A CONTINUOUS TIME APPROXIMATION TO THE UNSTABLE FIRST-ORDER AUTOREGRESSIVE PROCESS: THE CASE WITHOUT AN INTERCEPT

Pierre Perron
Princeton University and C.R.D.E.

Econometric Research Program
Research Memorandum No. 335

August 1988

*Assistant Professor, Department of Economics, Princeton University and Research Associate, Centre de Recherche et Développement en Economique (Université de Montréal). I have benefitted from comments by Peter Phillips. Research support is acknowledged from the Social Sciences and Humanities Research Council (Canada), the Fonds Concertés pour l'Aide à la Recherche (Québec) and the Natural Sciences and Engineering Research Council of Canada. This work was completed while the author was Assistant Professor at the Department of Economics, Université de Montréal.

Econometric Research Program
Princeton University
207 Dickinson Hall
Princeton, NJ 08544, USA

ABSTRACT

Consider the first–order autoregressive process $y_t = \alpha y_{t-1} + e_t$, y_0 a fixed constant, $e_t \sim$ i.i.d. $(0,\sigma^2)$ and let $\hat{\alpha}$ be the least–squares estimator of α based on a sample of size (T + 1) sampled at frequency h. Consider also the continuous time Orhnstein–Uhlenbeck process $dy_t = \theta y_t d_t + \sigma dw_t$ where w_t is a Weiner process and let $\hat{\theta}$ be the continuous time maximum likelihood (conditional upon y_0) estimator of $\hat{\theta}$ based upon a single path of data of length N. We first show that the exact distribution of $N(\hat{\theta} - \theta)$ is the same as the asymptotic distribution of $T(\hat{\alpha} - \alpha)$ as the sampling interval converges to zero. This asymptotic distribution permits explicit consideration of the effect of the initial condition y_0 upon the distribution of $\hat{\alpha}$. We use this fact to provide an approximation to the finite sample distribution of $\hat{\alpha}$ for arbitrary fixed y_0. The moment–generating function of $N(\hat{\theta} - \theta)$ is derived and used to tabulate the distribution and probability density function. We also derive the mean and mean–square error of $\hat{\theta}$ as well as the power function. In each case, the adequacy of the approximation to the finite sample distribution of $\hat{\alpha}$ is assessed for values of α in the vicinity of one. The approximations are, in general, found to be excellent.

Key Words: Orhnstein–Uhlenbeck process, moment–generating function, near–integrated processes, unit root, distribution theory.

1. INTRODUCTION

Consider the following first–order stochastic difference equation:

(1.1) $y_t = \alpha y_{t-1} + e_t$

where $y_0 = b$ a fixed constant and the e_t are independently and identically distributed $N(0,\sigma^2)$ variates. The unrestricted maximum likelihood estimate of α (conditional upon the initial observation y_0) based upon a sequence of observations of size T + 1, $\{y_t\}_0^T$, is the least squares estimator

$$\hat{\alpha} = \Sigma_{t=1}^T y_t y_{t-1}(\Sigma_{t=1}^T y_{t-1}^2)^{-1}.$$

The distribution of $\hat{\alpha}$ has been extensively studied. One topic of concern has been the adequacy of various asymptotic approximations to the finite sample distribution.

Mann and Wald (1943) and Rubin (1950) showed that $T^{1/2}(\hat{\alpha} - \alpha)(1 - \alpha^2)^{-1}$ has a limiting N(0,1) distribution when $|\alpha| < 1$. White (1958) showed that when $|\alpha| > 1$, the limiting distribution of $|\alpha|^T(\alpha^2 - 1)^{-1} (\hat{\alpha} - \alpha)$ is Cauchy provided that $y_0 = 0$. White also considered the case $|\alpha| = 1$ and showed that the limiting distribution of $T(\hat{\alpha} - \alpha)$ can be expressed in terms of the ratio of two functionals of a Weiner process (see also Phillips (1987)). In the latter case Rao (1976) obtained a computable expression.

The case of the unit root $\alpha = 1$ has attracted a great deal of attention. The critical values of the asymptotic distribution of $T(\hat{\alpha} - 1)$ have been derived by Dickey (1976) by simulation methods. Evans and Savin (1981a) obtained tabulated values of both the cumulative distribution function and the probability density function of the limiting distribution of $(T/\sqrt{2})(\hat{\alpha} - 1)$ by numerically integrating the limiting moment generating function derived by White (1958). A variety of procedures have also been suggested which permit a considerable weakening of the conditions imposed on the innovation sequence $\{e_t\}$ while at the same time using the same critical values derived in the earlier studies; see, for example, Dickey and Fuller (1979, 1981), Said and Dickey (1984) and Phillips (1987a).

The major point which emerges from these studies is that the asymptotic distributions have a discontinuity at one. Because the exact distribution of $\hat{\alpha}$ is continuous for all values of α, this suggests that the limiting distribution inadequately approximates the finite sample distribution for α near the discontinuity point $\alpha = 1$. This inadequacy has been well–documented by Evans and Savin (1981b, 1984) and by Phillips (1977) who also showed that Edgeworth expansions also perform poorly for α less than but close to one.

Recently, a new class of models which specifically deal with the presence of a root close to, but not necessarily equal to, one, have been studied. Phillips (1988) introduced the concept of a near–integrated random process where the autoregressive parameter is defined by:

(1.2) $\alpha = \exp(c/T)$.

Here, the real–valued constant c is a measure of the deviation from the unit root case. The models (1.1) and (1.2) may also be described as having a root local to unity (see Cavanagh (1986)): as the sample size increases, the autoregressive parameter converges to unity. When c is negative, the process $\{y_t\}$ is said to be (locally) stationary and when c is positive, it is said to be (locally) explosive. An expression for the limiting distribution of $T(\hat{\alpha} - \alpha)$ under (1.2) has been derived by Phillips (1988), Cavanagh (1986) and Chan and Wei (1987), and the analysis has been extended to a multivariate framework in Jeganathan (1987). This class of models has been quite useful in studying various problems such as the power of tests of a unit root under local alternatives (Phillips (1987b), Phillips and Perron (1988) and Perron (1986)), the derivation of confidence intervals when α is near unity (Cavanagh (1986)) and the calculation of the power of tests of a unit root with a continuum of observations (Perron (1987)).

Tabulations of the limiting distribution of $T(\hat{\alpha} - \alpha)$ under (1.2) have been obtained by Nabeya and Tanaka (1987), Cavanagh (1986) and Perron (1988) using different procedures. These studies also provide some measures of the adequacy of this limiting distribution to the finite sample distribution of $\hat{\alpha}$ when α is in the vicinity of 1. They show the approximation to be quite good in the case where $y_0 = 0$.

A feature of substantial interest is that the limiting distribution of $T(\hat{\alpha} - \alpha)$ under the near–integrated process (1.2) is invariant to the value of the initial observation y_0. As thoroughly documented by Evans and Savin (1981b), the finite sample distribution of $\hat{\alpha}$ is very sensitive to the value of y_0 when α is near 1 even for quite large sample sizes. Hence, the approximation provided by the 'near–integrated' limiting distribution is adequate only in the special case where $y_0 = 0$. When $y_0 \neq 0$, the approximation fails to capture the substantial effect a non–zero initial condition has on the finite sample distribution.

The purpose of this paper is to present a rather different approach in deriving an asymptotic approximation which permits the explicit consideration of the effects of different initial conditions on the distribution of $T(\hat{\alpha} - \alpha)$. Instead of the usual framework where the sample size increases to infinity keeping a fixed sampling interval, we use the so–called continuous records asymptotic analysis where T is increased to infinity keeping the span of the data fixed, i.e. by letting the sampling interval converge to zero at the same rate as T increases to infinity. This method is closely related to the exact distribution of estimators in continuous time.

Consider the following continuous time Orhnstein–Uhlenbeck diffusion process

(1.3) $$dy_t = \theta y_t dt + \sigma_0 dw_t; \qquad y_0 = b, \; t > 0.$$

w_t is the standard Weiner process. The maximum likelihood estimator of θ, conditional upon $y_0 = b$, is

$$\hat{\theta} = \int_0^N y_t dy_t / \int_0^N y_t^2 dt$$

where N is the span of the data. The discrete time representation of y_t in (1.3) is given by

(1.4) $\qquad y_{th} = \exp(\theta h)y_{(t-1)h} + u_{th}$

where $u_{th} \sim N(0,\sigma_0^2(e^{2\theta h} - 1)/2\theta)$ and h is the sampling interval. (1.4) is in the form of (1.1) with $\alpha = \exp(\theta h)$.

We derive the limiting distribution of $T(\hat{\alpha} - \alpha)$ under (1.4) as $h \to 0$ and show that it is identical to the exact distribution of $N(\hat{\theta} - \theta)$ under (1.3). Furthermore, this distribution contains an explicit dependence upon the initial condition y_0 and reduces to the near–integrated asymptotic distribution in the special case where $y_0 = 0$. These distributional results are considered in Section 2 which also discusses some interesting corollaries.

Section 3 derives the joint moment–generating function of $(N^{-1}\int_0^N y_t dw_t, N^{-2}\int_0^N y_t^2 dt)$. The expression obtained contains as a special case the limiting joint moment–generating function derived by Phillips for the 'near–integrated' framework by setting $y_0 = 0$. The result of White (1958) and Evans and Savin (1981a) is a special case obtained by setting $y_0 = 0$ and $\theta = 0$.

Section 4 uses the results of Section 3 to derive and tabulate the cumulative distribution and probability density function for selected combinations of the parameters θ and y_0, and provides some measure of the adequacy of the approximation to the discrete–time finite sample distribution of $\hat{\alpha}$.

Section 5 derives values for the bias and mean–squared error functions and discusses their use in approximating the exact mean and variance of $\hat{\alpha}$. The issues of power are discussed in Section 6 with special emphasis on the power of tests for a unit root against alternatives close to one. Finally, Section 7 offers remarks about the usefulness of our method to other frameworks. A mathematical appendix contains the proofs of some theorems.

Throughout the paper, we provide a measure of the adequacy of the asymptotic approximation to the finite sample counterpart. The various percentage points of the distribution, the moments of the distribution and the power functions are considered in turn. The results are quite striking, showing a very close approximation.

2. SOME DISTRIBUTIONAL RESULTS

We begin by restating our stochastic framework more formally. We consider an observable process $\{y_t, t \geq 0\}$ defined on a probability space $(\Omega, F, \mu_y^\theta)$. We denote by $\{F_t, t \geq 0\}$ a non–decreasing family of sub–σ–algebras of F such that y_t is F_t measurable. We denote by $\{w_t, t \geq 0\}$ the standard Weiner process; that is, a stochastic process with independent increments defined on a probability space (Ω, F, μ_w) with $w_0 = 0$ and where $w(t) - w(s)$ is $N(0, |t - s|)$. w_t is F_t measurable and the process $w_t(s) = w(t + s) - w(t)$, $(s \geq 0)$, is independent of F_t for any fixed $t \geq 0$. The measure μ_y^θ is induced by the following diffusion type process postulated for y_t:

$$dy_t = \theta y_t dt + \sigma dw_t \quad ; \quad y_0 = b, t \geq 0. \tag{2.1}$$

(2.1) is the standard Orhnstein–Uhlenbeck process. θ and σ are unknown parameters with $-\infty < \theta < \infty$ and $\sigma > 0$. The unique solution (μ_y^θ measure) in the mean squared sense to $\{y_t\}$ is given by (see, e.g., Arnold (1974)):

$$y_t = \exp(\theta t)b + \sigma \int_0^t \exp(\theta(t - s))dw_s. \tag{2.2}$$

The integral that appears in (2.2) is a stochastic integral. After Phillips (1988), we adopt the following notation: $J_\theta(t) = \int_0^t \exp(\theta(t - s))dw_s$. The solution to y_t is then written as:

$$y_t = \exp(\theta t)b + \sigma J_\theta(t). \tag{2.3}$$

Note that $J_\theta(t) \sim N(0,(e^{2\theta t} - 1)/2\theta)$. Our concern is the estimation of the unknown parameter θ given a single sample path of observations $\{y_t, 0 \leq t \leq N\}$, where

N is the span of the data. The analog to the least–squares procedure in continuous time yields the following estimator:

(2.4) $$\hat{\theta}_N(y) = \int_0^N y_t dy_t / \int_0^N y_t^2 dt.$$

$\hat{\theta}(y)$ is also the unrestricted maximum likelihood estimator when b = 0 (see, e.g., Liptser and Shiryayev (1978), 17.1.1). For simplicity of notation, we shall simply write $\hat{\theta} \equiv \hat{\theta}_N(y)$ and analyze the distributional properties of the standardized estimator $N(\hat{\theta} - \theta)$.

The discrete–time representation of the process y_t is easily shown to be given by:

(2.5) $$y_{th} = \exp(\theta h) y_{(t-1)h} + u_{th} \quad ; \quad y_0 = b, \ t \geq 0$$

where $u_{th} \sim N(0, \sigma^2(e^{2\theta h} - 1)/2\theta)$ and h is the sampling interval. In this discrete–time framework the goal is to estimate the unknown quantity $\alpha_h = \exp(\theta h)$ given a sequence of observations $\{y_{th}, t = 0, 1, 2, ..., T\}$ where $T = N/h$ is the total number of observations available (minus one given the initial condition y_0). The least square estimator (and maximum likelihood estimator conditional upon y_0) is

(2.6) $$\hat{\alpha}_h = \Sigma_{t=1}^T y_{th} y_{(t-1)h} / \Sigma_{t=1}^T y_{(t-1)h}^2$$

We focus on the asymptotic distribution of $T(\hat{\alpha}_h - \alpha_h)$ as $h \to 0$ given a fixed span N. For simplicity, we consider a limiting sequence $\{h = h_1, h_2, ..., h_n\}$ such that $T = N/h$ is integer–valued and we require that $h_n \to 0$ as $n \to \infty$.

The 'near–integrated' framework analyzed by Phillips (1988), Nabeya and Tanaka (1987), Perron (1988) and others specify the following process for y_t:

(2.7) $$y_t = \alpha \, y_{t-1} + e_t, \qquad y_0 = b, \ t \geq 0$$

with $e_t \sim N(0,\sigma^2)$ and $\alpha = \exp(c/T)$. Here, the statistic of interest is $T(\hat{\alpha} - \alpha)$ where $\hat{\alpha}$ is the least–squares estimator of α in (2.7) and the asymptotic distributional results are obtained letting $T \to \infty$. Note the close relationship between the 'near–integrated' framework and the discrete–time representation of the Orhnstein–Uhlenbeck process. The models are essentially the same if we interpret the parameter c as θN since one can write $\alpha_h = \exp(\theta h) = \exp(\theta N/T)$.

The main result of this section ties together:

- the exact distribution of $N(\hat{\theta} - \theta)$ given as sample path of observations $\{y_t\}_0^N$;
- the asymptotic distribution of $T(\hat{\alpha}_h - \alpha_h)$ with the continuous records asymptotic;
- the $(T \to \infty)$ asymptotic distribution of $T(\hat{\alpha} - \alpha)$ in the near–integrated framework.

Theorem 1

Let $A(\gamma,c) = \gamma\int_0^1 \exp(cr)dw_r + \int_0^1 J_c(r)dw_r$

and $B(\gamma,c) = \gamma^2(\exp(2c) - 1)/2c + \gamma\int_0^1 \exp(cr)J_c(r)dr + \int_0^1 J_c(r)^2dr,$

where $J_c(r) = \int_0^r e^{c(r-s)}dw_s$ and w_r the standard Weiner process defined on C(0,1).

i) Let $\{y_t, t \geq 0\}$ be a continuous–time stochastic process generated by (2.1) and let $\hat{\theta}$ be the estimator of θ defined by (2.4), then

$$N(\hat{\theta} - \theta) \overset{d}{=} A(\gamma,c)/B(\gamma,c) \equiv Z(\gamma,c)$$

with $\gamma = b/\sigma N^{1/2}$ and $c = \theta N$; and where $\overset{d}{=}$ signifies equality in distribution.

ii) Let $\{y_t, t \geq 0\}$ be a continuous time stochastic process generated by (2.1) and let $\hat{\alpha}_h$ be defined by (2.6) with $\{y_{th}, t = 0, ..., T\}$ generated by (2.5), then as $h \to 0$ with $T \to \infty$ and N fixed:

$$T(\hat{\alpha}_h - \alpha_h) \to Z(\gamma,c)$$

with, again, $\gamma = b/\sigma N^{1/2}$ and $c = \theta N$; and where $\to$ denotes weak convergence in distribution.

iii) Let $\{y_t, t = 1, ..., T\}$ be a stochastic process generated by (2.7) and $\hat{\alpha}$ be the least–squares estimator of α from that regression, then as $T \to \infty$:

$$T(\hat{\alpha} - \alpha) \to Z(0,c).$$

The proof of part (iii) is given in Phillips (1987b). The proof of part (ii) is closely related to the proof of Theorem 3.1 in Phillips (1987b) and is presented in the appendix.

To prove part (i), we consider a scale transformation from $t \ \varepsilon \ (0,N)$ to $r \ \varepsilon \ (0,1)$ such that $t \to Nr = t$. Then $J_\theta(t) = \int_0^t e^{\theta(t-s)}dw_s \sim N(0,(e^{2\theta t} - 1)/2\theta)$ can be expressed as $J_\theta(t) = N^{1/2}J_c(r)$ where $c = \theta N$. Using (2.3), we obtain the solution

$$y_r = \exp(cr)b + \sigma N^{1/2}J_c(r) \quad , \quad 0 \leq r \leq 1, \tag{2.8}$$

with initial condition $y_0 = b$. Now, the normalized estimator of $\hat{\theta}$ can be expressed as:

$$N(\hat{\theta} - \theta) = [\int_0^N y_t dy_t - \theta \int_0^N y_t^2 dt][N^{-1}\int_0^N y_t^2 dt]^{-1}$$

$$= \sigma \int_0^N y_t dw_t [N^{-1}\int_0^N y_t^2 dt]^{-1} \tag{2.9}$$

using (2.1). Consider first the numerator of (2.9). Using (2.8) and the fact that $w(t) = N^{1/2}w(r)$, we have:

$$\int_0^N y_t dw_t = \int_0^1 [\exp(cr)b + \sigma N^{1/2} J_c(r)] d(N^{1/2} w_r)$$

$$(2.10) \qquad = N^{1/2} b \int_0^1 \exp(cr) dw_r + N\sigma \int_0^1 J_c(r) dw_r.$$

The denominator of (2.9) can similarly be transformed as follows:

$$\int_0^N y_t^2 dt = \int_0^1 [\exp(cr)b + N^{1/2}\sigma J_c(r)]^2 d(Nr)$$

$$(2.11) \qquad = Nb^2(\exp(2c) - 1)/2c + 2N^{3/2}\sigma b \int_0^1 \exp(cr) J_c(r) dr$$

$$+ N^2\sigma^2 \int_0^1 J_c(r)^2 dr,$$

using the fact that $\int_0^1 \exp(2cr)dr = (\exp(2c) - 1)/2c$. The result follows from (2.9) to (2.11).

The theorem suggests several avenues of investigation. First, the small–h (or continuous records) asymptotic distribution of $T(\hat{\alpha}_h - \alpha_h)$ is the same as the exact distribution of the normalized continuous time estimator $N(\hat{\theta} - \theta)$. This justifies studying the distributional properties of $N(\hat{\theta} - \theta)$ as an approximation to the exact distribution of the discrete–time least–square estimator $T(\hat{\alpha} - \alpha)$. Since the stochastic representation of the variable $Z(\gamma,c)$ is explicitly affected by the value of the initial condition $y_0 = b$, we can hope to capture the effects of different values of b on the finite sample distribution of the discrete–time least–squares estimator. More precisely, the distribution of the random variable $Z(\gamma,c)$ depends only upon two parameters, namely $c = \theta N$ and $\gamma = b/\sigma N^{1/2}$. We shall therefore characterize the distributional properties of $N(\hat{\theta} - \theta)$, in latter sections, directly as functions of these two quantities.

Second, we obtain from Theorem 1 that the asymptotic distribution of $T(\hat{\alpha} - \alpha)$ in the near–integrated case is a special case of the exact distribution of $N(\hat{\theta} - \theta)$ obtained by setting $y_0 = b = 0$. Hence, it appears that when $b \neq 0$, we can expect a better approximation to the exact distribution of $\hat{\alpha}$ using the continuous record asymptotic rather than the near–integrated framework since in the latter case the effect of b vanishes asymptotically.

Several known distributional results can be obtained as special cases of the exact distribution of $N(\hat{\theta} - \theta)$. First when $b = 0$ we have

$$N(\hat{\theta} - \theta) \overset{d}{=} Z(0,c) = \int_0^1 J_c(r)dw(r) / \int_0^1 J_c(r)^2 dr$$

which is the asymptotic distribution of $T(\hat{\alpha} - \alpha)$ in the near–integrated case. Secondly, when $b = \theta = 0$, we have:

$$N(\hat{\theta} - \theta) = (1/2)(w(1)^2 - 1) / \int_0^1 w(r)^2 dr$$

which corresponds to the limiting distribution of $T(\hat{\alpha} - \alpha)$ in the unit root case where $\alpha = 1$, a result derived by Phillips (1987).

It is also straightforward to obtain different asymptotic results. Consider first the behavior of $N(\hat{\theta} - \theta)$ as $N \to \infty$. If $\theta < 0$, we have $c \to -\infty$ and if $\theta > 0$, $c \to +\infty$. Hence, the limiting distribution of $N(\hat{\theta} - \theta)$ as $N \to \infty$ depends upon the limiting distribution of the various functionals in $Z(\gamma,c)$ as $c \to \pm\infty$. These are stated in the following lemma.

<u>Lemma 1</u> Let $J_c(r) = \int_0^r e^{c(r-s)} dw_s$, then

i) $(-2c)\int_0^1 J_c(r)^2 dr \to 1$, as $c \to -\infty$;

ii) $(-2c)^{1/2}\int_0^1 J_c(r)dw_r \to N(0,1)$, as $c \to -\infty$;

iii) $(2c)^2 e^{-2c}\int_0^1 J_c(r)^2 dr \to \eta^2$, as $c \to +\infty$;

iv) $(2c)e^{-c}\int_0^1 J_c(r)^2 dw_r \to \xi\eta$, as $c \to +\infty$;

v) $(2c)^{3/2}e^{-2c}\int_0^1 \exp(cr)J_c(r)dr \to \eta$, as $c \to +\infty$;

vi) $(2c)^{1/2}e^{-c}\int_0^1 \exp(cr)dw_r \to 1$, as $c \to +\infty$;

where ξ and η are independent N(0,1) variates. Parts (i) through (iv) of Lemma 1 are proved in Phillips (1987b) and the proof of parts (v) and (vi) are presented in the appendix.

Using Lemma 1, it is straightforward to deduce the following results concerning the asymptotic distribution of $N(\hat{\theta} - \theta)$ as $N \to \infty$:

Corollary 1

i) If $\theta < 0$, then for any fixed $y_0 = b$:

$$N^{1/2}(\hat{\theta} - \theta) \to N(0, -1/2\theta), \qquad \text{as } N \to \infty;$$

ii) If $\theta > 0$, then for any fixed $y_0 = b$:

$$e^{\theta N}(\hat{\theta} - \theta)/2\theta \to [d\eta + \xi\eta]/[d + \eta]^2, \qquad \text{as } N \to \infty;$$

where η and ξ are independent N(0,1) variates and $d = b(2\theta)^{1/2}/\sigma$.

Remark: As a special case to part (ii), we have if $y_0 = b = 0$ and $\theta > 0$:

$$e^{\theta N}(\hat{\theta} - \theta)/2\theta \to \text{Cauchy}, \qquad \text{as } N \to \infty.$$

Part (i) of Corollary 1 is a standard result in the literature, see, for example, Brown and Hewitt (1975) and Basawa and Rao (1980). The above results provide a simple alternative derivation. Part (ii) appears not to have been derived previously. Note, however, that when $y_0 = b = 0$, the limiting distribution of $e^{\theta N}(\hat{\theta} - \theta)/2\theta$ is the same as the limiting distribution of $|\alpha|^T(\alpha^2 - 1)^{-1}(\hat{\alpha} - \alpha)$ when $|\alpha| > 1$ (see White (1958)).

It is also possible to study the behavior of $N(\hat{\theta} - \theta)$ as $b \to \infty$. It is a straightforward consequence of Theorem 1 that $N(\hat{\theta} - \theta) \to 0$ as $b \to \infty$. Hence, one can expect the bias and variance of $\hat{\theta}$ to decrease as b increases. The same behavior should hold concerning the discrete-time estimator $\hat{\alpha}$. These conjectures are verified in

Section 5. It is also possible to obtain the following approximating distribution of $N(\hat{\theta} - \theta)$ for large γ (i.e. large b):

$$N(\hat{\theta} - \theta) \sim \int_0^1 \exp(cr)dw_r[\gamma(\exp(2c) - 1)/2c]^{-1}$$

Now $\int_0^1 \exp(cr)dw(r) \sim N(0,(e^{2c} - 1)/2c)$ (see the appendix), hence,

$$N(\hat{\theta} - \theta) \sim N(0,2c/\gamma^2(e^{2c} - 1)).$$

Given that the distribution of $N(\hat{\theta} - \theta)$ approximates the distribution of $T(\hat{\alpha}_h - \alpha_h)$, we have the following approximation to the discrete–time estimator $\hat{\alpha}_h$ for large b (or equivalently small σ asymptotic):

$$T(\hat{\alpha}_h - \alpha_h) \sim N(0,2c/\gamma^2(e^{2c} - 1)). \tag{2.12}$$

The approximation (2.12) generalizes a result derived by Phillips (1987) who considered the unit root case. Indeed, by taking the limit of (2.12) as $c \to 0$, we have $T(\hat{\alpha}_h - 1) \sim N(0,1/\gamma^2)$ which correspond to expression (27) in Phillips (1987).

As noted in Phillips (1987) for the unit root case, the usual $(T \to \infty)$ asymptotic distribution obscures the dependence of the distribution of $\hat{\alpha}$ on the parameter y_0/σ. The same feature emerges in the $(T \to \infty)$ asymptotic distribution for the near–integrated case. The $(h \to 0)$ asymptotic distribution highlights the effects of the initial condition and may be a better guide to the finite sample behavior of $T(\hat{\alpha} - \alpha)$ for values of α in the vicinity of 1.

In order to use these distributional results as approximations to the finite sample behavior, we need a method to compute the distribution function of the random variable $Z(\gamma,c)$ or equivalently of $N(\hat{\theta} - \theta)$. To this effect, the next section derives a computable expression for this distribution.

3. THE EXACT DISTRIBUTION OF $N(\hat{\theta} - \theta)$

Given that the exact distribution of $N(\hat{\theta} - \theta)$ depends only upon the parameters $c = \theta N$ and $\gamma = b/\sigma N^{1/2}$, it is useful to transform the original model. The transformation $t \in (0,N) \to r \in (0,1)$ with $t = Nr$ yields

$$N(\hat{\theta} - \theta) = \sigma N^{-1} \int_0^N y_t dw_t \,/ N^{-2} \int_0^N y_t^2 dt$$

$$= \sigma N^{-1/2} \int_0^1 y_r dw_r \,/ N^{-1} \int_0^1 y_r^2 dr$$

where y_r is defined by (2.8). Now, let $x_r = y_r / \sigma N^{1/2}$. Then,

(3.1) $$N(\hat{\theta} - \theta) = \int_0^1 x_r dw_r \,/ \int_0^1 x_r^2 dr$$

where

(3.2) $$x_r = \exp(cr)\gamma + J_c(r).$$

The expression (3.2) is the solution x_r of the following stochastic differential equation:

(3.3) $$dx_r = cx_r dr + dw_r, \quad x_0 = \gamma, \quad 0 \le r \le 1,$$

where $c = \theta N$ and $\gamma = b/\sigma N^{1/2}$.

To study the exact distribution of $N(\hat{\theta} - \theta)$, we derive the joint moment generating function of $(\int_0^1 x_r dw_r, \int_0^1 x_r^2 dr)$ given that x_r is a random variable in the probability space $(\Omega, F, \mu_y^\theta)$ generated according to the diffusion process (3.3). We denote this joint moment generating function:

$$M_{c,\gamma}(v,u) = E[\exp(v \int_0^1 x_r dw_r + u \int_0^1 x_r^2 dr)]$$

where the expectation is taken with respect to the measure $\mu_y^\theta \equiv \mu_x^c$. The main result of this section is an expression for $M_{c,\gamma}(v,u)$ which is contained in the following theorem.

Theorem 2

Let $$\psi_c(v,u) = \left[\frac{2\lambda e^{-(v+c)}}{[\lambda + v + c]e^{-\lambda} + [\lambda - (v + c)]e^{\lambda}}\right]^{1/2}$$

where $\lambda = (c^2 + 2cv - 2u)^{1/2}$

then

$$M_{c,\gamma}(v,u) = \psi_c(v,u) \exp\{-(\gamma^2/2)(v + c - \lambda)(1 - \exp(v + c + \lambda)\psi_c^2(v,u))\}.$$

Remark: when $\gamma = 0$, $M_{c,\gamma}(v,u) = \psi_c(v,u)$ which is the joint moment generating function of $(\int_0^1 J_c(r)dw_r, \int_0^1 J_c(r)^2 dr)$ derived by Phillips (1987b).

The proof of the theorem is inspired by the development in Lemma 17.3 of Liptser and Shiryayev (1978). Denote by μ_x^c and μ_x^λ the measures corresponding to the processes x^c and x^λ generated by the following stochastic differential equations:

(3.4) $dx_r^c = cx_r^c dr + dw_r, \quad x_0^c = \gamma$

(3.5) $dx_r^\lambda = \lambda x_r^\lambda dr + dw_r, \quad x_0^\lambda = \gamma.$

The measures μ_x^c and μ_x^λ are equivalent (Liptser and Shiryayev (1978), Theorem 7.19) and the density or Radon–Nikodym derivative $d\mu_x^c/d\mu_x^\lambda$ evaluated with the random process x^λ is given by:

$$\frac{d\mu_x^c}{d\mu_x^\lambda}(x^\lambda) = \exp\{(c-\lambda)\int_0^1 x_r^\lambda dx_r^\lambda - \frac{(c^2-\lambda^2)}{2}\int_0^1 (x_r^\lambda)^2 dr\}.$$

Hence, denoting by E_c, the expectation operator taken with respect to the measure μ_x^c, we have:

$$M_{c,\gamma}(v,u) = E_c[\exp\{v\int_0^1 x_r dw_r + u\int_0^1 x_r^2 dr\}]$$

$$= E[\exp\{v\int_0^1 x_r^c dw_r + u\int_0^1 (x_r^c)^2 dr\}]$$

$$= E[\exp\{v\int_0^1 x_r^\lambda dw_r + u\int_0^1 (x_r^\lambda)^2 dr\}] \cdot \frac{d\mu_x^c}{d\mu_x^\lambda}(x^\lambda)$$

$$= E[\exp\{v\int_0^1 x_r^\lambda dw_r + u\int_0^1 (x_r^\lambda)^2 dr$$

$$+ (c-\lambda)\int_0^1 x_r^\lambda dx_r^\lambda - \frac{(c^2-\lambda^2)}{2}\int_0^1 (x_r^\lambda)^2 dr\}].$$

Using the fact that $\int_0^1 x_r^\lambda dw_r = \int_0^1 x_r^\lambda dx_r^\lambda - \lambda\int_0^1 (x_r^\lambda)^2 dr$ with (3.5) and rearranging:

$$M_{c,\gamma}(v,u) = E[\exp\{(v+c-\lambda)\int_0^1 x_r^\lambda dx_r^\lambda + [u - vc - (c^2-\lambda^2)/2]\int_0^1 (x_r^\lambda)^2 dr\}].$$

Now, let $\lambda = (c^2 + 2cv - 2u)^{1/2}$ and $a = v + c - \lambda$. Then, $[u - vc - (c^2\lambda^2)/2] = 0$ and

$$M_{c,\gamma}(v,u) = E[\exp\{a\int_0^1 x_r^\lambda dx_r^\lambda\}].$$

By Ito's lemma applied to the function $(x_r^\lambda)^2$, we have $\int_0^1 (x_r^\lambda) dx_r^\lambda = (x_1^\lambda - x_0^\lambda - 1)/2$. Hence,

$$M_{c,\gamma}(v,u) = E[\exp\{(a/2)((x_1^\lambda)^2 - (x_0^\lambda)^2)\}]$$

(3.6) $$= \exp\{(-a/2)(\gamma^2 + 1)\}E[\exp\{(a/2)(x_1^\lambda)^2\}].$$

since $x_0^\lambda = \gamma$, a fixed constant. Now, the solution to (3.5) is given by:

$$x_r^\lambda = \exp(\lambda r)\gamma + \int_0^r e^{\lambda(r-s)}dw_s$$

$$= \exp(\lambda r)\gamma + J_\lambda(r)$$

and, hence,

$$x_1^\lambda = \exp(\lambda)\gamma + J_\lambda(1).$$

Now $J_\lambda(1) \sim N(0,(e^{2\lambda} - 1)/2\lambda)$ and therefore x_1^λ is $N(\exp(\lambda)\gamma, (e^{2\lambda} - 1)/2\lambda)$. Let $s^2 = (e^{2\lambda} - 1)/2\lambda$ and $q = \exp(\lambda)\gamma$ then $(x_1^\lambda)^2/s^2$ is distributed as a non–central chi–squared variate with non–centrality parameter q. Then

$$E[\exp\{(a/2)(x_1^\lambda)^2\}] = E[\exp\{(a/2)s^2(x_1^\lambda/s)^2\}]$$

$$= (1 - as^2)^{-1/2}\exp[(aq^2/2)(1 - as^2)^{-1}]$$

and

$$M_{c,\gamma}(v,u) = \exp\{(-a/2)(\gamma^2 + 1)\}(1 - as^2)^{-1/2}\exp\{(aq^2/2)(1 - as^2)^{-1}\}.$$

Upon substitution for $s = (e^{2\lambda} - 1)/2\lambda$ and $a = v + c - \lambda$, it is straightforward to verify that

$$\psi_c(v,u) = (1 - as^2)^{-1/2}\exp(-a/2).$$

It follows

$$M_{c,\gamma}(v,u) = \psi_c(v,u)\exp\{(-a\gamma^2/2) + (aq^2/2)(1 - as^2)^{-1}\}$$

$$= \psi_c(v,u)\exp\{(-a\gamma^2/2)(1 - \exp(2\lambda + a)\psi_c(v,u)^2)\}$$

which reduces to the expression in Theorem 1 upon substitution for $a = v + c - \lambda$. □

The moment–generating function derived in Theorem 2 allows us to compute the cumulative distribution function as well as other distributional quantities using numerical integration. The next sections present the results of such computations.

4. THE COMPUTATION OF THE C.D.F. AND P.D.F. OF $N(\hat{\theta} - \theta)$

Denote the joint characteristic function of $(\int_0^1 x_r dw_r, \int_0^1 x_r^2 dr)$ by $cf_{c,\gamma}(v,u)$. Then,

$$cf_{c,\gamma}(v,u) = M_{c,\gamma}(iv,iu) \tag{4.1}$$

$$= E[\exp(iv\int_0^1 x_r dw_r + iu\int_0^1 x_r^2 dr)].$$

The distribution function of $N(\hat{\theta} - \theta)$ can be obtained as follows. Let $F_{c,\gamma}(z) = P[N(\hat{\theta} - \theta) \le z]$ and recall that $N(\hat{\theta} - \theta) = \int_0^1 x_r dw_r / \int_0^1 x_r^2 dr$. Then by Theorem 1 of Gurland (1948), we have

$$F_{c,\gamma}(z) = \frac{1}{2} - \frac{1}{2\pi i} \lim_{\substack{\varepsilon_1 \to 0 \\ \varepsilon_2 \to \infty}} \int_{\varepsilon_1 < |v| < \varepsilon_2} \frac{cf_{c,\gamma}(v, -vz)}{v} dv$$

$$= \frac{1}{2} - \frac{1}{2\pi} \int_0^\infty \mathrm{AIMAG}\left[\frac{cf_{c,\gamma}(v, -vz)}{v}\right] dv \tag{4.2}$$

where AIMAG$(\cdot)$ denotes the imaginary part of the complex number. Further, the density function is derived as follows:

$$f_{c,\gamma}(z) = \frac{d}{dz} F_{c,\gamma}(z)$$

$$= \frac{1}{2\pi i} \lim_{\substack{\varepsilon_1 \to 0 \\ \varepsilon_2 \to \infty}} \int_{\varepsilon_1 < |v| < \varepsilon_2} \left. \frac{\partial cf_{c,\gamma}(v,u)}{\partial u} \right|_{u = -vz} dv$$

$$= \frac{1}{\pi} \int_0^\infty \mathrm{Real}[cf_{c,\gamma}(v, -vz)(A1 + A2/2)]\, dv \tag{4.3}$$

where $A1 = (-\gamma^2/2)\ \{(1 - \exp(iv + c + \lambda)[\psi_c(iv, -ivz)]^2)$

$$+ (iv + c - \lambda)\exp(iv + c + \lambda)\psi_c(iv, - ivz)(-\lambda^{-1} + A2)\}$$

and $$A2 = -\lambda^{-2} + D1/D2$$

$$D1 = \lambda^{-1}e^{-\lambda}[\lambda + (iv + c) -1] + \lambda^{-1}e^{\lambda}[\lambda - (iv + c) +1]$$

$$D2 = [\lambda + (iv + c)]e^{-\lambda} + [\lambda - (iv + c)]e^{\lambda}$$

$$\lambda =(c^2 + 2iv(c + z))^{1/2}$$

and Real($\cdot$) denotes the real part of the complex number.

The expressions (4.2) and (4.3) can be numerically integrated to obtain values for the cumulative distribution function and the probability density function. When calculating these functions, we evaluate the integrals in the range $(0 + \varepsilon, \bar{V})$ where $\bar{V}$ is an upper bound set such that the integrand evaluated at $\bar{V}$ is less than ε (ε was set at 1.0E–07 in each integration). The integrals are then evaluated in this range using the subroutine DCADRE of the International Mathematical and Statistical Library (IMSL) (the error criterion for this routine integration was also set at 1.0E–07). Special care, however, must be applied to the integration since it involves the square root of a complex valued function. The use of the principal value of the square root may not ensure the continuity of the intergrand. We must therefore integrate over the Reiman surface consisting here of two planes. The method used is described in more details in the Appendix to Perron (1988).

Figures 1 through 3 present the cumulative distribution functions of $N(\hat{\theta} - \theta)$ for $c = -5$, 0 and 2 respectively. Each figure contains four curves corresponding to the following values for γ: 0, 0.5, 1.0 and 2.0. Figures 4 through 6 contain the corresponding results for the probability density function. Each curve was constructed by evaluating the c.d.f. (or p.d.f.) at 100 equidistant points in the relevant range.

Several striking features emerge from these graphs. First, the distributions are sharply non–normal. Second, the shapes are significantly affected by changes in the values of γ and c. The most important feature is that the distribution becomes more closely centered and more concentrated near zero as γ and c increase. As c increases, the distribution becomes less skewed to the left and shows much less variability around zero. A similar behavior occurs as γ increases. Another interesting feature is the effect of increasing γ for different values of c. For a given change in γ, the effect on the shape of the distribution is more pronounced the larger c is. In other words, the effects of different values of γ are smaller when θ is smaller (i.e. "more stationary") or when N is larger, provided $\theta < 0$. When $\theta > 0$, the effect of the initial condition is considerable.

To summarize the main features of the various distributions, Table 1 presents selected percentage points of the distribution of $N(\hat{\theta} - \theta)$ for various values of c and γ.

To assess the adequacy of the asymptotic approximation to the finite sample distribution, we proceed as follows. We simulated the least–squares estimator of $\hat{\alpha}$ from an AR(1) model with N(0,1) errors. The number of replications was set at 20,000. We obtained the percentage points of the distribution of $T(\hat{\alpha} - \alpha)$ for selected values of T, α and $y_0 = b$. The correspondence between the exact distribution and the asymptotic approximation is obtained by setting $c = T\ln(\alpha)$ (since $\alpha = \exp(\theta N/T)$) and $\gamma = b/T^{1/2}$ since $\sigma^2 = 1$. Table 2 presents the results for T = 10, 25, 50 and 100 with c = –5, –2, 0 and 2 and γ = 0.5, 1.0, 2.0. We omit the case where γ = 0.0 since it was thoroughly analysed in Nabeya and Tanaka (1987), Perron (1988) and Cavanagh (1986). These studies found the approximation to be adequate for various values of c. The approximation deteriorates as c becomes smaller.

In our more general context, the approximation remains quite good especially for values of T at 50 or 100. The approximation gets more accurate as 1) c increases, 2) γ increases and 3) if we consider the right tail rather than the left. It is also relatively less accurate in the extreme tails. For example, consider the case with c = 2, γ = 2 and the 5th percentage point. The asymptotic approximation is –0.3294 and the exact values range from –0.338 for T = 10 to –0.325 for T = 100, showing a high degree of accuracy even for a very small sample size.

On the other hand, when c is small, the convergence to the asymptotic value is slower. For example, consider c = –5, γ = 0.5 and the 5th percentage point. The asymptotic critical value is –8.755 while the exact values range from –5.442 (T = 10) to –8.314 (T = 100). However, even for small values of c, it gets more accurate as γ

increases. For example, when $c = -5.0$, $\gamma = 2.0$, the asymptotic critical value for the 95th percentage point is 1.7357 and the exact values range from 1.464 ($T = 10$) to 1.675 ($T = 100$).

Overall, the asymptotic distribution approximates rather well the major changes in the distribution as c and γ are varied and for sample sizes typically available the approximation is rather good.

5. THE BIAS AND MEAN–SQUARED ERROR OF $\hat{\theta}$

The moment–generating function derived in Section 3 can be used to derive the moments of $N(\hat{\theta} - \theta)$. Using Mehta and Swamy's (1978) result, we have:

$$\text{(5.1)} \qquad N \cdot \mathrm{Bias}(\hat{\theta}) = E[N(\hat{\theta} - \theta)] = \int_0^\infty \left. \frac{\partial M_{c,\gamma}(v,-u)}{\partial v} \right|_{v=0} du$$

and

$$\text{(5.2)} \qquad N^2 \cdot \mathrm{MSE}(\hat{\theta}) = E[N(\hat{\theta} - \theta)]^2 = \int_0^\infty u \left. \frac{\partial^2 M_{c,\gamma}(v,-u)}{\partial v} \right|_{v=0} du$$

where $M_{c,\gamma}(v,u) = E[\exp(v\int_0^1 x_r dw_r + u\int_0^1 x_r^2 dr)]$

is the joint moment–generating function of $(\int_0^1 x_r dw_r, \int_0^1 x_r^2 dr)$ derived in Theorem 2.

Straightforward computations give the following results:

$$\left. \frac{\partial M_{c,\gamma}(v,u)}{\partial v} \right|_{v=0} = M_{c,\gamma}(0,-u)[DH + [(DA/A) - (DB/B)]/2]$$

where
$$DH = -\gamma^2(1 - c/\lambda)(1 - \exp(c + \lambda)[\psi_c(0,-u)]^2)/2$$
$$+ (1/2)(c - \lambda)\gamma^2 \exp(c + \lambda)[\psi_c(0,-u)]^2\{(1 + c/\lambda) + DA/A - DB/B\}$$

$$A = 2\lambda e^{-c}$$

$$B = [\lambda + c]e^{-\lambda} + [\lambda - c]e^{\lambda}$$

$$DA = 2e^{-c}[c - \lambda^2]/\lambda$$

$$DB = (1 - c)[e^{-\lambda}(c + \lambda) + e^{\lambda}(c - \lambda)]/\lambda$$

$$\lambda = (c^2 + 2u)^{1/2}$$

and

$$\left.\frac{\partial^2 M_{c,\gamma}(v,-u)}{\partial v^2}\right|_{v=0} = M_{c,\gamma}(0,-u)\{DH + [(DA/A) - (DB/B)]/2\}^2$$

$$+ M_{c,\gamma}(0,-u)\{D2H + (D2A - D2B)/2\}$$

where $D2A = -2c^2/\lambda^4$

$$D2B = DBW/B - (DB)^2/B^2$$

$$D2W = ce^{-\lambda}[\lambda(c-2) + c(c-1) + c^2/\lambda]/\lambda^2$$

$$+ ce^{\lambda}[\lambda(c-2) - c(c-1) + c^2/\lambda]/\lambda^2$$

and

$$D2H = DH(1 - c/\lambda)/(c - \lambda) + H[c^2/\lambda^3(c-\lambda) - 1/\lambda^2]$$

$$+ [DH + (1 - c/\lambda)\gamma^2/2][(1 + c/\lambda) + DA/A - DB/B]$$

$$+ [H + (c - \lambda)\gamma^2/2][-c^2/\lambda^3 - 2c^2/\lambda^4 - DBW/B + (DB)^2/B^2].$$

Figures 7 and 8 present, respectively, the normalized bias and mean–squared error of $\hat{\theta}$ (i.e. the first two moments of $N(\hat{\theta} - \theta)$) as a function of $c = \theta N$ for various values of $\gamma = b/\sigma N^{1/2}$. The conjectures discussed in Section 2 are verified. The bias and mean–squared error are smaller (in absolute value) as c increases. Furthermore, both decrease (in absolute value) as γ increases. An increase in γ has a larger reduction effect on both the bias and the mean squared error when c is small. But this is mainly due to the fact that the bias and MSE are already quite small for large c when $\gamma = 0$.

Both the bias and MSE converge quite rapidly to zero as c and/or γ increase. On the other hand, the bias appears to attain an upper limit (in absolute) value when $c \to -\infty$ for any fixed γ and it approaches this limit fairly rapidly. The limit of the bias as $c \to -\infty$ is a decreasing (in absolute value) function of γ.

To assess the relevance of the asymptotic approximation to the exact moments of $\hat{\alpha}$ we consider applying our approximation to the experimental setting in Table III of Evans and Savin (1981b) which presents the exact mean and variance of a normalized estimator of $\hat{\alpha}$; $\alpha = 0.90, 0.99, 1.00, 1.01$ and 1.05, $y_0/\sigma = 0, 4, 16$ and $T = 25, 50, 100, 200$ and 400. The normalization is defined as $g(T)(\hat{\alpha} - \alpha)$ where

$$g(T) = \begin{cases} [T(1-\alpha^2)^{-1}]^{1/2} & |\alpha| < 1; \\ T/\sqrt{2} & |\alpha| = 1; \\ |\alpha|^T(\alpha^2 - 1)^{-1} & |\alpha| > 1. \end{cases}$$

To approximate the moments of $g(T)(\hat{\alpha} - \alpha)$, we consider the correspondence $c = T\ell n(\alpha)$ and $\gamma = y_0/\sigma T^{1/2}$ and use (5.1) and (5.2).

The results are presented in Tables 3 and 4 for the mean and the variance of $g(T)(\hat{\alpha} - \alpha)$ respectively. The values in parentheses are taken from Table III of Evans and Savin (1981b).

The approximation is, in general, very good and certainly captures most of the changes in the distribution as α, y_0/σ and T are varied. As expected, the approximation is better as α is closer to 1 and as T is larger. But it is also better as y_0/σ increases.

We can also compare our approximation to White's approximation when $y_0 = 0$ and $\alpha < 1$ (presented in Table III of Evans and Savin (1981b)). Our approximation is better when $\alpha = 0.99$ and slightly worse when $\alpha = 0.90$. Our approximation, however, has the advantage of capturing the effects of a non–zero initial condition.

Overall, our procedure appears to provide a very adequate approximation to the exact moments of $\hat{\alpha}$ when α is in the vicinity of one allowing the initial condition y_0/σ to be an arbitrary fixed value.

6. THE POWER FUNCTION

Using the statistic $N(\hat{\theta} - \theta_0)$, the results of Section 2 can also be used to analyze the power function for tests of the null hypothesis

(6.1) $\quad H_0\colon\ \theta = \theta_0$

against various alternatives. Denote by z^* the value such that $P_{\theta_0}[N(\hat{\theta} - \theta_0) \leq z^*] = \beta$. Then, the power function of a one-sided test with size β for testing $\theta = \theta_0$ against $\theta < \theta_0$ is given by

(6.2)
$$P_\theta[N(\hat{\theta} - \theta_0) < z^*] = P_\theta[N(\hat{\theta} - \theta) < z^* + N(\hat{\theta} - \theta_0)]$$
$$= P_\theta[Z(\gamma,c) < z^* + (c - c_0)]$$

with $c = N\theta$ and $c_0 = N\theta_0$. Expression (4.2) can be used to evaluate the power function (6.2) for various values of c, γ, c_0 and significance level β.

Much attention has recently been given to testing the null hypothesis of a unit root ($\alpha = 1$, $\theta = 0$). Letting $\theta_0 = 0$ and using (6.2), the power function of a one-sided test of the unit root hypothesis using the statistic $N\hat{\theta}$ is given by

(6.3)
$$P_\theta[N\hat{\theta} < z_0^*] = P_\theta[N(\hat{\theta} - \theta) < z_0^* - c]$$
$$= P_\theta[Z(\gamma,c) < z_0^* - c]$$

which can be evaluated using (4.2). Here, z_0^* is the critical value such that $P[Z(\gamma,0) \leq z_0^*] = \beta$ where β is the size of the test.

Interestingly, (6.3) gives the local asymptotic power function of a test for a unit root using the statistic $T(\hat{\alpha} - 1)$ (in discrete–time) where the sequence of local alternatives is given by

$$\alpha = \exp(c/T)$$

and the asymptotic is derived as $h \to 0$, i.e. using the continuous records asymptotic instead of the usual $(T \to \infty)$ asymptotic. To elaborate, consider the following derivation:

$$\begin{aligned} \lim_{h \to 0} P_\alpha[T(\hat{\alpha} - 1) < z_0^*] &= \lim_{h \to 0} P_\alpha[T(\hat{\alpha} - \alpha) < z_0^* + T(\alpha - 1)] \\ &= P[Z(\gamma,c) < z_0^* - c] \end{aligned}$$

since $T(\hat{\alpha} - \alpha) \to Z(\gamma,c)$ as $h \to 0$ (Theorem 1) and $T(\alpha - 1) \to c$ as $h \to 0$.

This framework generalizes the analysis of Phillips (1987b) who considers the local asymptotic power in the usual $(T \to \infty)$ asymptotic context. Here, the continuous records asymptotic $(h \to 0)$ permits explicit consideration of the effect of a non–zero initial condition. Our results specialize to those of Phillips by setting $\gamma = 0$. Note finally, that when $\gamma = 0$, the critical values z_0^* correspond to those derived by Dickey and Fuller (1979); see Fuller (1976).

Figure 9 presents the power function of a one–sided 5 % size test that $\theta = 0$ against both positive and negative alternative values of $c = \theta N$. Four curves are drawn corresponding to values of γ at 0, 0.5, 1.0 and 2.0. Several features emerge. First, as expected from the studies of Evans and Savin (1981b) and others, the power is much higher against explosive alternatives $(c > 0)$. Secondly, the power function is significantly influenced by different values of the startup condition γ. Indeed, the power increases as γ increases. There is an interesting analogy with the bias function (see figure 7) in that changes in γ have a larger effect on the power function for stationary alternatives. This is due to the fact that when $\gamma = 0$ the power function against explosive alternatives is already quite high compared to the power function under stationary alternatives.

To assess the adequacy of our asymptotic power function in the context of finite samples tests for a unit root against alternatives close to one, we considered the experimental framework adopted in Evans and Savin (1981b). The exact power of tests for a unit root against the alternatives $\alpha = 0.9, 0.95, 0.99, 1.01, 1.025$ and 1.05 can be found in their Table IV. For each value of α, they vary as the initial condition y_0/σ = 0, 2 and 4 and T = 25, 50, 100, 200, and 400. Our results are presented in Table 5 where, for ease of comparison, we have included in parentheses the exact values derived by Evans and Savin (1981b). The correspondence between the finite sample and asymptotic frameworks is again achieved by specifying $c = T\ \ell n(\alpha)$ and $\gamma = y_0/\sigma T^{1/2}$.

The results are quite striking. For almost all combinations of parameters the approximation is indeed excellent. The worst fit occurs for a small sample size (T = 25) and large values of α (1.05). The conclusions reached by Evans and Savin (1981b) are well–captured by the asymptotic approximations.

7. CONCLUDING COMMENTS

This paper has considered an alternative asymptotic approximation to the least-squares estimator of the parameter in a first-order autoregressive model with a fixed initial condition. This approach is different in that we consider a framework using the continuous records asymptotics ($h \to 0$) instead of the usual ($T \to \infty$) asymptotic theory. The main advantage over the ($T \to \infty$) near-integrated framework is that our method allows explicit consideration of the effects of non-zero initial conditions.

The continuous record asymptotic distribution for the discrete-time estimator is shown to be equivalent to the distribution of the continuous time estimator in an Orhnstein-Uhlenbeck process. The exact distribution of this continuous time estimator was derived. It was found to yield interesting theoretical results for continuous time estimation. More importantly, it has proved to be a good approximation to the finite sample distribution of the discrete-time least squares estimator.

This study indicates the usefulness of the continuous records asymptotic distributional theory which can be used in other contexts. One such possible extension is the possibility to approximate the finite sample distribution of the least-squares estimator $\hat{\alpha}$ when a random initial condition is specified. Indeed, equation (3.6) can be used to derive the moment-generating function of $N(\hat{\theta} - \theta)$ with a random initial condition as well. An extensive study of the case where the stochastic process $\{y_t\}$ is stationary is presented in Perron (1988b).

Further possible extensions include the case where a fitted mean is included in the regression. Here, the parameter affecting the shape of the distribution is a normalization of the mean of the variables instead of the initial condition per se (see Evans and Savin (1984)). Again, the analysis can distinguish between a fixed and random initial condition. Such an analysis is presented in a companion paper, Perron (1988c). It remains to be seen whether analogs to Theorem 2 can be obtained in the context of more general models and different statistics such as the t-ratio.

MATHEMATICAL APPENDIX

Proof of Theorem 1, part (ii)

The proof of Theorem 1, part (ii) is closely related to the proof of Theorem 3.1 in Phillips (1987b). We consider a triangular array of random variables $\{\{y_{nt}\}_{t=1}^{T_n}\}_{n=1}^{\infty}$. For a given n, the sequence $\{y_{nt}\}_{n=1}^{T_n}$ is generated by (see equation (2.5))

(A.1) $\quad y_{nt} = \exp(\theta h_n) y_{n(t-1)} + u_{nt}, \qquad t = 1, ..., T_n$

where the innovation sequence $\{u_{nt}\}_{t=1}^{T_n}$ is i.i.d. normal with mean 0 and variance $\sigma^2(\exp(2\theta h_n) - 1)/2\theta$. We have $T_n = N/h_n$ and $T_n \to \infty$, $h_n \to 0$ as $n \to \infty$.

Let the sequence $\{\{\varepsilon_{nt}\}_1^{T_n}\}_1^{\infty}$ be defined by $\varepsilon_{nt} = a(h_n)^{-1/2} u_{nt}$ where $a(h_n) = (\exp(2\theta h_n) - 1)/2\theta$. Then $\{\{\varepsilon_{nt}\}_1^{T_n}\}_1^{\infty}$ is a triangular array of i.i.d. $N(0,\sigma^2)$ variates. Note that $\lim_{h_n \to 0} h_n^{-1} a(h_n) = 1$. Now from (A.1), we have:

(A.2) $$y_{nt} = \textstyle\sum_{j=1}^{t} \exp((t-j)\theta h_n)\, u_{nj} + \exp(t\theta h_n)\, y_0$$

$$= a(h_n)^{1/2} \textstyle\sum_{j=1}^{t} \exp((t-j)\theta h_n)\, \varepsilon_{nj} + \exp(t\theta h_n)\, y_0.$$

Now define the following partial sums $Z_{nt} = \sum_{j=1}^{t} \varepsilon_{nj}$ with $Z_{n0} = 0$ and the random elements

$$X_{T_n}(r) = T_n^{-1/2}\sigma^{-1} Z_{[T_n r]} = T_n^{-1/2}\sigma^{-1} Z_{nj-1} \qquad (j-1)/T_n \le r < j/T_n$$

$$(j = 1, ..., T_n)$$

$$X_{T_n}(1) = T_n^{-1/2}\sigma^{-1} Z_{T_n}$$

then, (see Phillips (1987a)), $X_{T_n}(r) \to w(r)$, a Weiner process on C[0,1].

To prove Theorem 1,we shall derive the limiting distribution of the following quantities: a) y_{T_n}; b) $\sum_{t=1}^{T_n} y_{tn}$; c) $\sum_{t-1}^{T_n} y_{nt}^2$ and d) $\sum_{t=1}^{T_n} y_{nt-1}u_{nt}$.

A) y_{T_n}:

$$y_{T_n} = N^{1/2}T_n^{-1/2}h_n^{-1/2}\sum_{j=1}^{T_n} \exp((T_n - j)\theta h_n)u_{nj} + \exp(T_n\theta h_n)y_0$$

$$= N^{1/2}h_n^{-1/2}a(h_n)^{1/2}\sigma\sum_{j=1}^{T_n} \exp((1 - j/T_n)\theta N)\int_{(j-1)/T_n}^{j/T_n} dX_{T_n}(s) + \exp(\theta N)y_0$$

$$= N^{1/2}h_n^{-1/2}a(h_n)^{1/2}\sigma\sum_{j=1}^{T_n} \int_{(j-1)/T_n}^{j/T_n} \exp((1 - s)\theta N)dX_T(s) + \exp(\theta N)y_0$$

$$= N^{1/2}h_n^{-1/2}a(h_n)^{1/2}\sigma\int_0^1\exp((1 - s)\theta N)dX_{T_n}(s) + \exp(\theta N)y_0$$

$$\to N^{1/2}\sigma J_c(1) + \exp(\theta N)y_0, \qquad \text{as } n \to \infty.$$

B) $\sum_{t=1}^{T_n} y_{tn}$:

$$N^{-1/2}T_n^{-1}\Sigma y_{tn} = h_n^{-1/2}T_n^{-3/2}\sum_{t=1}^{T_n} \sum_{j=1}^{t} \exp((t - j)\theta h_n)u_{nj}$$

$$+ h_n^{1/2}T_n^{-3/2}\sum_{t=1}^{T_n} \exp(t\theta h_n)y_0$$

$$= \sigma a(h_n)^{1/2}h_n^{-1/2}T_n^{-1}\sum_{t=1}^{T_n} \sum_{j=1}^{t} \exp((t - j)\theta h_n)\int_{(j-1)/T_n}^{j/T_n} dX_{T_n}(s)$$

$$+ h_n^{-1/2} T_n^{-3/2} \sum_{t=1}^{T_n} \exp(t\theta h_n) y_0$$

$$= \sigma a(h_n)^{1/2} h_n^{-1/2} \sum_{t=1}^{T_n} \sum_{j=1}^{t} \int_{(t-1)/T_n}^{t/T_n} dr \int_{(j-1)/T_n}^{j/T_n} \exp((t-j)\theta h_n) dX_{T_n}(s)$$

$$+ h_n^{-1/2} T_n^{-3/2} \sum_{t=1}^{T_n} \exp(t\theta h_n) y_0$$

$$= \sigma a(h_n)^{1/2} h_n^{-1/2} \sum_{t=1}^{T_n} \sum_{j=1}^{t} \int_{(t-1)/T}^{t/T_n} \int_{(j-1)/T_n}^{j/T_n} \exp((r-s)\theta N) dX_{T_n}(s)\, dr$$

$$+ h_n^{-1/2} T_n^{-3/2} \sum_{t=1}^{T_n} \exp(t\theta h_n) y_0$$

$$= \sigma a(h_n)^{1/2} h_n^{-1/2} \int_0^1 \int_0^r \exp((r-s)\theta N) dX_{T_n}(s)\, dr$$

$$+ h_n^{-1/2} T_n^{-3/2} \sum_{t=1}^{T_n} \exp(t\theta h_n) y_0$$

$$= \sigma a(h_n)^{1/2} h_n^{-1/2} \int_0^1 \{X_{T_n}(r) + \theta N \int_0^r \exp((r-s)\theta N) dX_{T_n}(s)\} dr$$

$$+ h_n^{-1/2} T_n^{-3/2} \sum_{t=1}^{T_n} \exp(t\theta h_n) y_0$$

and $\quad N^{-1/2} T_n^{-1} \sum y_{nt} \to \sigma \int_0^1 J_c(r) dr + y_0 N^{-3/2}(\exp(\theta N) - 1)/\theta \qquad$ as $n \to \infty$.

C) $\sum_{t=1}^{T_n} y_{nt}^2$:

$$N^{-1} T_n^{-1} \sum_{t=1}^{T_n} y_{nt}^2 = N^{-1} T_n^{-1} \sum_{t=1}^{T_n} [\sum_{j=1}^{t} \exp((t-j)\theta h_n) u_{nj} + \exp(t\theta h_n) y_0]^2$$

$$= N^{-1}T_n^{-1}\sum_{t=1}^{T_n} a(h_n)(\sum_{j=1}^{t} \exp((t-j)\theta h_n)\varepsilon_{nj})^2$$

$$+ 2 N^{-1}T_n^{-1}y_0 a(h_n)^{1/2}\sum_{t=1}^{T_n} \exp(t\theta h_n)\sum_{j=1}^{t} \exp((t-j)\theta h_n)\varepsilon_{nj}$$

$$+ N^{-1}T_n^{-1}\sum_{t=1}^{T_n} \exp(2t\theta h_n)y_0^2$$

$$= \sigma^2 h_n^{-1}a(h_n)T_n^{-1}\sum_{t=1}^{T_n} (\sum_{j=1}^{t} \exp((t-j)\theta h_n)\int_{(j-1)/T_n}^{j/T_n} dX_{T_n}(s))^2$$

$$+ 2\sigma h_n^{-1}a(h)^{1/2}T_n^{-3/2}y_0\sum_{t=1}^{T_n} \exp(t\theta h_n)\sum_{j=1}^{t} \exp((t-j)\theta h_n)\int_{(j-1)/T_n}^{j/T_n} dX_{T_n}(s)$$

$$+ N^{-1}T_n^{-1}\sum_{t=1}^{T_n} \exp(2t\theta h_n)y_0^2$$

$$= \sigma^2 h_n^{-1}a(h_n)\sum_{t=1}^{T_n} \int_{(t-1)/T_n}^{t/T_n} dr(\sum_{j=1}^{t} \int_{(j-1)/T_n}^{j/T_n} \exp((r-s)\theta N)dX_{T_n}(s))^2$$

$$+ 2\sigma h_n^{-1}a(h_n)^{1/2}N^{-1/2}y_0\sum_{t=1}^{T_n} \sum_{j=1}^{t} \int_{(t-1)/T_n}^{t/T_n} \exp(r\theta N)dr\int_{(j-1)/T_n}^{j/T_n}$$
$$\exp((r-s)\theta N)dX_{T_n}(s)$$

$$+ N^{-1}T_n^{-1}\sum_{t=1}^{T_n} \exp(2t\theta h_n)y_0^2$$

$$= \sigma^2 h_n^{-1}a(h_n)\int_0^1 dr(\int_0^r \exp((r-s)\theta N)dX_{T_n}(s))^2$$

$$+ 2\sigma h_n^{-1/2}a(h_n)^{1/2}N^{-1/2}y_0\int_0^1 \exp(r\theta N)dX_{T_n}(s)$$

$$+ N^{-1}T_n^{-1}\sum_{t=1}^{T_n} \exp(2t\theta h_n)y_0^2$$

$$= \sigma^2 h_n^{-1} a(h_n)\int_0^1 (X_{T_n}(r) + \theta N\int_0^r \exp((r-s)\theta N)X_{T_n}(s)ds)^2 dr$$

$$+ 2\sigma h_n^{-1/2} a(h_n)^{1/2} N_n^{-1/2} y_0 \int_0^1 \exp(r\theta N)(X_{T_n}(r) + \theta N\int_0^r \exp((r-s)\theta N)X_{T_n}(s)ds)dr$$

$$+ N^{-1}T_n^{-1}\sum_{t=1}^{T_n} \exp(2t\theta h_n)y_0^2$$

$$\to \sigma^2\int_0^1 J_c(r)^2 dr + 2\sigma y_0 N^{-1/2}\int_0^1 \exp(rc)J_c(r)dr + y_0^2 N^{-2}(\exp(2\theta N) - 1)/2\theta$$

as $n \to \infty$

D) $\sum_{t=1}^{T_n} y_{nt-1}u_{nt}$:

Note that by squaring (A.1) we obtain

$$y_{nt}^2 = \exp(2\theta h_n)y_{nt-1}^2 + 2\exp(\theta h_n)y_{nt-1}u_{nt} + u_{nt}^2$$

$$y_{nt}^2 - y_{nt-1}^2 = (\exp(2\theta h_n) - 1)y_{nt-1}^2 + 2\exp(\theta h_n)y_{nt-1}u_{nt} + u_{nt}^2;$$

summing both sides over $t=1, \ldots, T_n$, we get

$$y_{T_n}^2 - y_0^2 = (\exp(2\theta h_n) - 1)\Sigma y_{nt-1}^2 + 2\exp(\theta h_n)\Sigma y_{nt-1}u_{nt} + \Sigma u_{nt}^2$$

and

$$N^{-1}\Sigma y_{nt-1}u_{nt} = (1/2)\exp(-\theta h_n)[N^{-1}y_{T_n}^2 - N^{-1}y_0^2$$

$$- T_n(\exp(2\theta h_n) - 1)N^{-1}T_n^{-1}\Sigma y_{nt-1}^2 - N^{-1}\Sigma u_{nt}^2].$$

Now, $N^{-1}\Sigma u_{nt}^2 = N^{-1}a(h)\Sigma\varepsilon_{nt}^2 = h_n^{-1}a(h)T_n^{-1}\Sigma\varepsilon_{nt}^2 \to \sigma^2$, and

$$N^{-1}\Sigma y_{t-1}u_{nt} \to (1/2)[(\sigma J_c(1) + N^{-1/2}\exp(\theta N)y_0)^2 - y_0^2/N - 2\theta N[\psi] - \sigma^2]$$

where $\psi = \sigma^2\int_0^1 J_c(r)^2 dr + 2\sigma y_0 N^{-1/2}\int_0^1 \exp(rc)J_c(r)dr + y_0^2 N^{-2}(\exp(2\theta N) - 1)/2\theta$

Using the fact that:

$$T_n(\hat{\alpha}_n - \alpha_n) = N^{-1}\Sigma_1^T y_{nt-1}u_{nt}\{N^{-1}T_n^{-1}\Sigma_1^T y_{nt-1}^2\}^{-1}$$

$$T_n(\hat{\alpha}_n - \alpha_n) \to (1/2)[(\sigma J_c(1) + N^{-1/2}\exp(\theta N)y_0)^2 - y_0^2/N - 2\theta N[\psi] - \sigma^2]/\psi$$

$$= [\int_0^1 J_c(r)dw_r + (y_0/\sigma N^{1/2})(\exp(c)J_c(1) - 2c\int_0^1 \exp(cr)J_c(r)dr]/$$

$$[\int_0^1 J_c(r)^2 dr + (2y_0/\sigma N^{1/2})\int_0^1 \exp(cr)J_c(r)dr$$

$$+ (y_0{}^2/N\sigma^2)(\exp(2c) - 1)/2c].$$

What remains to be shown is that in the numerator $\exp(c)J_c(1) - 2c\int_0^1 \exp(cr)J_c(r)dr = \int_0^1 \exp(cr)dw_r$. Now, from the proof of Lemma 1 below: $\int_0^1 \exp(cr)dw_r \sim N(0,(\exp(2c) - 1)/2c)$. Hence, we must show that $\exp(c)J_c(1) - 2c\int_0^1 \exp(cr)J_c(r)dr \sim N(0,(\exp(2c) - 1)/2c)$.

Note that $J_c(1) \sim N(0,(\exp(2c) - 1)/2c)$ and $\int_0^1 \exp(rc)J_c(r)dr \sim N(0,(\exp(4c) - 1 - 4c\exp(2c))/8c^3)$. Therefore $[\exp(c)J_c(1) - 2c\int_0^1 \exp(rc)J_c(r)dr]$ is Gaussian with mean 0 and variance given by $\exp(2c)\mathrm{Var}[J_c(1)] + 4c^2\mathrm{Var}[\int_0^1 \exp(rc)J_c(r)dr] - 2E[J_c(1)\int_0^1 \exp(rc)J_c(r)dr]$.

Now, $E[J_c(1)\int_0^1 \exp(rc)J_c(r)dr]$

$$= \int_0^1 \exp(rc)E[J_c(1)J_c(r)]dr$$

$$= \int_0^1 \exp(rc)[\exp((r+1)c) - \exp((1-r)c)]/2c \, dr$$

$$= \exp(c)[\exp(2c) - 1 - 2c]/(2c)^2.$$

Therefore $[\exp(c)J_c(1) - 2c\int_0^1 \exp(rc)J_c(r)dr]$ has variance given by $(\exp(2c) - 1)/2c$.

Proof of Lemma 1

To prove part (v), we first note that $w(s)$ is Gaussian with mean zero and variance s, then $J_c(r)$ and $\int_0^1 \exp(rc)J_c(r)dr$ are also Gaussian with mean zero for any c. Now $J_c(r) \sim N(0,(\exp(2rc) - 1)/2c)$ and using the fact that $E(w(s)w(t)) = \min(t,s)$, simple calculations yield

$$E[J_c(r)J_c(k)] = [\exp((r+k)c) - \exp((M-m)c)]/2c$$

where $M = \max(r,k)$ and $m = \min(r,k)$. This result can be used to show that

$$\int_0^1 \exp(rc)J_c(r)dr \sim N(0,v)$$

where $v = (\exp(4c) - 1 - 4c\exp(2c))/8c^3.$

Part (v) follows by noting that $(2c)^3\exp(-4c)v \to 1$ as $c \to +\infty$.

To prove part (vi), note that $\int_0^1 \exp(cr)dw_r = \exp(c)\int_0^1 \exp(-c(1-r))dw_r - \exp(c)J_{-c}(1)$. Now, $J_{-c}(1) \sim N(0,(\exp(-2c)-1)/(-2c))$, hence $\int_0^1 \exp(cr)dw_r \sim N(0,(\exp(2c)-1)/2c)$. The result follows by noting that $(2c)^{1/2}\exp(-c)\int_0^1 \exp(cr)dw_r = (2c)^{1/2}\exp(-c) \cdot N(0,(\exp(2c)-1)/2c) = N(0,(1-\exp(-2c))) \to N(0,1)$ as $c \to +\infty$.

BIBLIOGRAPHY

Arnold, L. (1974), Stochastic Differrential Equations: Theory and Applications, John Wiley and Sons, New York.

Basawa, I.V. and B.P. Rao (1980), Statistical Inference for Stochastic Processes, Academic Press, New York.

Brown, B.M. and J.I. Hewitt (1975), "Asymptotic Likelihood Theory for Diffusion Processes", Journal of Applied Probability, vol. 12, pp. 228–238.

Cavanagh, C. (1986), "Roots Local to Unity", mimeo, Harvard University.

Chan, N.H. and C.Z. Wei (1987), "Asymptotic Inference for Nearly Nonstationary AR(1) Processes", forthcoming Annals of Statistics.

Dickey, D.A. (1976), "Estimation and Hypothesis Testing for Nontstationary Time Series", Ph.D. Thesis, Iowa State University, Ames.

Dickey, D.A. and W.A. Fuller (1979), "Distribution of the Estimators for Autoregressive Time Series with a Unit Root", Journal of the American Statistical Association, vol. 74, pp. 427–431.

Dickey, D.A. and W.A. Fuller (1981), "Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root", Econometrica, vol. 49, pp. 1057–1072.

Evans, G.B.A. and N.E. Savin (1981a), "The Calculation of the Limiting Distribution of the Least Squares Estimator of the Parameter in a Random Walk Model", Annals of Mathematical Statistics, vol. 5, pp. 1114–1118.

Evans, G.B.A. and N.E. Savin (1981b), "Testing for Unit Roots: 1", Econometrica, vol. 49, pp. 753–779.

Evans, G.B.A. and N.E. Savin (1984), "Testing for Unit Roots: 2", Econometrica, vol. 52, pp. 1241–1269.

Fuller, Wayne A. (1976), Introduction to Statistical Time Series, John Wiley & Sons, New York.

Gurland, J. (1948), "Inversion Formulae for the Distribution of Ratios", Annals of Mathematical Statistics, vol. 19, pp. 228–237.

Jeganathan, P. (1987), "On the Asymptotic Behavior of Least Squares Estimators in AR Times Series with Roots Near the Unit Circle", Technical Report #150, Department of Statistics, The University of Michigan.

Liptser, R.S. and A.N. Shiryayev (1978), Statistics of Random Processes II: Applications, Springer–Verlag, New York.

Mann, H.B. and A. Wald (1943), "On the Statistical Treatment of Linear Stochastic Difference Equations", Econometrica, vol. 11, pp. 173–220.

Mehta, J.S. and P.A.V.B. Swamy (1978), "The Existence of Moments of Some Simple Bayes Estimators of Coefficients in a Simultaneous Equation Model", Journal of Econometrics, vol. 7, pp. 1–14.

Nabeya, S. and K. Tanaka (1987), "Asymptotic Distribution of the Estimator for the Nonstable First–order Autoregressive Model", mimeo, Hitotsubashi University.

Perron, P. (1986), "Tests of Joint Hypotheses for Time Series Regression with a Unit Root", Cahier de recherche no. 2086, C.R.D.E., Université de Montréal.

Perron, P. (1987), "Test Consistency with Varying Sampling Interval", Cahier de recherche no. 4187, C.R.D.E., Université de Montréal.

Perron, P. (1988a), "The Calculation of the Limiting Distribution of the Least Squares Estimator in a Near–Integrated Model", mimeo, Université de Montréal.

Perron, P. (1988b), "A Continuous–Time Approximation to the Stationary First–order Autoregressive Model", in preparation, Université de Montréal.

Perron, P. (1988c), "A Continuous–Time Approximation to the First–order Autoregressive Model: The Case with an Unknown Mean", in preparation, Université de Montréal.

Phillips, P.C.B. (1977), "Approximations to Some Finite Sample Distributions Associated with a First–order Stochastic Difference Equation", Econometrica, vol. 45, pp. 463–485.

Phillips, P.C.B. (1987a), "Time Series Regression with Unit Roots", Econometrica, vol. 55, pp. 277–302.

Phillips, P.C.B. (1987b), "Towards a Unified Asymptotic Theory for Autoregression", Biometrika, vol. 74, pp. 535–47.

Phillips, P.C.B. (1988), "Regression Theory for Near–Integrated Time Series", forthcoming in Econometrica.

Phillips, P.C.B. and P. Perron (1988), "Testing for a Unit Root in Time Series Regression", forthcoming in Biometrika.

Rao, M.M. (1978), "Asymptotic Distribution of an Estimator of the Boundary Parameter of an Unstable Process", Annals of Statistics, vol. 6, pp. 185–190.

Rubin, H. (1950), "Connsistency of Maximum Likelihood Estimates in the Explosive Case", in T.C. Koopmans (ed.) Statistical Inference in Dynamic Economic Models, Wiley, New York.

Said, S.E. and D.A. Dickey (1984), "Testing for Unit Roots in Autoregressive–Moving Average Models of Unknown Order", Biometrika, vol. 17, pp. 599–607.

White, J.S. (1958), "The Limiting Distribution of the Serial Correlation Coefficient in the Explosive Case", Annals of Mathematical Statistics, vol. 29, pp. 1188–1197.

Table 1

Selected Percentage Points of the Distribution of $N(\hat{\theta} - \theta)$

	1 %	2.5 %	5 %	10 %	90 %	95 %	97.5 %	99 %
c = –5.0								
γ = 0.0	–16.7378	–13.1743	–10.4633	–7.7242	2.4561	3.0637	3.5239	4.0342
= 0.5	–13.8955	–10.9832	–8.7550	–6.4899	2.2695	2.8257	3.2413	3.6788
= 1.0	–9.5405	–7.6095	–6.1140	–4.5703	1.9279	2.4017	2.7578	3.1215
= 2.0	–4.8739	–3.9531	–3.2206	–2.4413	1.3836	1.7357	2.0123	2.3035
c = –2.0								
γ = 0.0	–15.0342	–11.6602	–9.1377	–6.6406	1.4622	1.926*	2.3095	2.7849
= 0.5	–12.2749	–9.5498	–7.5038	–5.4726	1.3047	1.6839	2.005*	2.3921
= 1.0	–8.1023	–6.3512	–5.0258	–3.6952	1.0808	1.3688	1.5996	1.8594
= 2.0	–3.8162	–3.0438	–2.4456	–1.8281	0.7943	0.9940	1.1599	1.3310
c = –1.0								
γ = 0.0	–14.3901	–11.0813	–8.6182	–6.2072	1.1752	1.5887	1.9502	2.4005
= 0.5	–11.6484	–8.9854	–7.0027	–5.0566	1.007*	1.3498	1.6426	2.0031
= 1.0	–7.5197	–5.8323	–4.5689	–3.3203	0.8076	1.056*	1.2535	1.4819
= 2.0	–3.3477	–2.6348	–2.0925	–1.5445	0.5908	0.7512	0.8784	1.012*
c = 0.0								
γ = 0.0	–13.6919	–10.4399	–8.0383	–5.7133	0.9280	1.2854	1.6122	2.0325
= 0.5	–10.9570	–8.3535	–6.4315	–4.5709	0.7643	1.0481	1.3042	1.6338
= 1.0	–6.8480	–5.2202	–4.0196	–2.8568	0.5876	0.7780	0.9384	1.1302
= 2.0	–2.7392	–2.0886	–1.6093	–1.1454	0.4116	0.5271	0.6211	0.7260
c = 1.0								
γ = 0.0	–12.9229	–9.7183	–7.3710	–5.1243	0.7101	1.0079	1.2966	1.6832
= 0.5	–10.1753	–7.6177	–5.7469	–3.9627	0.5544	0.7771	0.9932	1.2871
= 1.0	–6.0273	–4.4386	–3.2793	–2.1696	0.4017	0.5366	0.6571	0.8093
= 2.0	–1.7104	–1.1842	–0.8640	–0.5948	0.2610	0.3360	0.3994	0.4725
c = 2.0								
γ = 0.0	–12.0557	–8.8762	–6.5587	–4.3431	0.5191	0.7585	1.0067	1.3550
= 0.5	–9.2554	–6.7080	–4.8444	–3.0206	0.3756	0.5401	0.7143	0.9680
= 1.0	–4.8763	–3.1253	–1.7404	–0.8782	0.2493	0.3357	0.4182	0.5294
= 2.0	–0.5897	–0.4307	–0.3294	–0.2374	0.1466	0.1896	0.2272	0.2712
c = 5.0								
γ = 0.0	–7.8066	–3.0640	–1.0714	–0.3734	0.1316	0.2212	0.3410	0.5473
= 0.5	–2.2400	–0.5435	–0.2103	–0.0969	0.0611	0.0974	0.1505	0.2571
= 1.0	–0.0826	–0.0579	–0.0438	–0.0315	0.0283	0.0385	0.0491	0.0650
= 2.0	–0.0275	–0.0225	–0.0186	–0.0142	0.0137	0.0177	0.0214	0.0259

Note: The integral (4.2) which evaluates $P[N(\hat{\theta} - \theta) \leq z]$ is undefined for $z = -c$ and the numerical integration is accordingly ill–behaved for values of z around –c. The * entry indicates where such a situation occurs. While all digits reported are accurate for the other entries, in these cases the accuracy is not as precise.

Table 2
Selected Percentage Points of the Distribution of $T(\hat{\alpha} - \alpha)$
$y_t = \alpha y_{t-1} + e_t$, $y_o/\sigma = T^{1/2}\gamma$, $\alpha = \exp(c/T)$

	1 %	2.5 %	5 %	10 %	90 %	95 %	97.5 %	99 %
c=–5.0								
γ=0.5								
T=10	–7.750	–6.555	–5.442	–4.225	2.065	2.556	2.962	3.436
T=25	–10.539	–8.711	–7.123	–5.421	2.165	2.722	3.143	3.578
T=50	–12.061	–9.530	–7.840	–5.935	2.213	2.772	3.206	3.646
T=100	–13.305	–10.487	–8.314	–6.266	2.212	2.777	3.193	3.614
γ=1.0								
T=10	–5.642	–4.756	–3.941	–3.050	1.693	2.079	2.395	2.713
T=25	–7.407	–6.053	–5.002	–3.874	1.815	2.260	2.610	2.930
T=50	–8.488	–6.930	–5.616	–4.228	1.862	2.335	2.677	3.038
T=100	–8.969	–7.193	–5.956	–4.464	1.869	2.341	2.713	3.056
γ = 2.0								
T=10	–3.282	–2.708	–2.245	–1.727	1.167	1.464	1.694	1.939
T=25	–4.048	–3.308	–2.708	–2.099	1.278	1.594	1.869	2.152
T=50	–4.462	–3.683	–3.031	–2.317	1.323	1.656	1.943	2.232
T=100	–4.696	–3.831	–3.117	–2.385	1.333	1.675	1.956	2.251
c=–2.0								
γ=0.5								
T=10	–7.854	–6.626	–5.414	–4.196	1.336	1.748	2.123	2.564
T=25	–10.132	–8.183	–6.585	–4.939	1.315	1.700	2.047	2.474
T=50	–10.951	–8.626	–6.962	–5.217	1.321	1.708	2.053	2.472
T=100	–11.691	–9.215	–7.319	–5.398	1.291	1.671	1.996	2.375
γ=1.0								
T=10	–5.583	–4.572	–3.712	–2.882	1.049	1.337	1.568	1.847
T=25	–6.757	–5.350	–4.362	–3.299	1.064	1.349	1.587	1.850
T=50	–7.518	–6.045	–4.751	–3.580	1.075	1.363	1.592	1.872
T=100	–7.732	–6.225	–4.935	–3.678	1.068	1.349	1.573	1.836
γ=2.0								
T=10	–3.015	–2.449	–1.967	–1.507	0.756	0.939	1.097	1.262
T=25	–3.402	–2.698	–2.211	–1.672	0.773	0.978	1.138	1.303
T=50	–3.620	–2.913	–2.382	–1.789	0.785	0.983	1.152	1.326
T=100	–3.779	–2.991	–2.437	–1.834	0.785	0.986	1.138	1.316
c=0.0								
γ=0.5								
T=10	–7.892	–6.392	–5.230	–3.897	0.864	1.188	1.508	1.915
T=25	–9.589	–7.578	–5.897	–4.347	0.791	1.098	1.372	1.765
T=50	–9.954	–7.873	–6.128	–4.454	0.797	1.092	1.348	1.688
T=100	–10.457	–8.159	–6.417	–4.587	0.772	1.052	1.313	1.608

Table 2 (continued)

	1 %	2.5 %	5 %	10 %	90 %	95 %	97.5 %	99 %
γ=1.0								
T=10	–5.253	–4.209	–3.364	–2.469	0.634	0.843	1.001	1.221
T=25	–5.892	–4.638	–3.677	–2.653	0.597	0.793	0.961	1.167
T=50	–6.557	–5.128	–3.995	–2.788	0.598	0.798	0.955	1.156
T=100	–6.629	–5.178	–4.063	–2.891	0.588	0.777	0.935	1.125
γ=2.0								
T=10	–2.433	–1.873	–1.462	–1.057	0.430	0.549	0.648	0.758
T=25	–2.589	–1.996	–1.569	–1.102	0.417	0.531	0.630	0.737
T=50	–2.715	–2.055	–1.582	–1.131	0.416	0.534	0.629	0.738
T=100	–2.721	–2.107	–1.613	–1.151	0.411	0.527	0.617	0.718
c=2.0								
γ=0.5								
T=10	–7.598	–5.796	–4.461	–2.832	0.459	0.655	0.867	1.163
T=25	–8.423	–6.302	–4.686	–2.962	0.406	0.583	0.770	1.054
T=50	–8.953	–6.493	–4.751	–3.016	0.401	0.565	0.745	1.008
T=100	–9.207	–6.595	–4.801	–3.000	0.377	0.550	0.718	0.957
γ=1.0								
T=10	–4.359	–2.564	–1.421	–0.788	0.290	0.386	0.481	0.611
T=25	–4.306	–2.870	–1.617	–0.833	0.263	0.356	0.445	0.558
T=50	–4.703	–3.036	–1.676	–0.860	0.262	0.348	0.432	0.543
T=100	–5.037	–3.163	–1.735	–0.887	0.252	0.340	0.418	0.526
γ=2.0								
T=10	–0.562	–0.437	–0.338	–0.247	0.165	0.214	0.257	0.311
T=25	–0.585	–0.432	–0.334	–0.241	0.154	0.200	0.237	0.288
T=50	–0.596	–0.445	–0.335	–0.242	0.151	0.196	0.235	0.279
T=100	–0.597	–0.424	–0.325	–0.238	0.149	0.192	0.226	0.272

Table 3

Expectations of $g(T)(\hat{\alpha} - \alpha)$

Based on the asymptotic distribution (exact values in parentheses)

$y(0)/\sigma = 0.0$

T	$\alpha = 0.90$		$\alpha = 0.99$		$\alpha = 1.00$		$\alpha = 1.01$		$\alpha = 1.05$	
25	−0.8933	(−0.7339)	−2.5718	(−2.2980)	−1.2597	(−1.1347)	−4.4482	(−4.0382)	−2.0080	(−1.8770)
50	−0.6416	(−0.5528)	−1.8457	(−1.7374)	−1.2597	(−1.1955)	−2.7778	(−2.6566)	−2.4096	(−2.3709)
100	−0.4570	(−0.4027)	−1.3342	(−1.2890)	−1.2597	(−1.2271)	−2.1295	(−2.0902)	−3.6704	(−3.7018)
200	−0.3240	(−0.2886)	−0.9675	(−0.9467)	−1.2597	(−1.2433)	−2.2841	(−2.2710)	−5.2877	(−5.3839)
400	−0.2293	(−0.2054)	−0.6975	(−0.6867)	−1.2597	(−1.2514)	−3.3706	(−3.3726)	–	–

$y(0)/\sigma = 4.0$

T	$\alpha = 0.90$		$\alpha = 0.99$		$\alpha = 1.00$		$\alpha = 1.01$		$\alpha = 1.05$	
25	−0.5509	(−0.4622)	−1.5492	(−1.4152)	−0.7480	(−0.6882)	−2.5895	(−2.3978)	−0.9968	(−0.9368)
50	−0.4878	(−0.4222)	−1.4021	(−1.3279)	−0.9467	(−0.9038)	−2.0429	(−1.9642)	−1.3231	(−1.2855)
100	−0.3943	(−0.3477)	−1.1555	(−1.1179)	−1.0835	(−1.0571)	−1.7800	(−1.7490)	−1.8444	(−1.8091)
200	−0.3001	(−0.2673)	−0.8981	(−0.8791)	−1.1657	(−1.1509)	−2.0108	(−1.9989)	−2.3509	(−0.0688)
400	−0.2205	(−0.1975)	−0.6712	(−0.6608)	−1.2110	(−1.2032)	−2.9477	(−2.9467)	–	–

$y(0)/\sigma = 16.0$

T	$\alpha = 0.90$		$\alpha = 0.99$		$\alpha = 1.00$		$\alpha = 1.01$		$\alpha = 1.05$	
25	−0.0805	(−0.0713)	−0.1677	(−0.1592)	−0.0732	(−0.0697)	−0.2244	(−0.2145)	−0.0455	(−0.0439)
50	−0.1060	(−0.0942)	−0.2605	(−0.2523)	−0.1509	(−0.1468)	−0.2556	(−0.2493)	−0.0308	(−0.0302)
100	−0.1288	(−0.1148)	−0.3678	(−0.3598)	−0.2968	(−0.2920)	−0.3150	(−0.3102)	−0.0060	(−0.0059)
200	−0.1423	(−0.1272)	−0.4311	(−0.4236)	−0.5120	(−0.5072)	−0.3955	(−0.3910)	−0.0001	(−0.0001)
400	−0.1399	(−0.1253)	−0.4287	(−0.4225)	−0.7480	(−0.7440)	−0.4107	(−0.4045)	–	–

Table 4

Standard Deviations of $g(T)(\hat{\alpha} - \alpha)$

Based on the asymptotic distribution (exact values in parentheses)

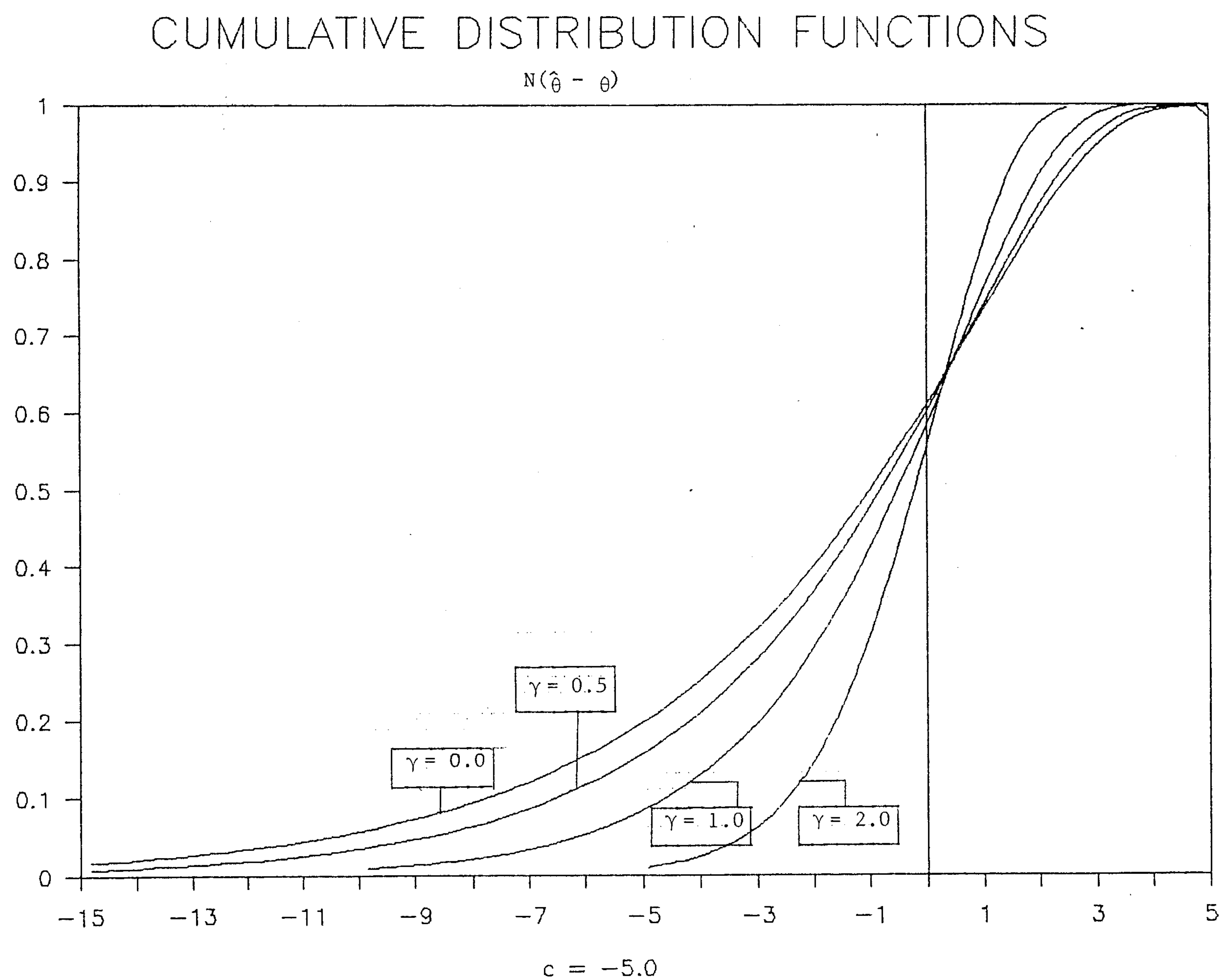
T	$\alpha = 0.90$		$\alpha = 0.99$		$\alpha = 1.00$		$\alpha = 1.01$		$\alpha = 1.05$	
$y(0)/\sigma = 0.0$										
25	1.7438	(1.5135)	4.5919	(4.1480)	2.2486	(2.0401)	7.9463	(7.2601)	3.7365	(3.5359)
50	1.4274	(1.3004)	3.3049	(3.1293)	2.2486	(2.1356)	4.9795	(4.7771)	5.4288	(5.4447)
100	1.2482	(1.1657)	2.4199	(2.3552)	2.2486	(2.1898)	3.8966	(3.8620)	18.4035	(18.8851)
200	1.1523	(1.0876)	1.8307	(1.8113)	2.2486	(2.2186)	4.6999	(4.7802)	228.4101	(234.2883)
400	1.1030	(1.0452)	1.4611	(1.4594)	2.2486	(2.2334)	12.0295	(12.3768)	–	–
$y(0)/\sigma = 4.0$										
25	1.2425	(1.0374)	2.8652	(2.6100)	1.3852	(1.2683)	4.8429	(4.4536)	2.1914	(2.0004)
50	1.2305	(1.0706)	2.5471	(2.4084)	1.7088	(1.6240)	3.7469	(3.5833)	3.8048	(3.5898)
100	1.1813	(1.0611)	2.1203	(2.0554)	1.9401	(1.8888)	3.3340	(3.2820)	13.2287	(12.7345)
200	1.1321	(1.0402)	1.7228	(1.6956)	2.0825	(2.0543)	4.3122	(4.3375)	161.2751	(5.6014)
400	1.0974	(1.0231)	1.4256	(1.4158)	2.1622	(2.1475)	11.2652	(11.4476)	–	–
$y(0)/\sigma = 16$										
25	0.4425	(0.3173)	0.6374	(0.5686)	0.2585	(0.2552)	0.8714	(0.8164)	0.2396	(0.2274)
50	0.5862	(0.4264)	0.7560	(0.6835)	0.4155	(0.4098)	0.8121	(0.7567)	0.2311	(0.2179)
100	0.7441	(0.5554)	0.8961	(0.8166)	0.6714	(0.6631)	0.9412	(0.8692)	0.2299	(0.2119)
200	0.8861	(0.6876)	1.0199	(0.9346)	1.0141	(1.0046)	1.5026	(1.3650)	0.6096	(0.2090)
400	0.9844	(0.8021)	1.0937	(1.0122)	1.3852	(1.3769)	4.0157	(3.6065)	–	–

Table 5
Power function for testing the hypothesis $\alpha = 1.00$ against alternatives near 1.00

Based on the asymptotic distribution (exact values in parentheses)
Significance level 0.05

T	$\alpha = 0.90$		$\alpha = 0.95$		$\alpha = 0.99$		$\alpha = 1.00$	$\alpha = 1.01$		$\alpha = 1.025$		$\alpha = 1.05$	
$y(0)/\sigma = 0.0$													
25	0.0789	(0.0779)	0.0494	(0.0506)	0.0456	(0.0472)	0.0500	0.0584	(0.0570)	0.0832	(0.0759)	0.1946	(0.1645)
50	0.1913	(0.1848)	0.0769	(0.0770)	0.0440	(0.0461)	0.0500	0.0730	(0.0706)	0.1992	(0.1843)	0.7073	(0.6904)
100	0.5893	(0.5615)	0.1835	(0.1818)	0.0461	(0.0462)	0.0500	0.1379	(0.1322)	0.7164	(0.7046)	0.9755	(1.0000)
200	0.9925	(0.9870)	0.5665	(0.5487)	0.0627	(0.0622)	0.0500	0.5330	(0.5179)	1.0000	(1.0000)	1.0000	(1.0000)
400	1.0000	(1.0000)	0.9897	(0.9862)	0.1290	(0.1284)	0.0500	0.9402	–	–	–	–	–
$y(0)/\sigma = 2.0$													
25	0.0920	(0.0918)	0.0524	(0.0536)	0.0451	(0.0467)	0.0500	0.0604	(0.0614)	0.0949	(0.0887)	0.2646	(0.2406)
50	0.2160	(0.2119)	0.0829	(0.0825)	0.0439	(0.0463)	0.0500	0.0763	(0.0745)	0.2331	(0.2198)	0.7612	(0.7513)
100	0.6232	(0.5960)	0.1951	(0.1915)	0.0465	(0.0469)	0.0500	0.1465	(0.1412)	0.7444	(0.7341)	0.9798	(1.0000)
200	0.9942	(0.9900)	0.5834	(0.5676)	0.0637	(0.0633)	0.0500	0.5510	(0.5369)	0.9790	(1.0000)	1.0000	(1.0000)
400	1.0000	(1.0000)	0.9908	(0.9877)	0.1307	(0.1304)	0.0500	0.9425	–	–	–	–	–
$y(0)/\sigma = 8.0$													
25	0.4933	(0.4853)	0.1541	(0.1565)	0.0447	(0.0448)	0.0500	0.0965	(0.0918)	0.3184	(0.3083)	0.8395	(0.8421)
50	0.7035	(0.6816)	0.2257	(0.2242)	0.0465	(0.0462)	0.0500	0.1532	(0.1475)	0.6648	(0.6616)	0.9835	(0.9852)
100	0.9494	(0.9350)	0.4135	(0.4069)	0.0548	(0.0554)	0.0500	0.3271	(0.3207)	0.9375	(0.9360)	0.9989	(1.0000)
200	1.0000	(0.9998)	0.8037	(0.7925)	0.0794	(0.0803)	0.0500	0.7457	(0.7430)	0.9951	(1.0000)	1.0000	(1.0000)
400	1.0000	(1.0000)	0.9986	(0.9980)	0.1590	(0.1610)	0.0500	0.9681	–	–	–	–	–

Figure 1



CUMULATIVE DISTRIBUTION FUNCTIONS
N($\hat{\theta}$ - θ)
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
−15
−13
−11
−9
−7
−5
−3
−1
1
3
5
γ = 0.0
γ = 0.5
γ = 1.0
γ = 2.0
c = −5.0

Figure 2

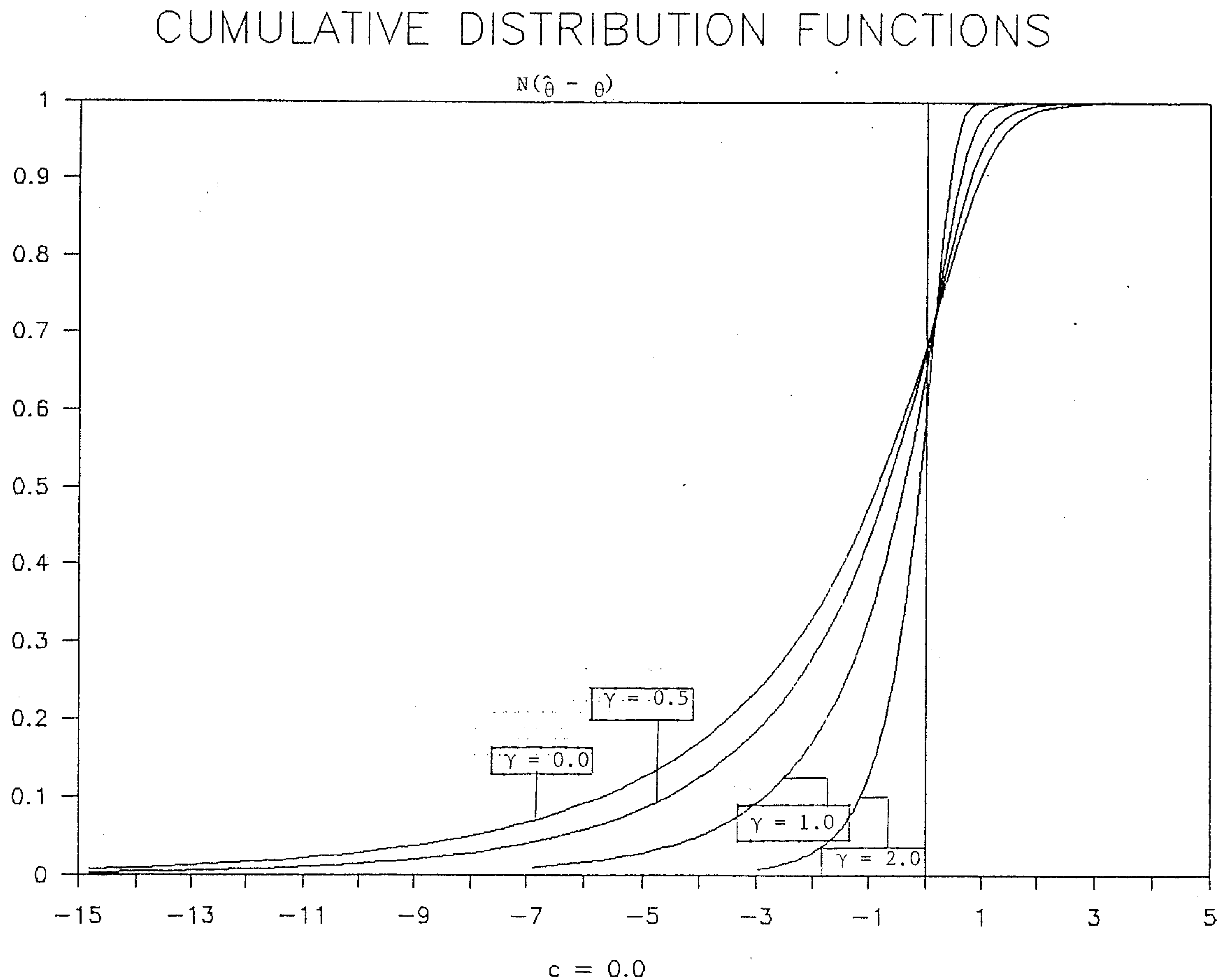

CUMULATIVE DISTRIBUTION FUNCTIONS
N(θ̂ - θ)
γ = 0.0
γ = 0.5
γ = 1.0
γ = 2.0
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
−15
−13
−11
−9
−7
−5
−3
−1
1
3
5
c = 0.0

Figure 3

CUMULATIVE DISTRIBUTION FUNCTIONS

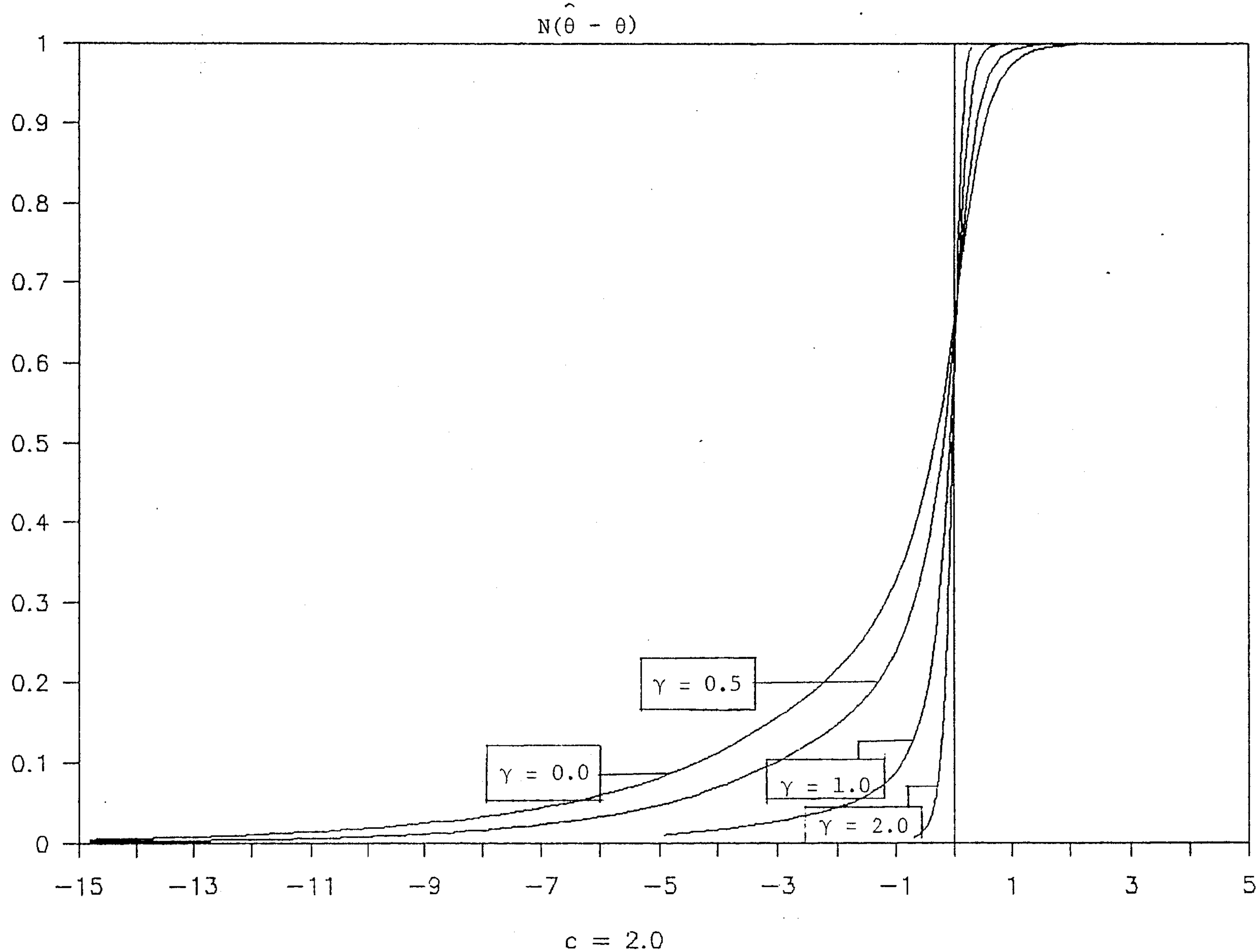


Figure 4

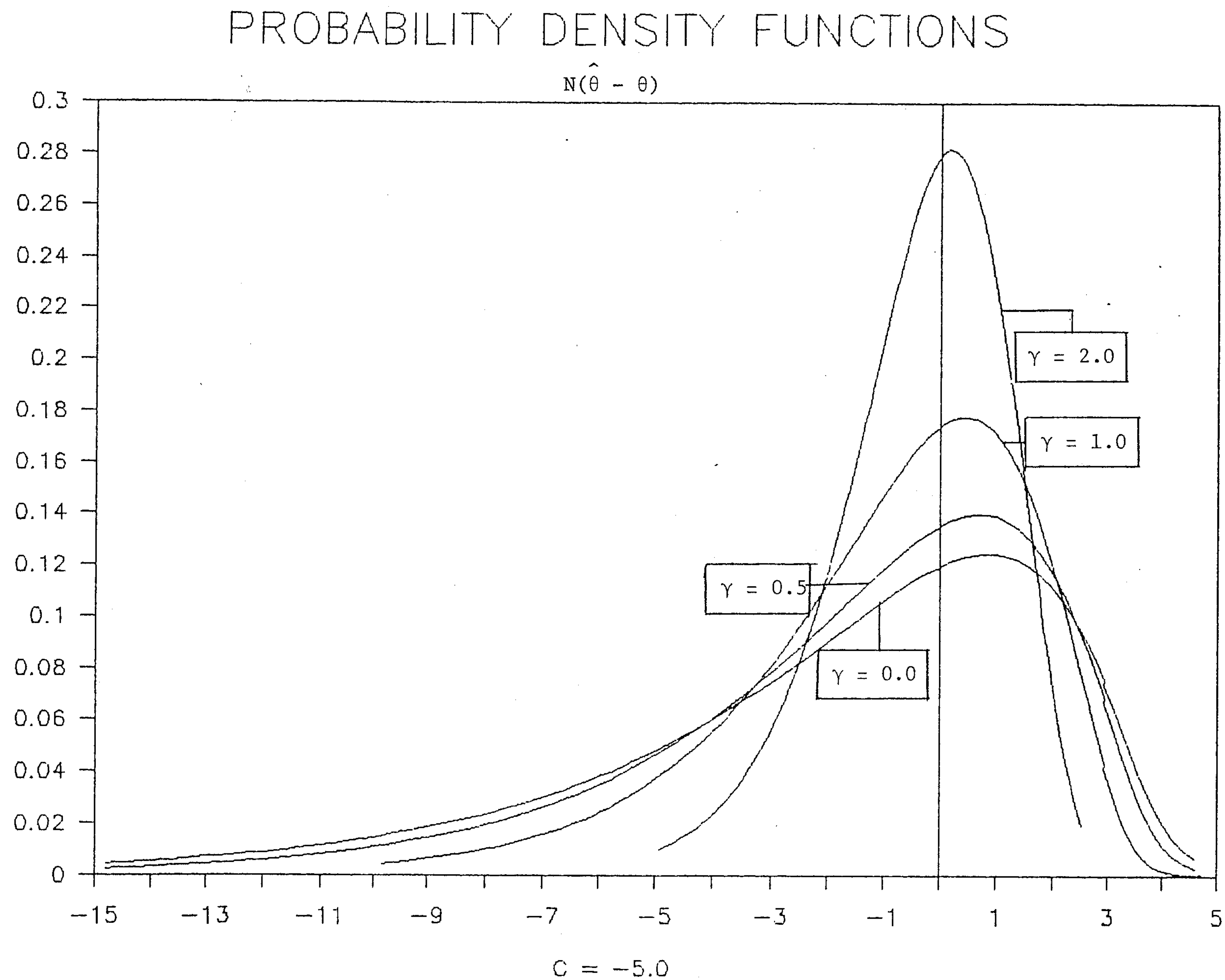

PROBABILITY DENSITY FUNCTIONS
N(θ̂ - θ)
0.3
0.28
0.26
0.24
0.22
0.2
0.18
0.16
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
γ = 2.0
γ = 1.0
γ = 0.5
γ = 0.0
−15
−13
−11
−9
−7
−5
−3
−1
1
3
5
C = −5.0

Figure 5

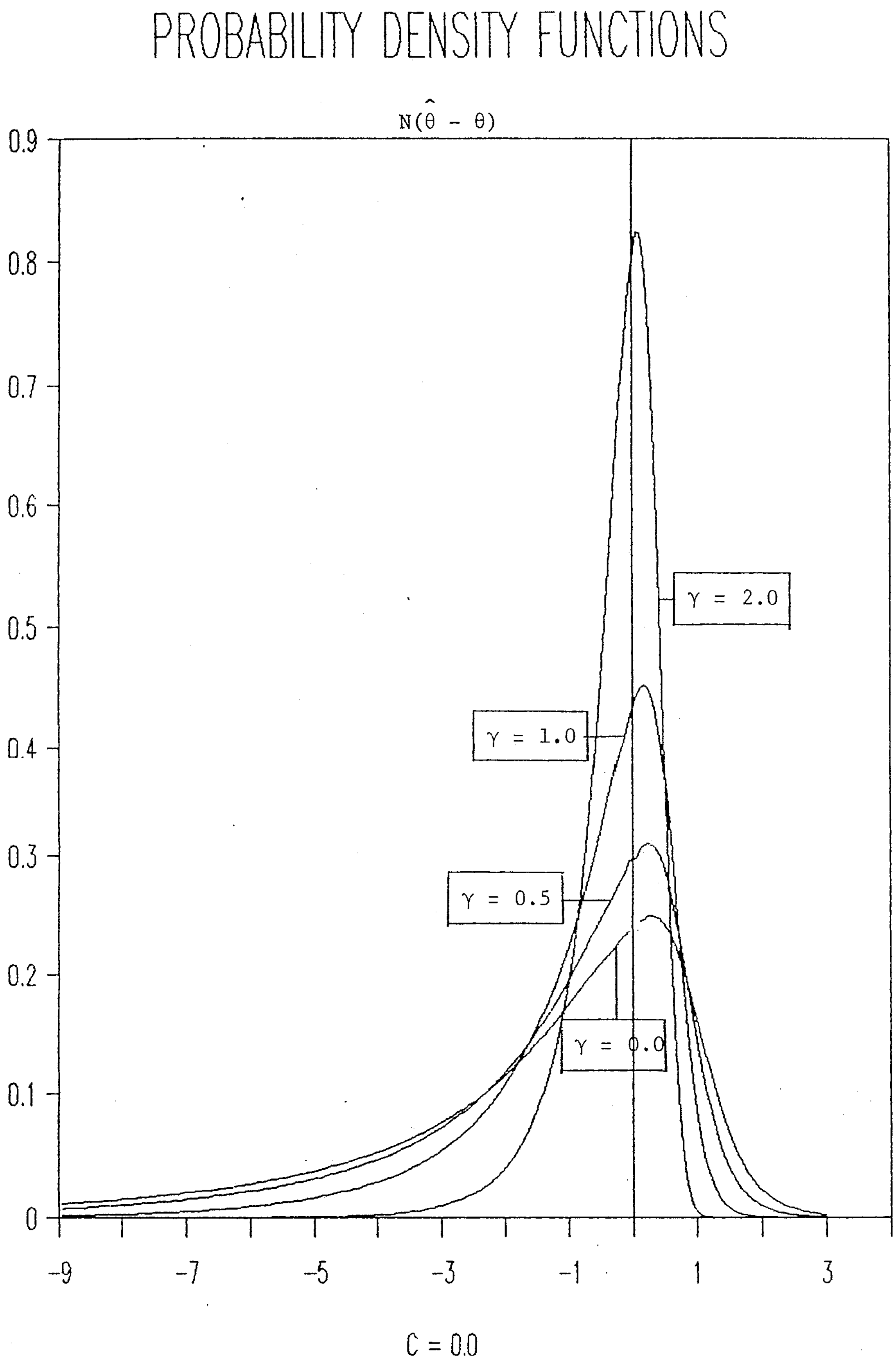
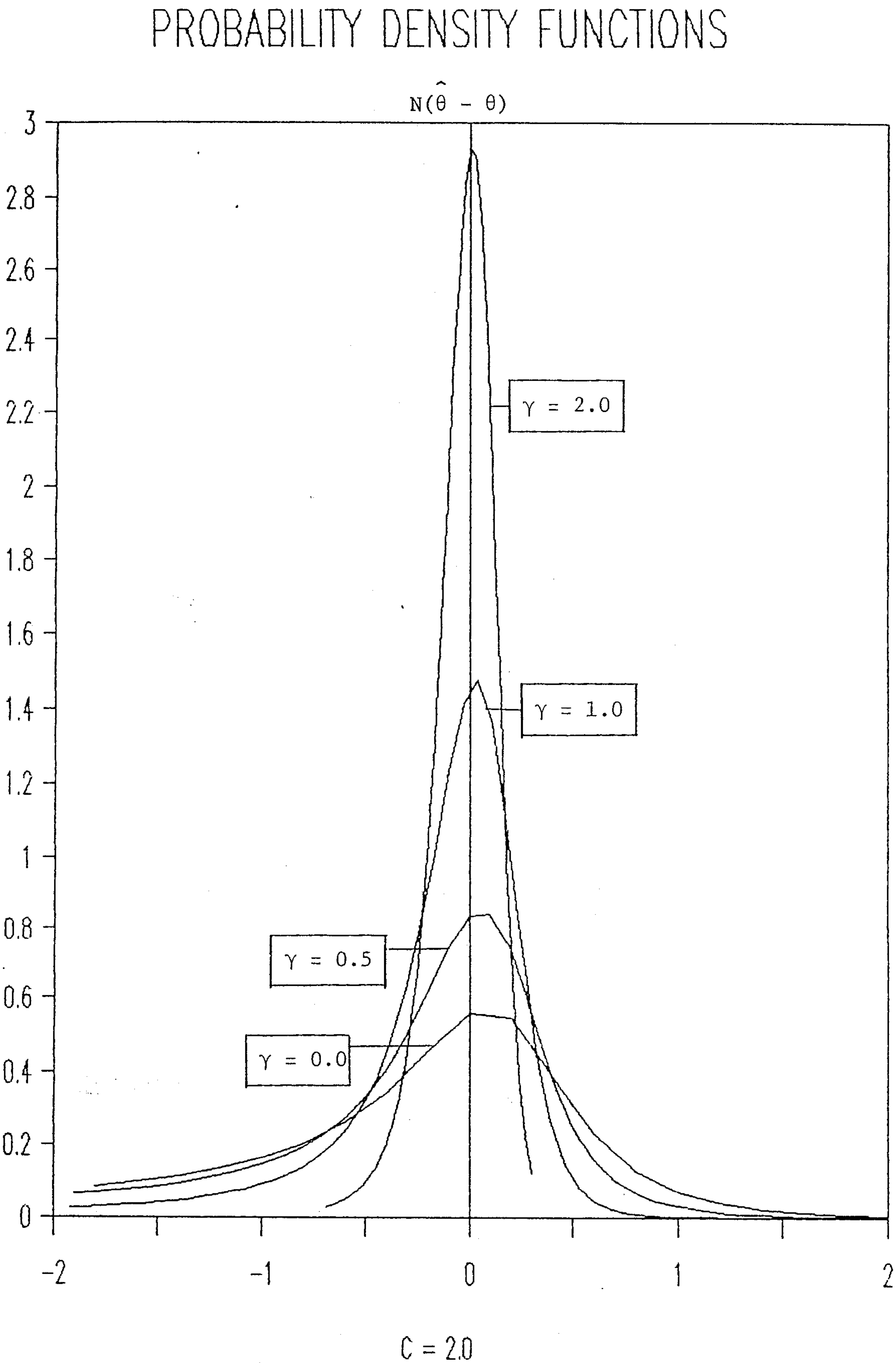
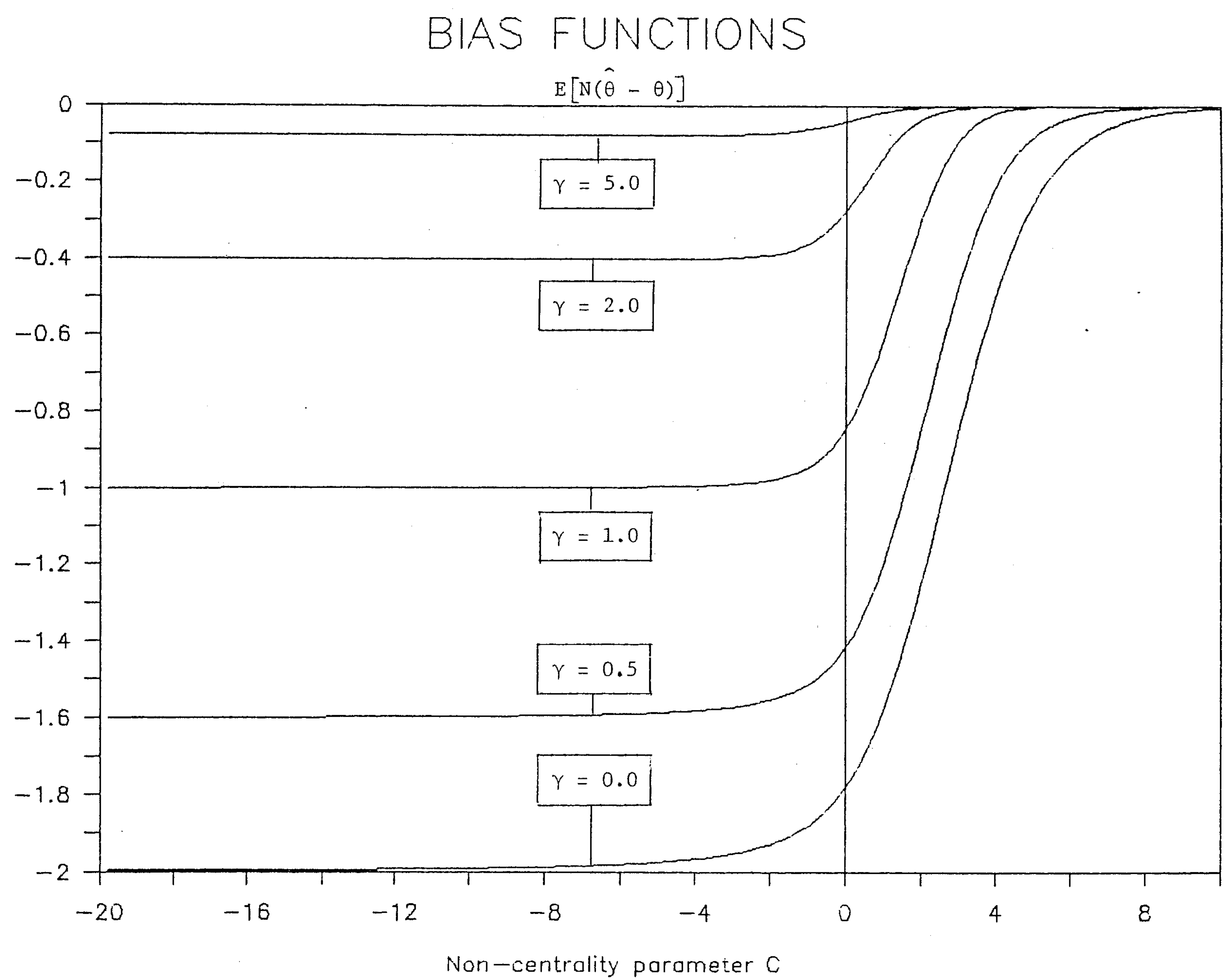

PROBABILITY DENSITY FUNCTIONS
$N(\hat{\theta} - \theta)$
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-9
-7
-5
-3
-1
1
3
$\gamma = 2.0$
$\gamma = 1.0$
$\gamma = 0.5$
$\gamma = 0.0$
C = 0.0

Figure 6



PROBABILITY DENSITY FUNCTIONS
$N(\hat{\theta} - \theta)$
3
2.8
2.6
2.4
2.2
2
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
γ = 2.0
γ = 1.0
γ = 0.5
γ = 0.0
-2
-1
0
1
2
C = 2.0

Figure 7



BIAS FUNCTIONS
$E[N(\hat{\theta} - \theta)]$
0
−0.2
−0.4
−0.6
−0.8
−1
−1.2
−1.4
−1.6
−1.8
−2
γ = 5.0
γ = 2.0
γ = 1.0
γ = 0.5
γ = 0.0
−20
−16
−12
−8
−4
0
4
8
Non−centrality parameter C

Figure 8

MEAN—SQUARED ERROR

$E[N(\hat{\theta} - \theta)]^2$

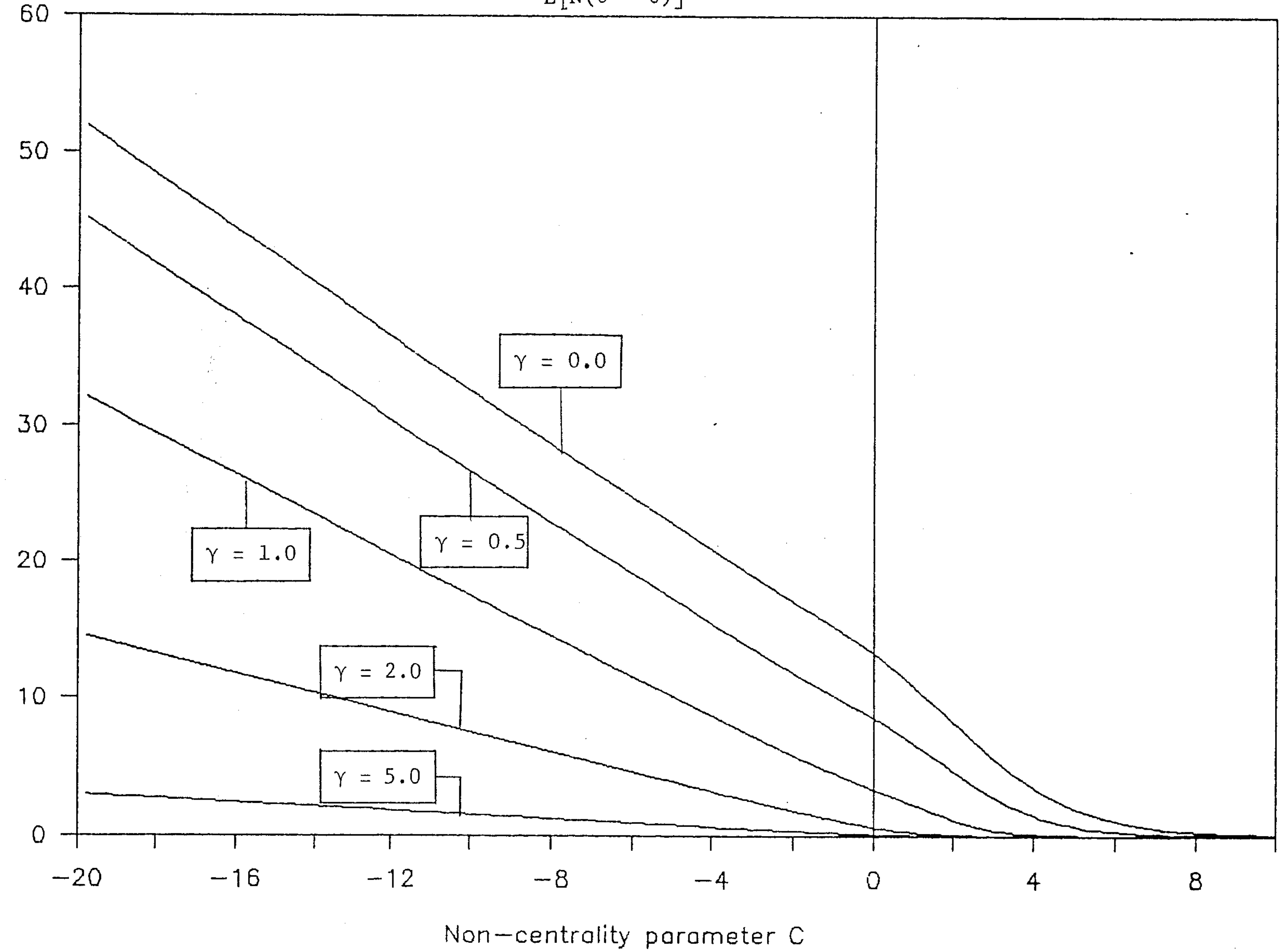
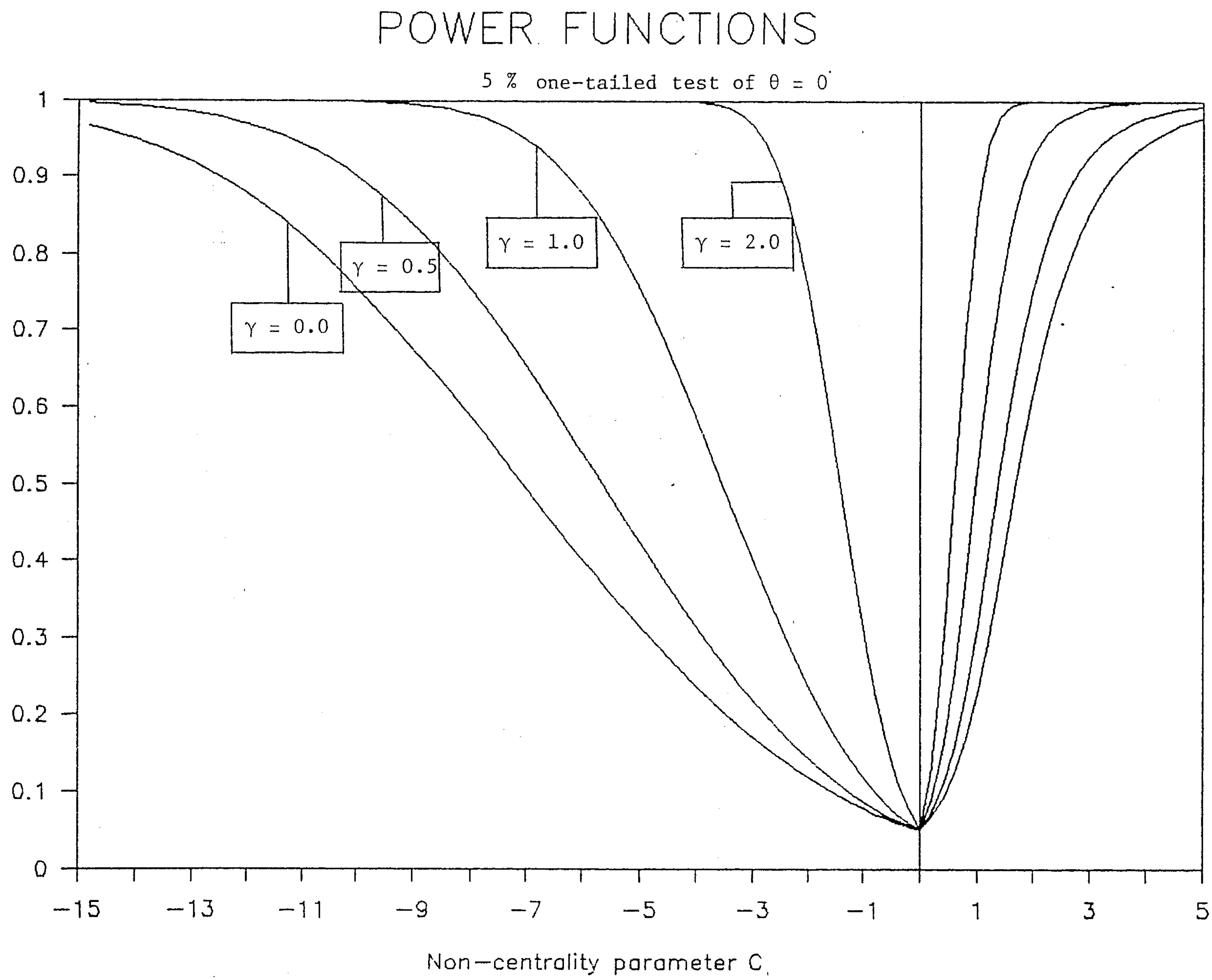


Figure 9


POWER FUNCTIONS
5 % one-tailed test of θ = 0
γ = 0.0
γ = 0.5
γ = 1.0
γ = 2.0
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
−15
−13
−11
−9
−7
−5
−3
−1
1
3
5
Non−centrality parameter C