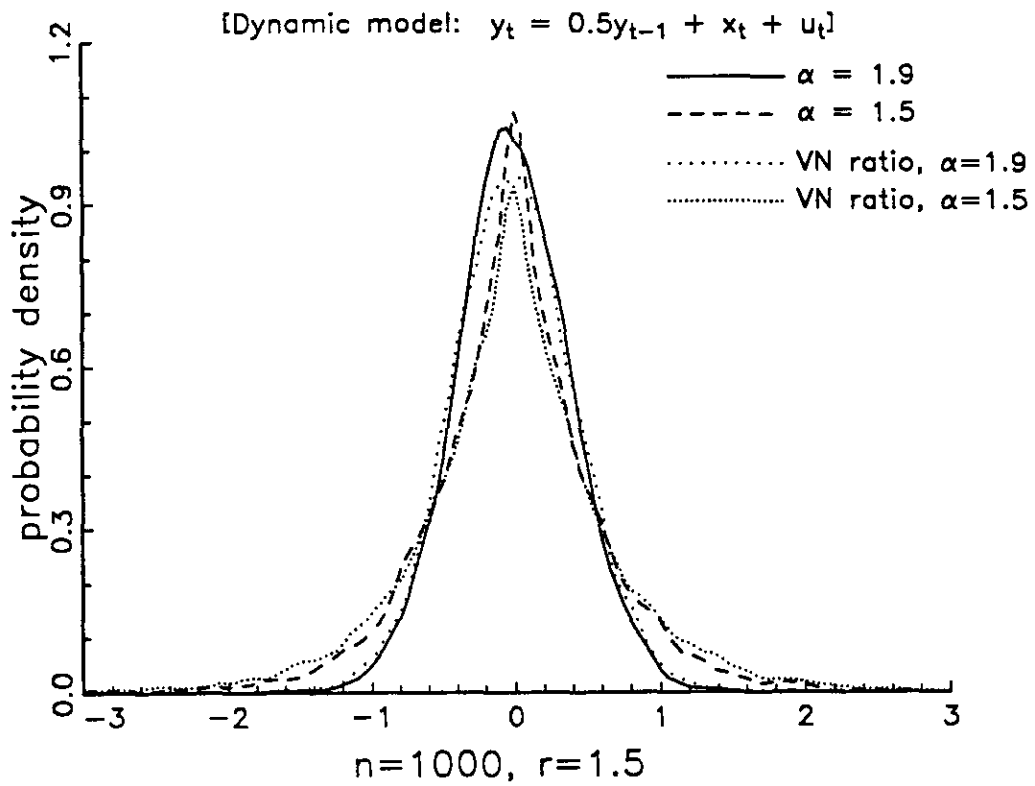


Fig. 7b: Large-sample distribution of d

