FRACTIONAL UNIT ROOT DISTRIBUTIONS

by

Fallaw Sowell*

DP#: 87-05

*Duke University
Durham, NC 27706

PRELIMINARY VERSION
COMMENTS SOLICITED

FRACTIONAL UNIT ROOT DISTRIBUTIONS

Fallaw Sowell

Duke University

July 1987

Revised, August 1987

This is a generalization of Chapter 5 of my Ph.D. thesis, which was supported by the Alfred P. Sloan Foundation through a Doctoral Dissertation Fellowship. I am grateful for several new ideas that were developed during a conversation with Sastry Pantula and Peter Bloomfield. John Geweke provided useful comments on an earlier version, for which I am also grateful.

ABSTRACT

Asymptotic distributions are derived for the ordinary least squares (OLS) estimate of a first order autoregression when the true series is the partial sum of errors that are fractionally integrated of order $1+d$ for $-1/2<d<1/2$. The limiting distributions are written as functions of fractional Brownian motions. A new family of distributions called fractional unit root distributions is introduced to describe the limiting distributions. The unit root distribution ($d=0$) is seen to be an atypical member of this family because its density is nonzero over the entire real line. For $-1/2<d<0$ the density of the fractional unit root distribution is nonzero only on the nonpositive half line, while if $0<d<1/2$ the density of the fractional unit root distribution is nonzero only on the nonnegative half line. Hence any misspecification of the order of differencing leads to radically different limiting distributions. The OLS estimate is shown to be consistent when $-1/2<d<1/2$. Also shown is how the rate of convergence to the limiting distribution depends on the differencing parameter d. Results are proven with functional limit theorems.

I. INTRODUCTION

Most economic time series posses nonstationary characteristics, which indicate standard econometric techniques are inappropriate. To alleviate this problem it is typical to estimate models using first differenced data. This approach to modeling is appropriate if the data were generated by a random walk. Several economic theorists have incorporated the random walk model such as Hall's (1978) consumption function model. A different approach which also uses this type of nonstationarity is the cointegration model, where a linear combination of nonstationary variables is stationary. Examples of cointegration models are Campbell's (1986) model of consumption and Campbell and Shiller's (1986) model of asset pricing. All these models use the random walk as their model for nonstationarity. Statisticians and econometricians have also used the random walk model of nonstationarity. Estimation and hypothesis testing of unit root distributions, which occur with random walks, have been the topic of several notable contributions to econometric theory; for example Dickey and Fuller (1979,1981), Evans and Savin (1981,1984), Sargan and Bhargava (1983), Phillips (1987).

A random walk has some of the nonstationary characteristics of economic time series, but it is only one of an infinite number of such models. Alternative models of nonstationarity do exist. One alternative model of nonstationarity is the fractional integration model that extend ARIMA(p,d,q) models to noninteger values of the differencing parameter d. Studies that have looked for fractional differencing (Granger and Joyeux (1981), Geweke and Porter-Hudak (1983)) have concluded that economic time series do posses fractional unit roots. Unfortunately, the techniques used to estimate these models were not as simple as first difference and apply OLS. Special estimation procedures coupled with no implications of fractional misspecification have led only to occasional use of fractional time series models.

These difficulties of fractional unit root modeling have been addressed in recent research (Sowell (1986)). This paper extends this research and presents the implications of ignoring fractional errors. It reveals that limiting distributions of fractionally integrated series are radically different than series integrated of order zero or one. The slow rate of convergence of OLS estimates of fractionally integrated models to their limiting distributions may explain why anomalies have not manifested themselves in previous work. However, the dissimilarities between the unit root distribution and fractional unit root distributions underscore the importance of considering fractional models.

The approach used to obtain the limiting distributions in this paper is similar to that of Phillips (1987). Phillips uses functional limit theorems to obtain unit root distributions. Phillips' (1987) result is a generalization of a result in White (1958) where the limiting distribution of the OLS estimate of an AR(1) model for a random walk of iid random variables is characterized by functions of Brownian motions. Phillips obtained the limiting distributions when the underlying random variables are strong mixing. The current result generalizes White's (1958) result to the case where the underlying series is fractionally integrated. Fractionally integrated series have greater dependency than allowed by strong mixing. A new class of functional limit theorems is used to obtain the limiting distributions.

Section II contains general properties of fractionally integrated time series and a related continuous stochastic process, fractional Brownian motion. A new family of distributions, fractional unit root distributions, is introduced in the third section. It characterizes the limiting distribution of the OLS estimate of the parameter of an AR(1) model when the series is $I(1+d)$, $-1/2<d<1/2$. Results are proven with a functional limit theorem which is shown to be applicable to fractionally integrated time series.

In the fourth section the convergence of $(\hat{\beta}-1)$ to its limiting distribution is studied through simulations. The last section is a summary of results and discusses directions for future research.

Normal errors are assumed for the proofs of theorems. Normality is assumed to simplify the proofs; the effect of weakening this assumption is noted in the third section. The two forms of convergence in probability and in distribution, will be denoted by $\rightarrow^{p}$ and $\Rightarrow$ respectively.

II. PRELIMINARY CONCEPTS

The fractional difference operator $(1-L)^{d}$ is defined by its Maclaurin series

$$(1-L)^{d} = \sum_{j=0}^{\infty} \frac{\Gamma(-d+j)}{\Gamma(-d)\Gamma(j+1)} L^{j} .$$

A series that is the partial sum of iid normal random variables (a random walk) will be called integrated of order one, I(1), because after applying the differencing operator $(1-L)^{1}$ the series is $IDN(0,\sigma^{2})$. Similarly, a series will be called integrated of order δ, if after applying the differencing operator, $(1-L)^{\delta}$, the series is $IDN(0,\sigma^{2})$. When d is not an integer the series is said to be fractionally integrated.

The following theorem summarizes results concerning fractionally integrated time series. (Proofs may be found in Granger and Joyeux (1980), Hosking (1981) or Sowell (1986).

THEOREM I

If the time series $\epsilon_t \sim I(d)$, i.e. $(1-L)^d \epsilon_t = u_t \sim IND(0,\sigma_u^2)$, then

a) ϵ_t is wide sense stationary if $d<1/2$.

b) ϵ_t has an invertible moving average representation if $d>-1/2$.

c) For $-1/2<d<1/2$, the spectral density of ϵ_t, $f_\epsilon(\lambda)$, is

$$f_\epsilon(\lambda) = (1-e^{-i\lambda})^{-d}(1-e^{i\lambda})^{-d}\sigma_u^2$$

and $f_\epsilon(\lambda) \sim \sigma_u^2 \lambda^{-2d}$ as $\lambda \to 0$.

d) For $-1/2<d<1/2$, the autocovariances of ϵ_t, $\gamma_\epsilon(s)=E[x_t x_{t-s}']$, are

$$\gamma_\epsilon(s) = \frac{\Gamma(1-2d)\Gamma(d+s)}{\Gamma(d)\Gamma(1-d)\Gamma(1-d+s)} \sigma_u^2$$

and as $s\to\infty$, $\gamma_\epsilon(s) \sim Ks^{2d-1}$, where K is the finite constant

$$K = \frac{\Gamma(1-2d)}{\Gamma(d)\Gamma(1-d)} \sigma_u^2.$$

This theorem highlights the important features of fractionally integrated series. Autocovariances are of the same sign as d for $s\geq 1$, i.e. the covariance between observations is of the same sign as the differencing parameter. The autocovariances show the strong dependency between observations which dies out at the rate N^{2d-1}. This long run dependence is recognizable by the behavior of the spectral density at low frequencies. As the frequency approaches zero the spectral density either goes to zero, if $d<0$, or infinity, if $d>0$. The low frequency characteristics show the ability of fractionally integrated series to describe long run behavior with a single parameter.

Fractional integration is a characteristic of discrete stochastic processes. Limiting functions of these discrete series will be described by a continuous stochastic processes, called fractional Brownian motion. Fractional Brownian motion was first introduced and studied by Mandelbrot and Van Ness (1968). It will be derived here from the familiar standard Brownian motion. Consider the measurable space $(f[0,1],\mathcal{F})$, where $f[0,1]$ is the set of real functions on the unit interval and $\mathcal{F}$ are the Borel sets of $f[0,1]$. Wiener measure is a probability measure defined on $(f[0,1],\mathcal{F})$, by the three properties for $X(t) \in f[0,1]$:

1) $P\{X(0)=0\} = 1$
2) $P\{[X(t)-X(s)] \le (t-s)^{-1/2}\alpha\} = \Phi(\alpha)$ for $s \le t$, where $\Phi(z)$ is the normal distribution function.
3) $[X(t_1)-X(t_0)],[X(t_2)-X(t_1)],\ldots,[X(t_n)-X(t_{n-1})]$ are independent for $0 \le t_0 \le t_1 \le \ldots \le t_n \le 1$.

An element drawn from $f[0,1]$ under Wiener measure is referred to as a Brownian motion and will be denoted $W(t)$.

If $g(x)$ is a continuously differentiable function, the stochastic integral

$$\int_a^b g(t)dW(t)$$

for $0 \le a \le b \le 1$ will be <u>defined</u> by the functional that assigns to a function $g(x)$ the random variable

$$g(b)W(b) - g(a)W(a) - \int_a^b g'(t)W(t)dt \quad .$$

Setting $g(t)=1$ yields

$$\int_0^t dW(x) = W(t) \ .$$

Which is one way of defining elements of f[0,1] which are drawn under Wiener measure.

If the last integral is generalized to a fractional integral other stochastic processes are defined. Fractional Brownian motion $W_d(t)$ is defined for $d \in (-1/2, 1/2)$ by

$$\frac{1}{\Gamma(d+1)} \int_0^t (t-x)^d dW(x) = W_d(t) \ .$$

$W_d(t)$ is an element of f[0,1] which is drawn under a fractional Wiener measure and is called a fractional Brownian motion. When d=0 this reduces to Wiener measure and Brownian Motion. Fractional Brownian motion is a continuous Gaussian stochastic process, with mean zero and covariance function

$$C(t,s) = E|W_d(t) - W_d(s)|^2 = |t-s|^{1+2d} \ .$$

The interested reader should see Jonas (1983) for a thorough presentation of the properties of fractional Brownian motions.

III. FIRST ORDER AUTOREGRESSIONS WITH FRACTIONALLY INTEGRATED SERIES

White (1958) characterizes the limiting distribution of the OLS estimate of a AR(1) model for a random walk, as the ratio of functions of Brownian motions, i.e. given the model

$$\begin{aligned} x_j &= \beta x_{j-1} + \epsilon_t \qquad \text{for } j=1,2,\ldots,N \\ x_j &= 0 \qquad\qquad\quad \text{for } j\le 0 \\ \text{where} \qquad \beta &= 1 \qquad\qquad\quad \epsilon_j \sim iid(0,\sigma_\epsilon^2<\infty), \end{aligned}$$

the least squares estimate, $\hat{\beta}$, has the following limiting distribution

$$N(\hat{\beta} - 1) \Rightarrow \frac{(1/2)(W(1)^2 - 1)}{\int_0^1 W(t)^2 dt} .$$

White's result is a special case (d=0). The more general question is the limiting distribution of $\hat{\beta}$ if $\epsilon_j \sim I(d)$ for $-1/2<d<1/2$. The answer follows from the following theorems. THEOREM II highlights a fundamental characteristic of a fractionally integrated time series the growth rate of the variance of the partial sums. THEOREM III shows that a functional limit theorem is applicable to a suitable normalized partial sum of a fractionally integrated time series. Finally, THEOREM IV defines fractional unit root distributions as the limiting distributions of $\hat{\beta}$ when $\epsilon_t \sim I(d)$. Proofs are collected at the end of the paper.

A fundamental characteristic of a fractionally integrated series is the growth of the variance of the partial sums of the series. If $0<d<1/2$, the series has strong positive correlation between observations and the covariances add significantly to the variance of the partial sum. When terms are added the variance grows faster than the typical linear rate. If $-1/2<d<0$, the series in strongly negatively correlated and the variance of the partial sums grows slower than the typical linear rate. This characteristic of a fractionally integrated series demonstrates that the distribution theory developed in Phillips (1987) is not general enough to deal with this situation. In Phillips (1987), Assumption 2.1 (c) requires the variance of the partial sums to grow at a linear rate, but fractionally integrated time series do not. The specific growth rates are presented in the following theorem.

THEOREM II

If $\epsilon_t \sim I(d)$ $-1/2<d<1/2$, i.e. $(1-L)^d \epsilon_t = u_t$ where $u_t \sim IND(0,\sigma_u^2)$, and

$$S_N = \begin{cases} \sum_{t=1}^{N} \epsilon_t & \text{for } 0<N, \\ 0 & \text{for } N\leq 0, \end{cases}$$

then $$\mathrm{Var}(S_N) = \frac{\sigma_u^2 \quad \Gamma(1-2d)}{(1+2d)\Gamma(1+d)\Gamma(1-d)}\left[\frac{\Gamma(1+d+N)}{\Gamma(-d+N)} - \frac{\Gamma(1+d)}{\Gamma(-d)}\right]$$

and $$\mathrm{Var}(S_N)/\,N^{1+2d} \rightarrow_p \frac{\sigma_u^2 \quad \Gamma(1-2d)}{(1+2d)\Gamma(1-d)\Gamma(1+d)} \qquad \text{as } N\rightarrow\infty .$$

THEOREM II shows $\mathrm{Var}(S_N)=O(N^{1+2d})$, an important characteristic of processes that are fractionally integrated. The variance of partial sums of iid variables (d=0) grows at the linear rate N. Each random shock is uncorrelated with the others and only adds its own variance to the variance of the partial sum. When $-1/2<d<0$, each shock is negatively correlated with the others. Therefore, the variance of the partial sum grows less than the variance of the individual shock. When d is near -1/2 the negative covariances almost totally offsets the added variance of the shock. For $0<d<1/2$ the shocks are positively correlated and the variance of the sum grows faster than the variance of a single shock. When d is near 1/2, the growth of the variance of the partial sum is almost quadratic.

This growth rate of the variance of partial sums is crucial in being able to apply a class of functional limit theorems. Unlike the functional limit theorem used in White (1958) and its generalizations used in Phillips (1987) the limiting stochastic process is not Brownian motion but is fractional Brownian motion. These functional limit theorems are presented in Davydov

(1970) and Taqqu (1975). The applicability of these functional limit theorems to fractionally integrated time series is noted in the following theorem.

THEOREM III

If $\epsilon_t \sim I(d)$ for $-1/2<d<1/2$, i.e. $(1-L)^d\epsilon_t=u_t$ where $u_t \sim IND(0,\sigma_u^2)$, and

$$S_N = \begin{cases} \sum_{t=1}^{N} \epsilon_t & \text{for } 0<N \\ 0 & \text{for } N\leq 0 \end{cases}$$

and $\sigma_N^2=Var(S_N)$ then $Z_N(t) = \sigma_N^{-1}S_{[Nt]} \Rightarrow W_d(t)$.

The normality assumption is not a necessary condition for the result in THEOREM III. The assumption can be replaced by the restrictions that u_t is stationary and $E|u_t|^r<\infty$ where $r=\max\{4,4+(8d/(1+2d))\}$. The fundamental result needed to prove THEOREM III with these assumptions is Theorem 2 in Davydov (1970). Normality is assumed here only to simplify the statements and proofs of the theorems. As was previously known for the case d=0, these limiting distributions for -1/2<d<1/2 do not depend on the moments of the random variables (beyond their existence).

With these theorems it is possible to generalize White's result to fractionally integrated error processes.

THEOREM IV[1]

Given the model

$$x_t = \beta x_{t-1} + \epsilon_t \text{ for } t=1,2,\dots,N$$

$$\beta = 1$$

$$x_t = 0 \text{ for } t \le 0,$$

$$\epsilon_t \sim I(d) \text{ for } -1/2<d<1/2, \text{ i.e. } (1-L)^d \epsilon_t = u_t \text{ where } u_t \sim IND(0,\sigma_u^2),$$

if the least squares estimator of β is denoted $\hat{\beta}$, then

$$(\hat{\beta}-1) = A_d + B_d$$

and

$$NA_d \Rightarrow \frac{\frac{1}{2}\left[W_d(1)\right]^2}{\int_0^1 \left[W_d(s)\right]^2 ds} \quad \text{and} \quad N^{1+2d}B_d \Rightarrow -\frac{\left[\frac{1}{2}+d\right]\frac{\Gamma(1+d)}{\Gamma(1-d)}}{\int_0^1 \left[W_d(s)\right]^2 ds} \quad .$$

The limiting distribution $N^{\min[1,1+2d]}(\hat{\beta}-1)$ is called a fractional unit root distribution. This limiting distribution is achieved by normalizing, by N if $d \ge 0$ and by N^{1+2d} if $d \le 0$. If $d>0$ then $N<N^{1+2d}$ and NB_d converges in probability to zero leaving asymptotically the distribution of NA_d. Conversely when $d<0$ the limiting distribution is that of $N^{1+2d}B_d$ because $N>N^{1+2d}$ and $N^{1+2d}A_d$ converges in probability to zero.

An immediate corollary is that the least squares estimate is consistent. The rate of convergence depends on the order of integration in a surprising way. It is well known that if the x_t series is I(0), the OLS estimate $\hat{\beta}$ converges at the rate $N^{1/2}$, and if the x_t series is I(1), convergence is at

---

[1] I thank Sastry Pantula for simplifying several steps of the proof and alleviating some unnecessary assumptions of an earlier version.

the rate N. This suggests that the rate of convergence increases with the order of integration. However, this is not the case. If a series is I(1+d) for $0 \le d < 1/2$, $\hat{\beta}$ converges at the rate N, the same for all d. If $-1/2 < d \le 0$, $\hat{\beta}$ converges at the rate N^{1+2d}. So if $-1/2 < d < -1/4$ the rate of convergence is even less than the I(0) case, $N^{1/2}$.

Fractional unit root distributions imply problems exist when testing fractionally integrated series for unit roots. If a series is fractionally integrated with $-1/2 < d < 0$, a unit root statistic that is normalized with N converges to infinity suggesting the hypothesis of I(0) is likely to be accepted when the series is more closely I(1). In general unit root distributions are not robust to any misspecification in the order of integration. NA_d is the distribution of a nonnegative random variable and $N^{1+2d}B_d$ is the distribution of a nonpositive random variable. Only when d=0 is the limiting density nonzero over the entire real line.

If normal errors are assumed the limiting distribution is independent of the variance of the errors. In general, fractional unit root distributions do not depend on the specific structure of the underlying distributions, e.g. variance, skewness, kurtosis, but only depend on general characteristics such as the existence of moments.

IV. $(\hat{\beta}-1)$ AND FRACTIONAL UNIT ROOT DISTRIBUTIONS

Fractional unit root distributions are defined as the limit of

$$N^{\min[1,1+2d]}(A_d + B_d).$$

The rate of convergence of this term to its fractional unit root distribution limit is slow for values of d near zero. The term is composed to two random variables: one which approaches a nontrivial limit distribution, and one which converges to zero. The term that converges to zero does so at the rate $N^{-|2d|}$

which is very slow when d is near zero. This implies that $(\hat{\beta}-1)$ converges to its limiting distribution at a very slow rate when d is near zero. This does not mean that the misspecification is not important for small values of d. Rather it implies that asymptotic theory may not be applicable. Whether or not an investigator should be concerned with fractional units roots is the question of how close the small sample distribution is to the unit root distribution. If an investigator should use fractional unit root distributions is a question of how close the small sample distribution is from its true limit the fractional unit root distribution. To help answer these questions 500 samples of 200 observations for I(1+d) series were generated and the OLS estimate of β was calculated.

To simulate a fractional integrated time series, first the covariance matrix Σ (a 200x200 matrix) was calculated using the equation in THEOREM I part d. The Cholesky decomposition $\Sigma = CC'$ was then calculated, where C is a lower triangular matrix. The simulated sample was then obtained by multiplying C and a 200x1 vector of computer generated independent standard normal random variables. The IMSL subroutine GGNML was used to generate the independent standard normal random variables. Samples were generated for eleven different values of d: -0.49, -0.4, -0.3, -0.2, -0.1, 0.0, 0.1, 0.2, 0.3, 0.4, 0.49 . The densities were then estimated by using a kernel estimator.

The kernel estimator used was

$$\hat{f}(x) = \frac{1}{(500)h} \sum_{j=1}^{500} \psi\left(\frac{x-\hat{\theta}_j}{h}\right) .$$

Where $\hat{\theta}_j$ are the estimated parameters and $\psi(y)$ is the density defined by $(15/16)((1 - y^2)^2)$ for $|y| \leq 1$ and zero elsewhere. Tapia and Thompson (1978) report that the smoothness properties of estimates using this kernel were "nearly indistinguishable" from those of the Gaussian kernel estimators. The value of h was chosen to minimize the integrated mean square error. This was done by using the h calculated by the tenth iteration of the procedure outlined in Tapia and Thompson (1978, p.67)

The estimated densities are plotted in FIGURE 1 thru FIGURE 4. The thick curve is the estimated density of

$$200^{\min[1,1+2d]}(\hat{\beta}-1).$$

This density is an estimate of the small sample density of $(\hat{\beta}-1)$ and asymptotically approaches the fractional unit root density. Superimposed on each plot is an estimate of the fractional unit root density. This estimate to the fractional unit root distribution is achieved by setting to zero that part of $(\hat{\beta}-1)=(A_d+B_d)$ which asymptotically goes to zero. If $d<0$, $200^{1+2d}A_d$ converges in probability to zero, hence the fractional unit root density is approximated by the density of $200^{1+2d}B_d$. Similarly, if $0<d$ the fractional unit root density is approximated by the density of $200A_d$, sense $200B_d$ converges in probability to zero. The estimate of the fractional unit root density is the thin curve in each plots in FIGURE 1 thru FIGURE 4. These estimates of the fractional unit root densities are presented by themselves in FIGURE 5 and FIGURE 6.

From THEOREM IV it is known that the limiting density of $\hat{\beta}$-1, i.e. θ, is nonnegative for $0<d<1/2$ and is nonpositive when $-1/2<d<0$, however; even with 200 observations, most of the estimated "small" sample densities have nonnegligible mass on both sides of zero. This implies that very large samples are required before asymptotic theory is applicable.

Several features should be noted. First, the small sample density estimates and the asymptotic density estimates are closer the further d is from zero (of course they are equal at zero). This is to be expected in general because the difference between N^{1+2d} and N is greater the further d is from zero. The second feature to notice is that as d approaches -1/2 the estimated densities appear to be converging to unit mass at zero. This is indeed the case because the numerator of $N^{1+2d}B_d$ goes to zero at d=-1/2. This type of behavior is not noted as d approaches 1/2 because the numerator of NA_d asymptotically has a chi-square distribution with one degree of freedom and hence is asymptotically independent of d.

The small sample densities for fractionally integrated series are quite different depending on the value of d. Hence, standard unit root tests have noticeably different size and power and are inappropriate when applied to fractionally integrated time series. If $-1/2<d<0$ the Dickey and Fuller (1979) test statistics do not posses a limiting distribution and tend to infinity as the number of observations increase. Because the longer negative tail of the unit root density is lost when $0<d<1/2$, it is conjectured that the test will accept the null $H_0:\beta=1$ more frequently then expected. In general, typical unit root tests do not have power against fractional alternatives.

Applicable tests would involve estimating the differencing parameter d. Two alternatives currently exist. Geweke and Porter-Hudak (1983) present an estimation technique that exploits the distinctive spectral density of a fractionally integrated series to estimate the d parameter. Alternatively, the unconditional maximum likelihood estimation of a univariate ARIMA(p,d,q) model, $-1/2<d<1/2$, could be used to estimate d. This univariate maximum likelihood procedure is a special case of the general results in Sowell (1986).

V. CONCLUSIONS

A new family of distributions called fractional unit root distributions was introduced to describe the limit distribution of the parameter of an AR(1) model when the series is I(1+d), $-1/2<d<1/2$. The unit root distribution is the special case $d=0$ and is an atypical member of the distribution family. The dissimilarities emphasis the importance of considering fractional unit roots when modeling economic time series.

Extensions of this work are currently being pursued in two directions. The question of the correct model of nonstationarity for economic time series is being investigated by maximum likelihood estimation of fractionally ARIMA models. It should be noted that the unit root model found to be so pervasive in economic time series by Nelson and Plosser (1982) is a special case of the fractionally ARIMA models. Because a fractional unit root accounts for low frequency (i.e. long run) characteristics, fractional models should produce better long run predictions.

Also currently being investigated, is if the current theorems can be generalized to more general error processes, similar to Phillips generalization of White's result to strong mixing processes. The question that needs to be answered for this result is "do the normalized partial sums converge to fractional Brownian motion, when the fractional difference operator applied to the original series produces an error series that is strong mixing?"

PROOFS OF THE THEOREMS

The proof of THEOREM II is simplified by the following.

LEMMA

For $N=1,2,\ldots$ and $a,b\in(-1,1)$

$$\sum_{k=1}^{N} \frac{\Gamma(a+k)}{\Gamma(b+k)} = \frac{1}{1+a-b}\left[\frac{\Gamma(1+a+N)}{\Gamma(b+N)} - \frac{\Gamma(1+a)}{\Gamma(b)}\right]$$

when $b=0$ this reduces to

$$= \frac{1}{1+a}\left[\frac{\Gamma(1+a+N)}{\Gamma(N)}\right].$$

PROOF The lemma follows by induction and the relation $x\Gamma(x)=\Gamma(x+1)$.■

PROOF OF THEOREM II.

From THEOREM I part d, $E[\epsilon_t\epsilon_{t-s}]=\gamma_\epsilon(s)=K\Gamma(d+s)/\Gamma(1-d+s)$ where K is the constant

$$K = \frac{\Gamma(1-2d)\ \sigma_u^2}{\Gamma(d)\Gamma(1-d)} \quad .$$

$E\epsilon_t=0$ so the variance of the partial sum can be written

$$\mathrm{Var}\left[\sum_{j=1}^{N}\epsilon_j\right] = K\left[\frac{N\Gamma(d)}{\Gamma(1-d)} + 2\sum_{k=1}^{N-1}\frac{(N-k)\Gamma(d+k)}{\Gamma(1-d+k)}\right]$$

$$= K\sum_{k=1}^{N}\left[-\frac{\Gamma(d)}{\Gamma(1-d)} + 2\sum_{m=1}^{k}\frac{\Gamma(d-1+m)}{\Gamma(-d+m)}\right].$$

Using the above LEMMA twice this reduces to

$$= \frac{K}{d(1+2d)}\left[\frac{\Gamma(1+d+N)}{\Gamma(-d+N)} - \frac{\Gamma(1+d)}{\Gamma(-d)}\right]$$

Using the definition of K, the fact that $x\Gamma(x)=\Gamma(1+x)$, and the result

$\Gamma(a+N)/\Gamma(b+N) \sim N^{a-b}$ as $n \to \infty$; the theorem is proven.■

PROOF OF THEOREM III.

Apply Lemma 5.1 from Taqqu (1975) with THEOREM II above.■

PROOF OF THEOREM IV.

If the variance of x_N is denoted by σ_N^2, then

$$(\hat{\beta}-1) = \frac{\dfrac{1}{N\sigma_N^2} \displaystyle\sum_{t=1}^{N} x_{t-1}\epsilon_t}{\dfrac{1}{N\sigma_N^2} \displaystyle\sum_{t=1}^{N} x_{t-1}^2} .$$

First consider the denominator,

$$\frac{1}{N\sigma_N^2} \sum_{t=1}^{N} x_t^2 = \frac{1}{N} \sum_{t=1}^{N} \left[\sigma_N^{-1}\, x_t\right]^2$$

$$= \sum_{t=1}^{N} \int_{\frac{(t-1)}{N}}^{t/N} \left[\, Z_N(s)\, \right]^2 ds$$

$$= \int_0^1 \left[\, Z_N(s)\, \right]^2 ds \Rightarrow \int_0^1 \left[W_d(s)\right]^2 ds$$

which holds for $-1/2<d<1/2$. The last equality follows from the Continuous Mapping Theorem and THEOREM III.

Now consider the numerator

$$\frac{1}{N\sigma_N^2} \sum_{t=1}^{N} x_{t-1}\epsilon_t = \frac{1}{2N\sigma_N^2} \sum_{t=1}^{N} \left[x_t^2 - x_{t-1}^2\right] - \frac{1}{2N\sigma_N^2} \sum_{t=1}^{N} \epsilon_t^2$$

$$= \frac{1}{2N}\left[\sigma_N^{-1} \ x_N\right]^2 - \frac{1}{2N\sigma_N^2} \sum_{t=1}^{N} \epsilon_t^2$$

$$= \frac{1}{2N} \left[Z_N(1)\right]^2 - \frac{1}{2N\sigma_N^2} \sum_{t=1}^{N} \epsilon_t^2$$

The first term when multiplied by N converges in distribution to $(1/2)[W_d(1)]^2$ for $-1/2<d<1/2$ again because of the Continuous mapping Theorem and THEOREM III. This proves the limiting distribution of the A_d term. The limiting distribution of the second term follows by first noting

$$\left[\frac{1}{N} \sum_{t=1}^{N} \epsilon_t^2\right] = E\epsilon_t^2 = \frac{\Gamma(1-2d)}{\Gamma(1-d)\Gamma(1-d)}\sigma_u^2$$

by the Ergodic Theorem and THEOREM I part d. Finally, using THEOREM II,

$$N^{1+2d}/\sigma_N^2 \to^p \frac{(1+2d)\Gamma(1-d)\Gamma(1+d)}{\Gamma(1-2d)}\sigma_u^{-2}$$

Which shows the limiting distribution of B_d and proves the theorem.■

REFERENCES

Billingsley, Patrick. Convergence of Probabilty Measures. New York: John Wiley and Sons, 1968.

Campbell, John Y. "Does Saving Anticipate Declining Labor Income? An Alternative Test of the Permanent Income Hypothesis," NBER Working Paper No. 1805, 1986.

Campbell, John Y., and Shiller, Robert J. "Cointegration and Tests of Present Value Models," Cowles Foundation Discussion Paper No. 785, 1986.

Davydov, Yu. A. "The Invariance Principle for Stationary Processes," Theory of Probability and Its Applications 15, 1970: 487-489.

Dickey, David A., and Fuller, Wayne A. "Distribution of the Estimator for Autoregressive Time Series with a Unit Root." Journal of the American Statistical Association 74, 1979: 427-431.

Dickey, David A., and Fuller, Wayne A. "Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root." Econometrica 49, 1981: 1057-1072.

Evans, G.B.A., and Savin, N.E. "Testing for Unit Roots: 1." Econometrica 49, 1981.

Evans, G.B.A., and Savin, N.E. "Testing for Unit Roots: 2." Econometrica 52, 1984.

Geweke, John F., and Porter-Hudak, Susan. "The Estimation and Application of Long Memory Time Series Models." Journal of Time Series Analysis 4, 1982: 221-238.

Granger, C.W.J., and Joyeux, Roselyne. "An Introduction to Long-Memory Time Series Models and Fractional Differencing." Journal of Time Series Analysis 1, 1980: 15-29.

Hall, R.E. "Stochastic Implications of the Life Cycle Permanent Income Hypothesis: Theory and Evidence." Journal of Political Economy 86, 1978.

Hosking, J.R.M. "Fractional Differencing." Biometrica 68. 1981: 165-76.

Jonas, Andrew Barrande. Persistent Memory Random Processes, Ph.D. Thesis, Department of Statistics Harvard University. 1983.

Mandelbrot, Benoit B., and Van Ness, John W. "Fractional Brownian Motions, Fractional Brownian Noises and Applications." SIAM Review 10, 1968: 422-437.

Nelson, Charles R., and Plosser, Charles I. "Trends and Random Walks in Macroeconomic Time Series." Journal of Monetary Economics 10, 1982.

Phillips, P.C.B. "Time Series Regression with a Unit Roots." Econometrica 55, 1987.

Sargan J.D., and Bhargava, A. "Testing Residuals from Least Squares Regresssion for Being Generated By a Gaussian Random Walk." Econometrica 51, 1983.

Sowell, Fallaw B. Fractionally Integrated Vector Time Series, Ph.D. Thesis, Department of Economics Duke University, 1986.

Tapia, Richard A., and Thompson, James R. Nonparametric Probability Density Estimation. Baltimore: The Johns Hopkins University Press, 1978.

Taqqu Murad S. "Weak Convergence to Fractional Brownian Motion and to the Rosenblatt Process." Z. Wahrscheinlichkeitstheorie verw. Gebiete 31, 1975: 287-302.

White, John S. "Limiting Distribution of the Serial Correlation Coefficient in the Explosive Case." Annals of Mathematical Statistics 30, 1958: 1188-1197.

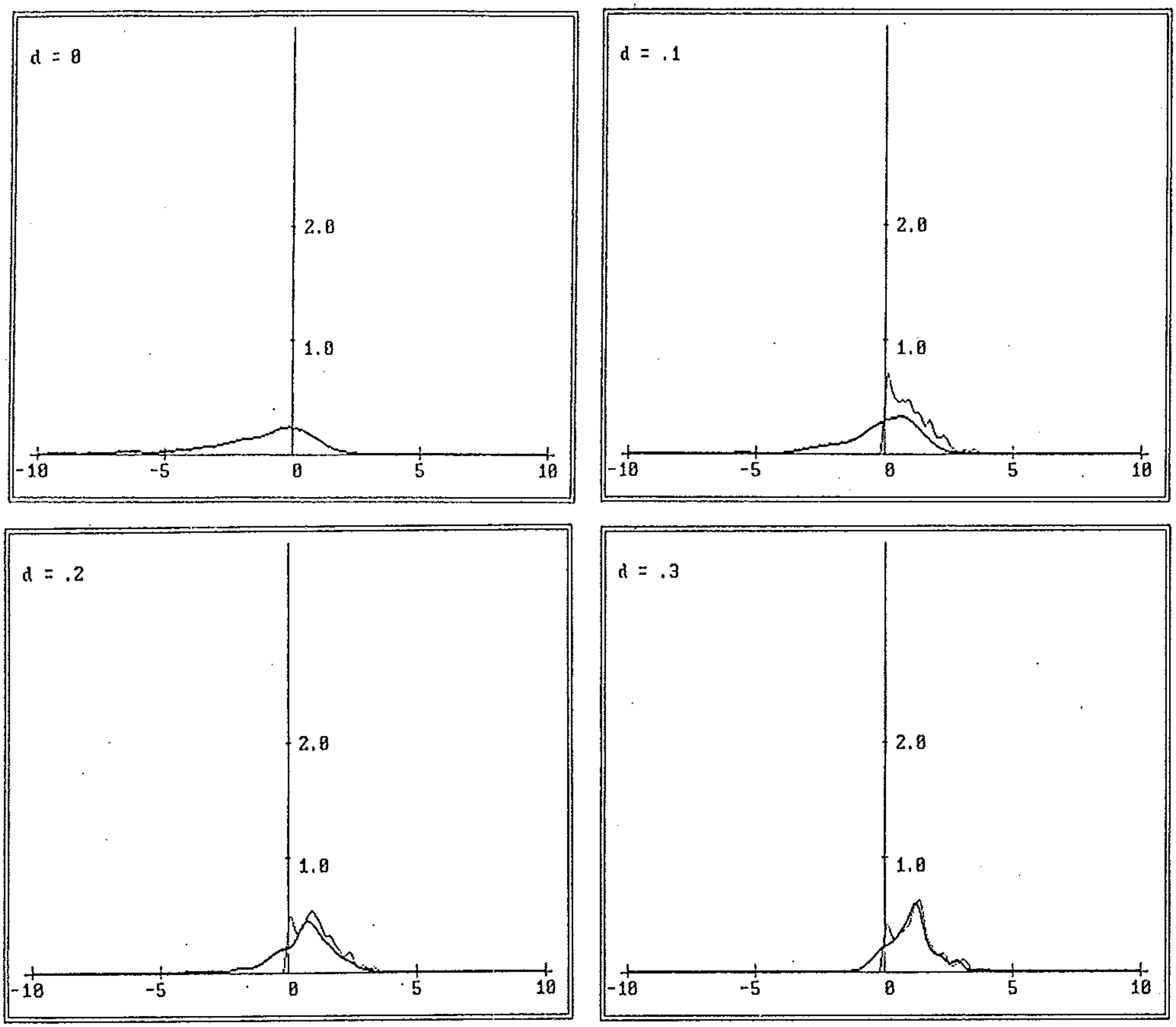

FIGURE 1

Estimates of the small sample density and of the asymptotic density of $N^{\min[1,1+2d]}(\hat{\beta}-1)$, for N = 200 and d = 0.0, 0.1, 0.2, 0.3. The thick density is the estimate of the small sample density and is the kernel estimate from the 500 samples of $200(\hat{\beta}-1)$. The thin density (for $d \neq 0$) is the estimated asymptotic density and is the kernel estimate from the same 500 samples of $200A_d$.

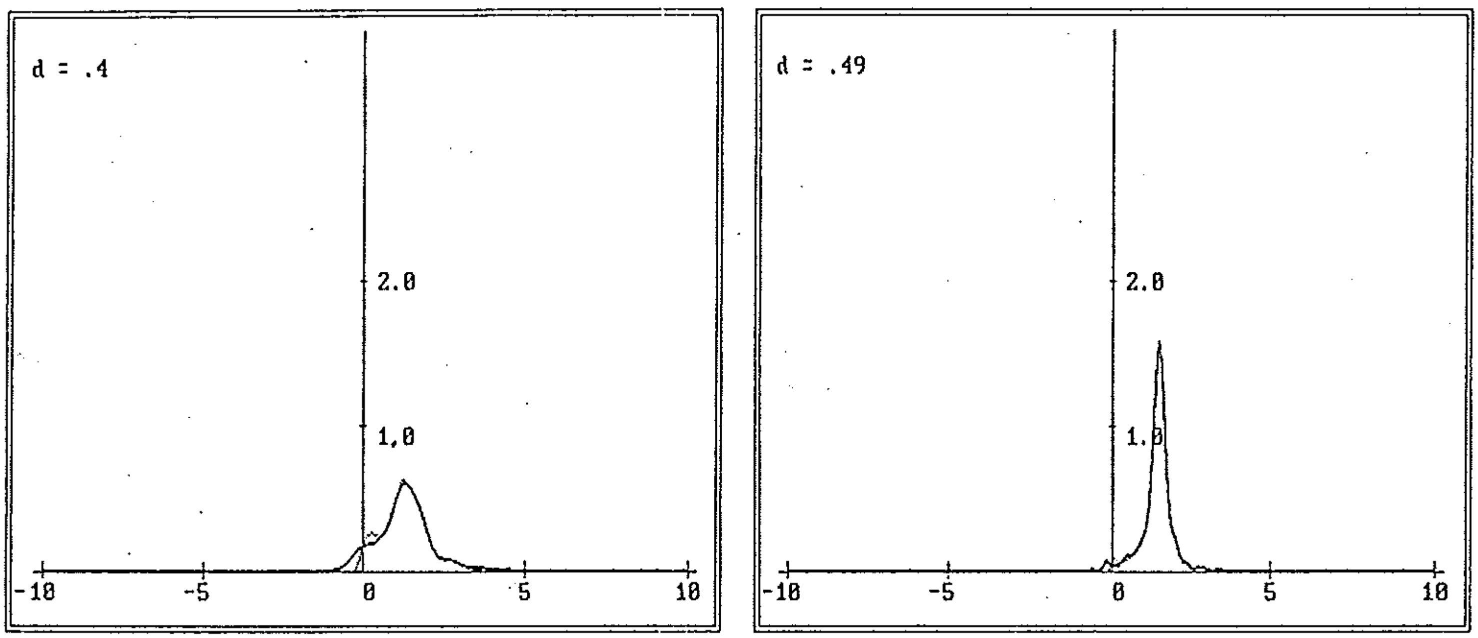

FIGURE 2

Estimates of the small sample density and of the asymptotic density of $N^{\min[1,1+2d]}(\hat{\beta}-1)$, for N = 200 and d = 0.4 and 0.49. The thick density is the estimate of the small sample density and is the kernel estimate from the 500 samples of $200(\hat{\beta}-1)$. The thin density is the estimated asymptotic density and is the kernel estimate from the same 500 samples of $200A_d$.

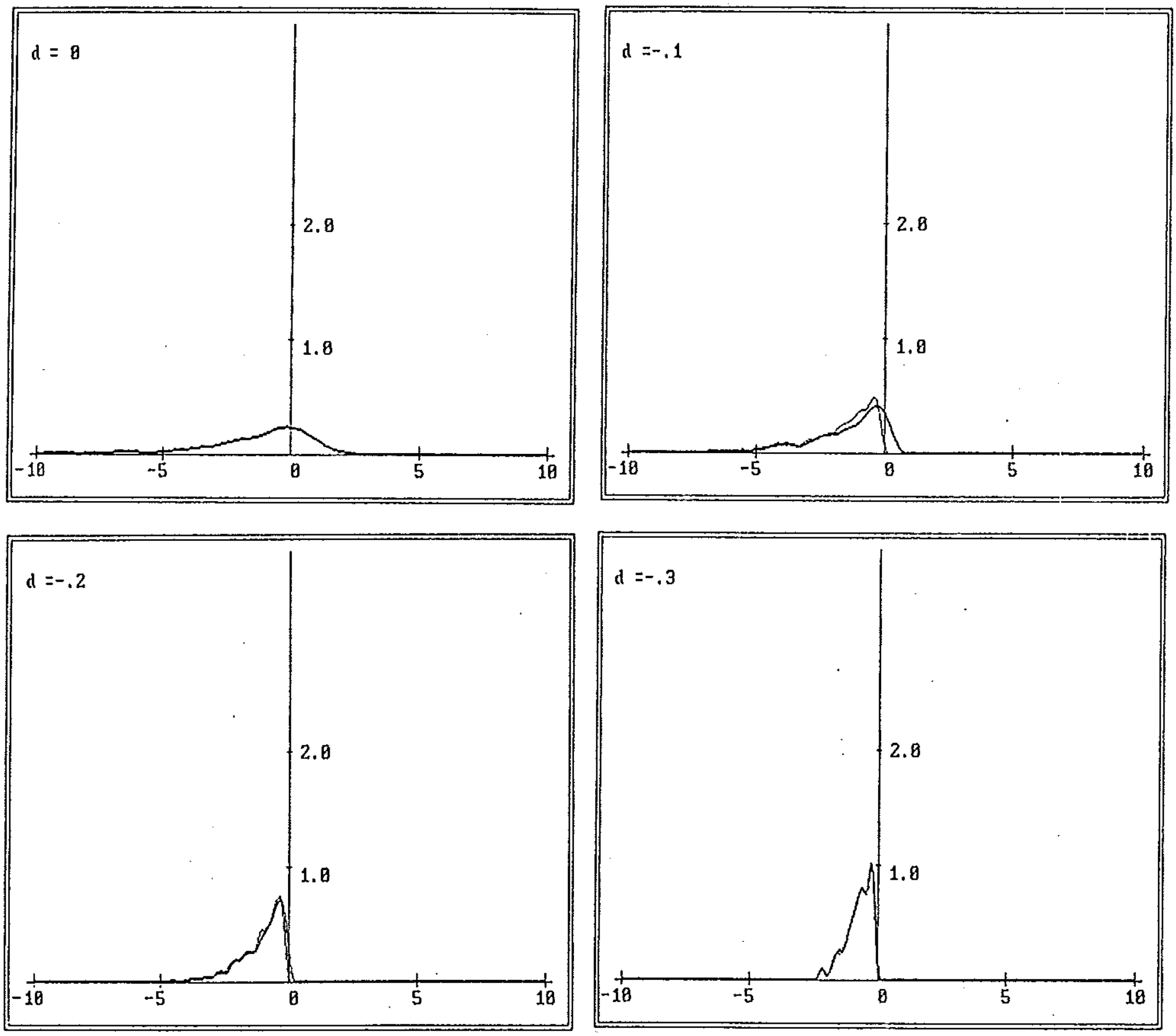

FIGURE 3

Estimates of the small sample density and of the asymptotic density of $N^{\min[1,1+2d]}(\hat{\beta}-1)$, for N = 200 and d = 0.0, -0.1, -0.2, -0.3. The thick density is the estimate of the small sample density and is the kernel estimate from the 500 samples of $200(\hat{\beta}-1)$. The thin density (for $d \neq 0$) is the estimated asymptotic density and is the kernel estimate from the same 500 samples of $200^{1+2d}B_d$.

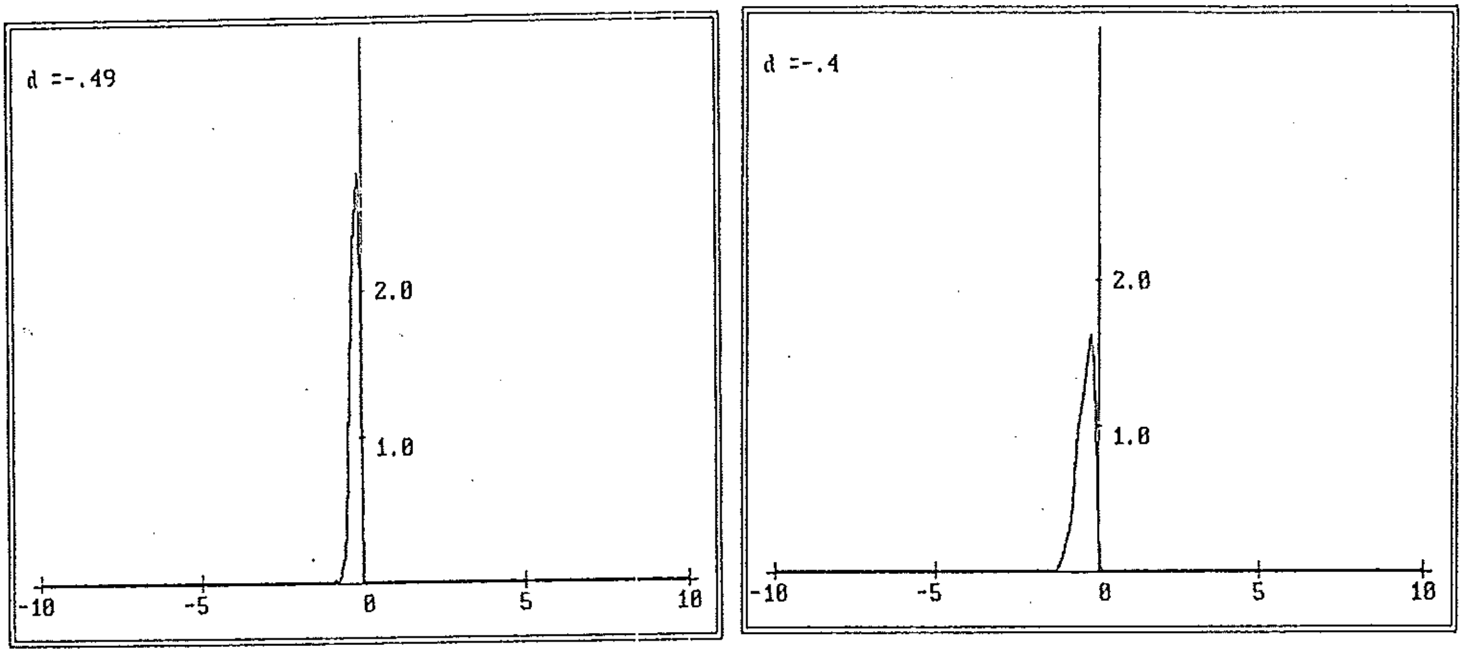

FIGURE 4

Estimates of the small sample density and of the asymptotic density of $N^{\min[1,1+2d]}(\hat{\beta}-1)$, for N = 200 and d = -0.4 and -0.49. The thick density is the estimate of the small sample denisty and is the kernel estimate from the 500 samples of $200^{1+2d}(\hat{\beta}-1)$. The thin density is the estimated asymptotic density and is the kernel estimate from the same 500 samples of $200^{1+2d}B_d$.

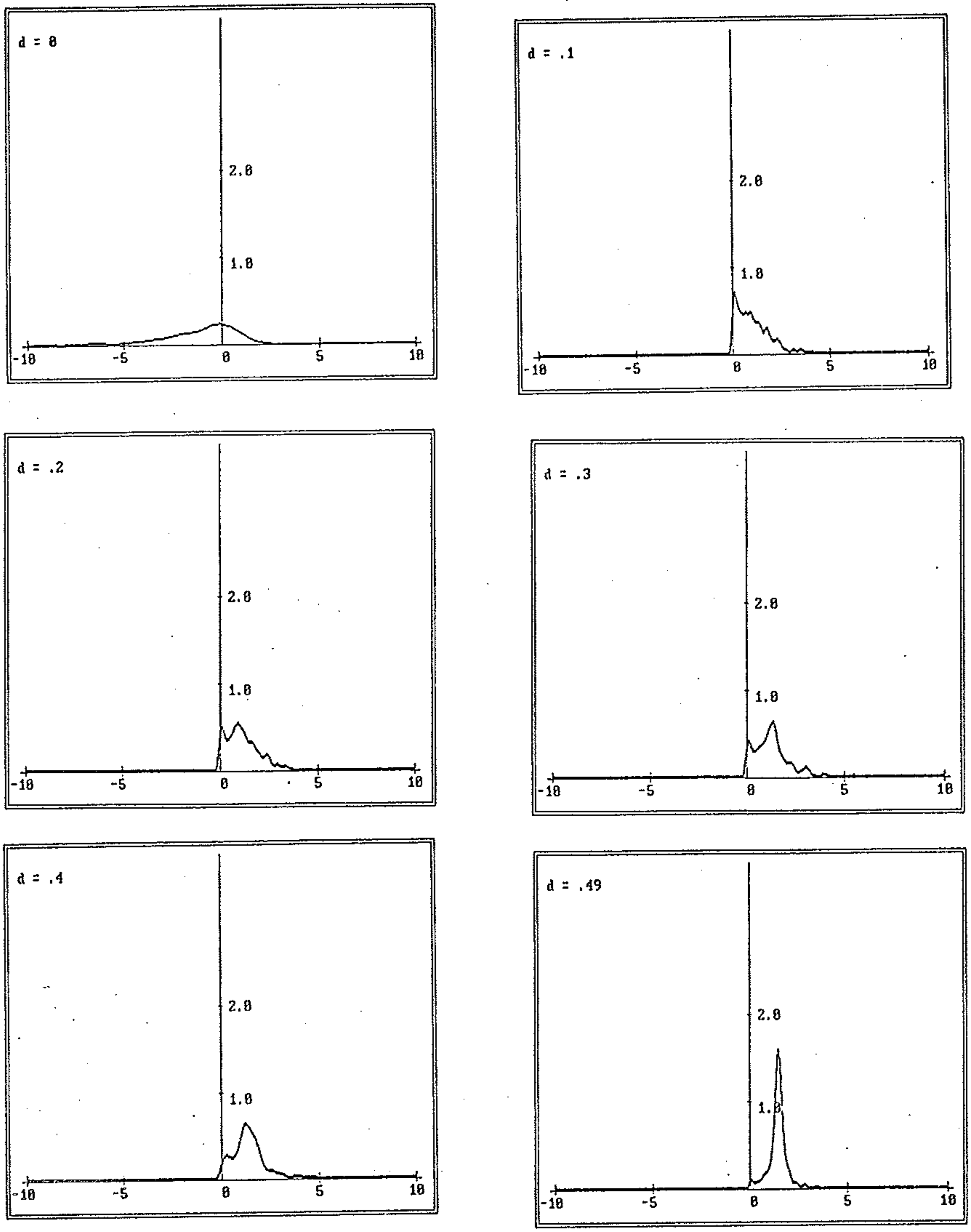

FIGURE 5

Estimates of the fractional unit root density for d = 0.0, 0.1, 0.2, 0.3, 0.4, 0.49. The plotted density is the kernel estimate of the density for $200(A_d+B_d)$ when d = 0.0. For $d \neq 0.0$ the plotted density is the kernel estimate of the density of $200A_d$.

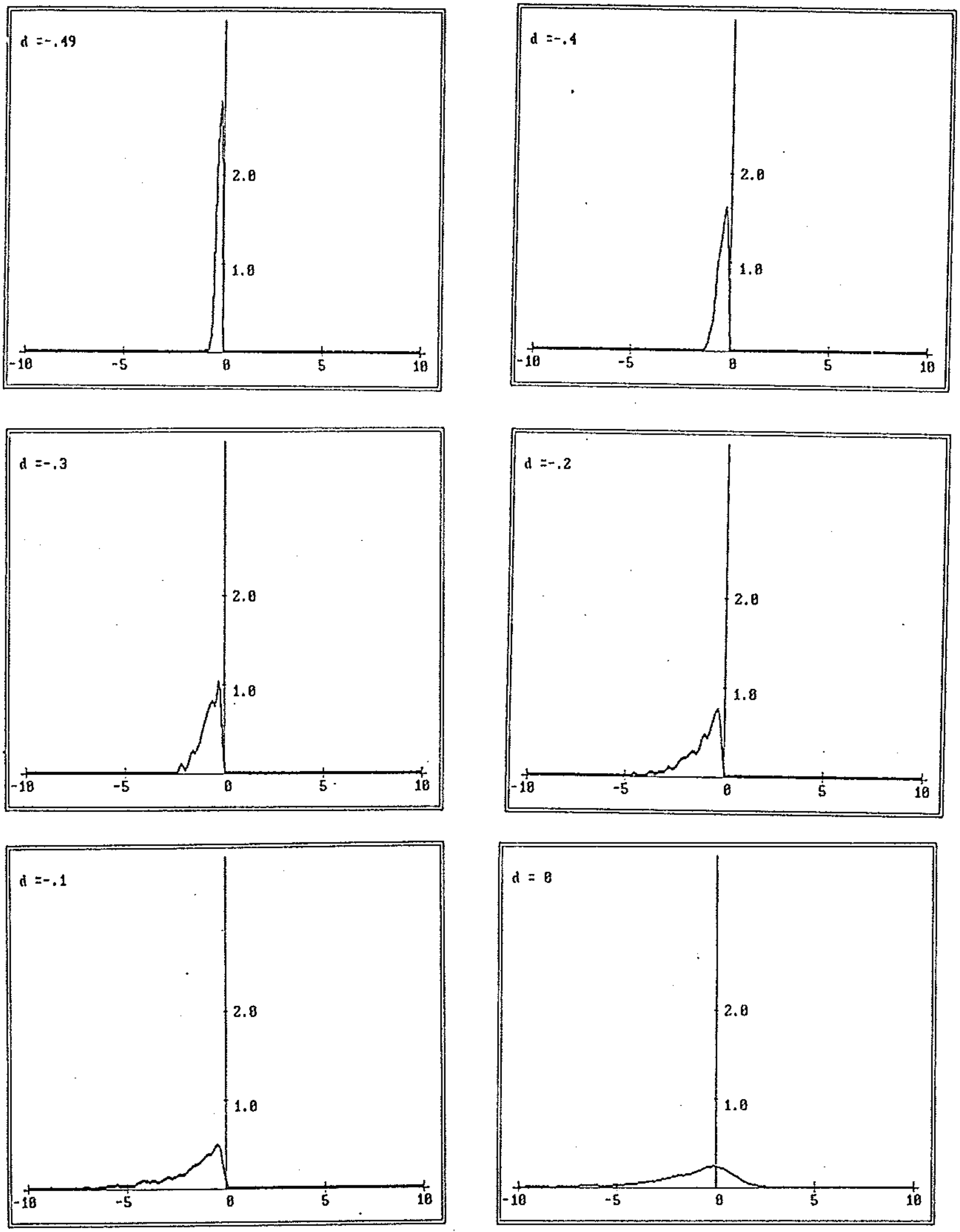

FIGURE 6

Estimates of the fractional unit root density for d = -0.0, -0.1, -0.2, -0.3, -0.4, -0.49. The plotted density is the kernel estimate of the density for $200(A_d+B_d)$ when d = 0.0. For $d \neq 0.0$ the plotted density is the kernel estimate of the density of $200B_d$.