COWLES FOUNDATION FOR RESEARCH IN ECONOMICS

AT YALE UNIVERSITY

Box 2125, Yale Station
New Haven, Connecticut 06520

COWLES FOUNDATION DISCUSSION PAPER NO. 683

FINITE SAMPLE ECONOMETRICS USING ERA's

Peter C. B. Phillips

November 15, 1983

FINITE SAMPLE ECONOMETRICS USING ERA'S

by

P. C. B. Phillips*

Cowles Foundation for Research in Economics
Yale University

0. ABSTRACT

The paper considers approximate distribution theory as a way to deliver practical improvements over asymptotic methods. Trials of the adequacy of asymptotic series based approximations are conducted and the results exhibit the need for further improvements. An alternative approach is suggested which utilizes a family of extended rational approximants (ERA's). ERA's build on the strength of primitive exact theory or asymptotic series; they allow us to blend information from diverse analytic numerical and experimental sources. The ultimate objective of the approach is to incorporate directly into software constructive functional approximants relevant to statistical practice. An illustration is provided.

---

*This is the text of an invited lecture delivered to the 51st annual meeting of the Japan Statistical Society.

My thanks go to Alison Gold for her help in preparing the graphical illustrations, to Glena Ames for her skill and effort in typing the manuscript and to the NSF for research support under Grant #SES 8218792.

1. INTRODUCTION

My subject today is concerned with attempts to improve upon asymptotic methods in statistical inference. This is a field where scholars in mathematical statistics and in econometrics have made many recent advances. My point of view in this lecture will be that of an econometrician who wants to see such improvements used in empirical work.

Let us start by considering the alternatives. On the one hand we have exact distribution theory. Exact results are available for many statistical estimators and tests, particularly under Gaussian assumptions; and the frontiers of the field are expanding rapidly with modern developments on the algebra of functions of multiple matrix arguments. The final results of such distribution theory often have a generality and a mathematical elegance that is very appealing. But the complexity of exact theory makes numerical work very difficult. Even now, only the simplest exact results are used in empirical work. Researchers in the field inevitably recognize that much of the motivation for their work comes from the intellectual challenge of unlocking the mathematical form of a distribution in a previously unknown case. They understand that exact results are seldom useful in the operational transmission to the empirical researcher of knowledge concerning the sampling properties of common statistical procedures. These considerations combined with the mathematical difficulties of exact theory lead many scholars to the view that approximations must provide the vehicle by which such a transmission is accomplished, as indeed they have in the past through the use of crude asymptotic results. Thus, the introduction of better approximations is seen as a way to deliver practical improvements over asymptotic methods in statistical inference.

The most significant progress on approximations in the last ten years has been based on asymptotic series expansions. These expansions provide refinements of crude asymptotic theory and they apply in very general situations extending to time series problems and nonlinear regressions. They can even be worked out by computer symbolic algebra and built into regression programs. Considerable headway has been achieved on all aspects of this work, ranging from issues of mathematical validity through to problems of practical implementation. Already we are approaching the time when these expansions could be routinely applied in empirical research. But, like new pharmaceutical medicines, these approximations should be well tried in laboratory type conditions before they are dispensed on the general public, here the empirical investigator.

The first part of this paper considers two simple trials of the adequacy of asymptotic series based approximations. These trials help to exhibit certain strengths of the asymptotic series approach. They also indicate the weak features of the approach where the results are not sufficiently reliable for use in empirical research.

The need for further improvements in approximations motivates the second part of the paper. Specifically, we are often interested in improvements which can repair the inadequacies of the asymptotic series approach. It is in this area that some of my own research has recently concentrated. The natural intellectual origin for work on this problem lies in the diverse field of applied mathematics known today as approximation theory. For many years, physicists have drawn upon the rich vein of technique in this field to assist in the solution of some of their more complex problems. Those mathematical statisticians most familiar with the methods of mathematical physics have also made good use of the field. The most celebrated of these are Henry Daniels, Harold Jeffreys and L. C. Hsu who explored the use in statistics of saddlepoint (SP) approximations and the Laplace method. To my knowledge there has been little or no attempt to implement some of the other promising techniques of approximation theory within mathematical statistics. My own work endeavors to take such a step in a systematic way. The approach that I have developed brings into play a new family of extended rational approximants (which we refer to under the acronym ERA). ERA's may build on the strengths of a primitive exact theory or a less flexible approximation method such as an asymptotic expansion; and they allow us to simultaneously blend information from diverse analytic, numerical and experimental sources. In this way the approach seeks to achieve a meaningful integration of earlier directions of research. Its ultimate objective is to incorporate directly into regression software constructive functional approximants to distributions which are relevant to testing problems in statistical practice. This can be done with ERA's in much the same way as computer function routines presently embody approximants (usually rational functions) for the commonly used transcendental functions.

The second part of the paper is devoted to an exposition and illustrative application of the new approach.

2. TWO TRIALS OF ASYMPTOTIC EXPANSIONS

2.1. Testing Linear Restrictions in the Multivariate Linear Model

We work with the linear model

$$y_t = Ax_t + u_t \ , \quad t = 1, \ldots, T \tag{1}$$

where y_t is a vector of n dependent variables, A is an $n \times m$ matrix of parameters, x_t is a vector of independent variables and the u_t are i.i. $N(0,\Sigma)$ errors with Σ positive definite. The hypothesis under consideration takes the form:

$$H_0 : D\,\mathrm{vec}(A) = d \tag{2}$$

where D is a $q \times nm$ matrix of known constants, d is a known vector and $\mathrm{vec}(A)$ stacks the columns of A.

In econometrics, this framework is a common one. The matrix D can take many different forms and it is necessary to allow for this degree of generality rather than restrict D to a Kronecker product form as is common practice in multivariate statistical analysis [2]. For example, in applied demand analysis where y_t is a vector of expenditures or budget shares on different commodity groups and x_t is a vector of prices and total expenditure (or income) the theory of consumer demand implies certain restrictions on the parameter matrix A. When the model is based on the Rotterdam system [10] or almost ideal system [1] the leading submatrix of A of dimension $n \times n$ is symmetric. The hypothesis (2) then involves $q = n(n-1)/2$ restrictions.

This is an important case for empirical econometric practice when the restriction coefficient matrix D cannot be reduced to a Kronecker representation. A typical value of n in applied studies of aggregate consumer demand lies in the range 7-14, although cases as high as $n = 40$ have been studied empirically. When $n = 10$ the symmetry conditions of demand analysis lead to $q = 45$ restrictions in H_0. Thus, H_0 can achieve considerable parameter parsimony in econometric studies of demand; but its validity is important to subject to statistical test. The fact that sample sizes in such studies are often less than 50 observations makes the need for a good sampling theory more acute in such a context. For, in such cases when H_0 severely reduces the dimension of the parameter space, asymptotic theory and (as we shall see) its series refinements can lead to very low inferential accuracy in statistical testing.

Unrestricted estimation of (1) leads to $A^* = Y'X(X'X)^{-1}$ where $Y' = [y_1, \ldots, y_T]$, $X' = [x_1, \ldots, x_T]$; and the covariance matrix of $\text{vec}(A^*)$ is $N \otimes \Sigma$ where $N = (n_{ij}) = (X'X)^{-1}$. We consider the commonly used Wald statistic for testing H_0 given by:

$$(3) \qquad W = (D\,\text{vec}\,A^* - d)'\left\{D(N \otimes \Sigma^*)D'\right\}^{-1}(D\,\text{vec}\,A^* - d)$$

where $\Sigma^* = T_m^{-1}Y'(I - P_X)Y$, $P_X = X(X'X)^{-1}X'$ and $T_m = T-m$.

The Edgeworth expansion to $O(T^{-1})$ of the cumulative distribution function (cdf) of W is given by the following formula:

$$(4) \qquad P(W \leq x) = F_q\left(x\left[1 - \frac{a_0 + a_1x/(q+2)}{2T_mq}\right]\right) + o(T^{-1})$$

where $F_q(\ )$ denotes the cdf of a χ_q^2 variate and the coefficients that appear in (4) are given by the tensor sums:

$$(5) \qquad a_0 = n_{ij}n_{k\ell}(\alpha_{jk}\alpha_{i\ell} + \alpha_{jki\ell} - \alpha_{ijk\ell})$$

$$(6) \qquad a_1 = n_{ij}n_{k\ell}(\alpha_{jk}\alpha_{i\ell} + \alpha_{jki\ell} + \alpha_{ijk\ell})$$

and $\quad \alpha_{rs} = \text{tr}(\Sigma D_r'Q^{-1}D_s)$,

$\alpha_{rstu} = \text{tr}(\Sigma D_r'Q^{-1}D_s\Sigma D_t'Q^{-1}D_u)$,

$Q = D(N \otimes \Sigma)D'$,

$D = [D_1 \vdots D_2 \vdots \cdots \vdots D_m]$,

$D_i = q \times n$ submatrix of restriction coefficients relating to the i^{th} independent variable.

The expansion (4) was first obtained by Rothenberg [9] in unpublished work and special cases of it have been known in multivariate analysis for some time. Appendix A gives a direct derivation of (4) using a method and a notation different from Rothenberg's. The special cases reported in [2] can be deduced immediately from (4) and its noncentral extension for local alternatives.

When $D = g' \otimes D_*$ for some m vector g and $q \times n$ matrix D_* , the restrictions (2) take the simpler form $D_*Ag = d$. We assume D_* to have full rank $q < n$ and the Wald statistic (3) then reduces to the form:

$$(7) \qquad W = (D_*A^*g - d)'[D_*\Sigma^*D_*']^{-1}(D_*A^*g - d)/g'Ng$$

$$= \ell'[D_*\Sigma^*D_*']^{-1}\ell , \quad \ell = (D_*A^*g - d)/(g'Ng)^{1/2}$$

which we identify as Hotelling's T^2 statistic. Specifically, $(T_m - q + 1)W/qT_m$ has an F_{q,T_m-q+1} distribution.

In this special case the coefficient formulae (5) and (6) reduce to $a_0 = q^2$ and $a_1 = q(q+2)$, and the expansion (4) has the form

$$(8) \qquad P(W \leq x) = F_q\left(x\left[1 - \frac{q+x}{2T_m}\right]\right) + o(T^{-1}) .$$

We deduce from (8) that, if x_α is the critical value for the χ_q^2 distribution at the upper $100\alpha\%$ level, then

$$(9) \qquad x_\alpha^* = x_\alpha\left\{1 + \frac{q + x_\alpha}{2T_m}\right\}$$

is the corresponding critical value for (8) correct to $O(T^{-1})$.

With these formulae comparisons of the asymptotic, Edgeworth corrected and exact critical values for tests of various sizes are simple to perform. The table below provides some numerical computations for empirically relevant values of T_m and q at the 5% level of significance.

Asymptotic, Edgeworth Corrected and Exact Critical Values

for the Wald Statistic (7) at a nominal 5% significance level

T_m \ q	1	5	8
10	x_α = 3.84 x_α^* = 4.77 Exact = 4.96	x_α = 11.07 x_α^* = 19.96 Exact = 36.58	x_α = 15.51 x_α^* = 33.74 Exact = 108.53
20	x_α = 3.84 x_α^* = 4.31 Exact = 4.35	x_α = 11.07 x_α^* = 15.52 Exact = 17.81	x_α = 15.51 x_α^* = 24.62 Exact = 40.00
30	x_α = 3.84 x_α^* = 4.15 Exact = 4.17	x_α = 11.07 x_α^* = 14.03 Exact = 14.94	x_α = 15.51 x_α^* = 21.58 Exact = 32.67

x_α = asymptotic critical value for test at $100\alpha\%$ nominal significance level.

x_α^* = Edgeworth corrected (to $O(T^{-1})$) critical value for same test.

These numerical computations make it clear that the number of restrictions q plays a critical role in the adequacy of asymptotic based approximations. For q small (unity or close to unity) the error on the asymptotic critical value is less than 25% even for quite small degrees of freedom T_m . Moreover, for these cases, a large fraction of the error is removed by the Edgeworth correction. Thus for $q = 1$, $T_m = 10$ the error on the asymptotic is 22½% while the Edgeworth correction reduces this to 3.8%, which is acceptable by most standards. On the other hand when

q is moderate or large (relative to T_m) the error on the asymptotic formula can be very substantial indeed. Moreover, while the Edgeworth correction works to remove the error in this case the error that remains after the correction is still substantial. Thus, when $q = 8$, $T_m = 10$ the asymptotic error is 86% and the error after the Edgeworth correction is 69%. Even when T_m trebles in value to 30 the latter error is still as high as 34%. By most standards these incorrect critical values would lead to seriously misleading inferences in empirical work.

Evidently, the approach to the limiting χ^2 distribution is much slower as q gets large. As the error on the asymptotic formula increases the Edgeworth correction has more work to do in order to narrow the gap to the exact critical value. It seems reasonable to deduce from these computations the following general verdict concerning the adequacy of Edgeworth corrections: asymptotic expansions of low order (to $O(T^{-1})$) do well and can provide significant improvements when the error on the crude asymptotic formula is not too large (in rough terms under 25%); but when the asymptotic error is large (above 75%) the asymptotic expansion is usually poor and can, in some cases, be worse than the crude asymptotic.

We will see an instance of the last mentioned phenomena in our second example.

2.2. The Distribution of the Serial Correlation Coefficient

We will consider the autoregression

$$(10) \quad y_t = \alpha y_{t-1} + u_t , \quad t = \ldots -1, 0, 1, \ldots$$

in which the u_t are i.i.d. $N(0, \sigma^2)$. Our attention will concentrate on the distribution of the least squares estimator of α in (10) given by $\hat{\alpha} = \Sigma_{t=1}^{T} y_t y_{t-1} / \Sigma_{t=1}^{T} y_{t-1}^2$. This statistic is a non circular serial correlation coefficient and it can be used to estimate the correlation between consecutive observations in an ordered sample. Unlike most correlation coefficients, however, its distribution is supported over the entire real line and approximations to the distribution which remain adequate in the tails are difficult to obtain.

In earlier work [3, 4] the author derived the following explicit expression for the Edgeworth series expansion of the distribution of $T^{-1/2}(1-\alpha^2)^{-1/2}(\hat{\alpha}-\alpha)$:

$$(11) \quad P\left(\frac{\sqrt{T}(\hat{\alpha}-\alpha)}{\sqrt{1-\alpha^2}} \leq x\right) = \Phi(x) + \varphi(x)\left[T^{-1/2}\alpha(1-\alpha^2)^{-1/2}(1+x^2) + (1/4T)(1-\alpha^2)^{-1}\left\{(1-\alpha^2)x + (1+\alpha^2)x^3 - 2\alpha^2x^5\right\}\right] + O(T^{-3/2})$$

where $\Phi(x)$ and $\varphi(x)$ denote the cdf and pdf of the $N(0,1)$ distribution, respectively. The corresponding series expansion for the density of $T^{-1/2}(1-\alpha^2)^{-1/2}(\hat{\alpha}-\alpha)$ is obtained by differentiation of (11).

As explored in [3] the approximations implied by (11) deteriorate in performance as $\alpha \uparrow 1$. This is to be expected since the asymptotic normality of $\hat{\alpha}$ (on which (11) is based) holds for $|\alpha| < 1$ but not for $\alpha = 1$ nor for $|\alpha| > 1$. The following Figures exhibit this phenomena by graphs which trace the asymptotic and Edgeworth approximations (to $O(T^{-1})$) against the exact density for $T = 5, 10, 20, 50$ and $\alpha = 0, 0.9$.

Figures 1, 2 and 3 about here

When $\alpha = 0$, we see from Figures 1, 2 and 3 that the approach to normality is reasonably fast. This is partly explained by the symmetry of the exact distribution and the fact that the error on the asymptotic distribution is $O(T^{-1})$ for this case. Thus when $T = 20$ (Figure 3) the error on the asymptotic is generally quite small relative to the exact density. The Edgeworth correction is so good in this case that the exact and approximate densities are almost indistinguishable on the scale of Figure 3. When $T = 5$, however, the errors on the asymptotic are more substantial and the corresponding errors on the Edgeworth approximation (shown in Figure 1) are of the same magnitude but generally of different sign, due to overcorrection. We may conclude that for $\alpha = 0$ the Edgeworth approximation performs well for moderate and large sample sizes but needs refinement for small T.

When $\alpha = 0.9$ the situation changes dramatically. We see from Figures 4 and 6 that the asymptotic delivers a very poor approximant not only for T as small as 10 but also for T as large as 50. The rate of convergence to normality is here very slow indeed. Correspondingly the Edgeworth approximations (shown in Figures 4, 5 and 6) are also poor and are particularly unreliable in the tail regions where they are worse than the crude asymptotic approximant. Only in the central body of the distribution for $T = 50$ is the approximant adequate. This visually apparent difference in the rate of convergence is explained by the fact that at $x = 0$ the errors on the Edgeworth approximant are of $O(T^{-2})$ rather than $O(T^{-3/2})$. We conclude that for the case of $\alpha = 0.9$ the errors on the asymptotic distribution are sufficiently large (even for large sample sizes) to make the Edgeworth corrections quite unreliable.

Figures 4, 5 and 6 about here

---

3. AN ALTERNATIVE APPROACH BASED ON CONSTRUCTIVE FUNCTIONAL APPROXIMANTS

In [5] and [6] the author has recently been exploring mechanisms by which we can build on the strengths of less flexible methods such as asymptotic expansions and primitive exact theory with the object of providing improvements in distribution approximants where they are most needed. Before we describe the approach taken in these papers it is worth pointing out that improvements inevitably carry a cost. Just as asymptotic expansions yield improvements over crude asymptotic theory by utilizing more information about the underlying random processes, so too do the methods I will describe. However, they have the additional advantage of a flexibility which enables them to blend information from diverse analytic and numerical sources.

3.1. pdf Approximation

We start by considering the problem of approximating a certain probability density function (pdf) in $C_0[\infty - , \infty]$, the class of continuous, positive valued functions that vanish at infinity. For the direct approximation of densities, we introduce the following family of extended rational approximants (ERA's) of maximal degrees m and n :

$$\text{(12)} \quad era(x;\gamma) = s(x)[m/n](x;\gamma) = s(x)\frac{\gamma_0 + \gamma_1 x + \ldots + \gamma_m x^m}{1 + \gamma_{m+1} x + \ldots + \gamma_{m+n} x^n}$$

where $s(x)$ is a given coefficient function usually in $C_0[-\infty,\infty]$ and γ is the vector of rational coefficients.

The first steps in the practical implementation of the approach require important elements of judgment in three areas: (i) choice of the coefficient function $s(x)$; (ii) selection of the degree of the rational function $[m/n]$; and (iii) determination of the coefficients in γ . The problem is one of constructive functional approximation within the general family defined by (12). The solution to this problem in any particular case will rely intimately on the information that is available about the true distribution. Typically, we will want the approximant to embody as much analytic and reliable numerical data about the distribution as possible. This directly affects the choice of $s(x)$ and the prescribed degree of $[m/n]$. As argued in [5], theoretical and practical considerations often suggest a specialization of the family (12) to that in which numerator and denominator polynomials are of the same degree (i.e., $m = n$) .

Primitive analytic structures will frequently provide a good choice of $s(x)$. Obvious examples are: (i) knowledge that the true distribution has finite moments up to a certain order ($< \mu$, say, where μ = supremal moment exponent), in which case we may set $s(x) = [1 + |x|^{\mu+1}]^{-1}$; knowledge that a primitive form of the distribution is of a simple type, such as the Cauchy distribution in the case of the limited information maximum likelihood estimator (see [7]) leading to the form $s(x) = [\pi(1 + x^2)]^{-1}$; knowledge of the Edgeworth approximation to a certain order, say

$$(13) \quad ed(x) = \varphi(x)\left\{1 + T^{-1/2}q_1(x) + T^{-1}q_2(x)\right\} , \quad \varphi(x) = (2\pi)^{-1/2}\exp[-x^2/2]$$

in which case we may set $s(x) = ed(x)$ except for those regions in which $ed(x)$ becomes non positive where we may splice in a scaled primitive density such as the standard normal or univariate t . In addition to analytic structures like the above we may also use numerical data in the construction of $s(x)$, an example being nonparametric density estimates based on experimental simulations. These choices of $s(x)$ are discussed in greater detail with some numerical illustrations in [5] and [6].

The main role of the rational coefficients γ in $[m/n]$ is to build on the strengths of $s(x)$ as a primitive approximant and to allow the direct use of numerical data in the approximant when $s(x)$ is furnished by rigid analytic formulae such as Edgeworth series. In [5] the author explored the use of multiple point Padé procedures based on Taylor series expansions of the true pdf at certain points including infinity. For example we may utilize the two series expansions:

$$(14) \quad \text{pdf}(x) = \alpha_0 + \alpha_1 x + \alpha_2 x^2 + \ldots , \quad x \to 0$$

$$(15) \quad \text{pdf}(x) = |x|^{-\mu-1}\left\{\beta_0 + \beta_1/x + \beta_2/x^2 + \ldots\right\} , \quad |x| \to \infty$$

to select γ so that to a certain order

$$\text{era}(x;\gamma) = s(x)[m/n](x;\gamma)$$

has the same asymptotic behavior at $x = 0, \pm\infty$ as (14) and (15). The examples in [5] and [6] show that this multiple point Padé procedure can work very well even for low degree rational approximants ($m, n \leq 4$ say). Direct numerical inversion of the characteristic function also provides pointwise computations of $\text{pdf}(x)$ that can be used in such Padé procedures. The example we give later makes systematic use of this latter approach.

[6] develops a theory of approximation for the class $C_0[-\infty,\infty]$ in terms of the uniform error norm:

$$\|\text{pdf} - \text{era}\| = \sup_x |\text{pdf}(x) - \text{era}(x)| .$$

This theory establishes the existence of a unique best uniform approximant

$era^* = era(x;\gamma^*)$ to any pdf $\in C_0[-\infty,\infty]$, where γ^* is such that

$$\|pdf - era^*\| = \inf_\gamma \sup_x |pdf(x) - era(x;\gamma)| .$$

Denseness and characterization theorems are also established. The latter theorem verifies the following error alternation property of the best approximant: if the number of alternations of the error $e(x) = pdf(x) - era(x)$ (i.e., the number of consecutive points at which $e(x)$ attains a maximum with alternate changes in sign) is at least $N = n+m+2$ then $era(x;\gamma)$ is the best approximant to $pdf(x)$ in the family defined by (12). These theorems extend the classical theory of approximation (given, for instance, in [8]) to density approximation over the entire real line.

3.2. cdf Approximation

The following modification of the family of approximants given by (12) allows us to treat cdf and tail probability functions directly:

$$(16) \quad Era(x;\gamma) = S(x)[n/n](x;\gamma) = S(x)\frac{\gamma_0 + \gamma_1 x + \ldots + \gamma_n x^n}{1 + \gamma_{n+1} x + \ldots + \gamma_{2n} x^n} , \quad n \text{ even}$$

where $S(x)$ is a primitive cdf, perhaps directly of the form $\int_{-\infty}^{x} s(t)dt$, with the properties that $S \uparrow 1$ as $x \uparrow \infty$, $\downarrow 0$ as $x \downarrow -\infty$ and ≥ 0 for all x . It will sometines also be useful to set $\gamma_n = \gamma_{2n}$ if we wish to preserve the correct asymptotic (as $x \to \infty$) behavior. Approximants of the form (16) yield density approximations by simple differentiation and they can be used directly for the tail area probability calculations that are needed for statistical testing.

3.3. Composite Functional Approximation

The research reported in [5] and [6] is currently being extended to treat the inherent multidimensionality of most practical problems by functionalization of the rational coefficients in (12) and (16). This work is being conducted jointly with P. C. Reiss of Stanford University. The process can be demonstrated by considering the simplest problem in which a cdf is parameterized by T (the sample size) over a discrete point set $\{T \geq T_0\}$. Let this function be denoted by $cdf_T(x)$ and suppose that $S_T(x)$ is a primitive cdf, also indexed by T , suitable for use in (16). Further, let

$$\text{(17)} \quad Era\Big(x;\ \gamma(T)\Big) = S_T(x)[n/n]\Big(x;\ \gamma(T)\Big)$$

be the best uniform approximant to $cdf_T(x)$ for a given T , where $\gamma(T)$ is the vector of rational coefficients. We may now construct a composite approximant to $cdf_T(x)$ by direct approximation to $\gamma(T)$ over T . Specifically, we introduce the family

$$\text{(18)} \quad [r/s]_i(\eta) = \frac{\alpha_{i0} + \alpha_{i1}\eta + \ldots + \alpha_{ir}\eta^r}{1 + \beta_{i1}\eta + \ldots + \beta_{is}\eta^s}$$

of rational coefficient function approximants for $\gamma_i(T)$; $\eta = T^\nu$ is a suitably chosen power of T as the argument of the function. Frequently $\eta = T^{-1/2}$ will be a good choice in practical applications.

We now let $[r/s]_i^*$ denote the best uniform approximant to $\gamma_i(T)$ over the discrete point set $\{T \geq T_0\}$. Then

$$\text{(19)} \quad Era_T(x) = Era(x;\ [r/s]^*)$$

is called the best composite approximant to $cdf_T(x)$ over $x \in (-\infty,\infty)$ and $\{T \geq T_0\}$.

3.4. Preservation of Asymptotics

It is interesting to explore ways in which the composite approximant (19) may be formulated to preserve the validity of asymptotic series approximations. This is important from the point of view of the refinement of asymptotic methods and from a practical standpoint. The latter is significant because the point set over which the functionalization (18) is relevant itself becomes finite, say $\{T_0 \leq T \leq T^0\}$ when (19) builds on a primitive asymptotic theory, such as an Edgeworth series, which is known to be adequate for $T > T^0$.

The simplest procedure involves the refinement of Edgeworth asymptotic series to improve their performance in regions of the distribution and for parameter domains where there are inherent weaknesses (see Section 2 for examples). Suppose we are working with a statistic α_T for which $T^{1/2}(\alpha_T - \alpha)$ has a limiting normal distribution and admits a valid Edgeworth series under general conditions at least to $O(T^{-1})$. We may write

$$(20) \quad P\left(T^{1/2}(\alpha_T - \alpha) \leq z\right) = \Phi\left(z/\sigma + T^{-1/2}p_1(z/\sigma) + T^{-1}p_2(z/\sigma)\right) + o(T^{-1})$$

and then

$$(21) \quad Ed_T(x) = \Phi\left(T^{1/2}(x-\alpha)/\sigma + T^{-1/2}p_1(T^{1/2}(x-\alpha)/\sigma) + T^{-1}p_2(T^{1/2}(x-\alpha)/\sigma)\right)$$

is the associated Edgeworth approximation to the distribution of α_T .

Now set $S_T(x) = Ed_T(x)$ as the primitive approximant in (16) and the resulting ERA for the distribution of α_T has the general form:

(22) $\quad \mathrm{Era}(x;\gamma) = \mathrm{Ed}_T(x)[n/n](x;\gamma)$.

As usual n is an even integer (say $n = 2m$) and we confine our attention to positive, bounded rational functions $[n/n]$ over $(-\infty,\infty)$. Decomposing $[n/n]$ into partial fractions and noting that there are no real poles we may write

(23) $$[n/n](x;\gamma) = c_0 + \sum_{j=1}^{m}\left\{\frac{c_j}{x - d_j} + \frac{\bar{c}_j}{x - \bar{d}_j}\right\}$$

where c_0 is real, $c_j = c_{j1} + ic_{j2}$ and $d_j = d_{j1} + id_{j2}$.

Best approximation in the class (22) to a given $\mathrm{cdf}_T(x)$ leads to the functional dependence $\gamma = \gamma(T)$ in the rational coefficients, as in (17) above. The reparameterization suggested by (23) is useful in the development of composite approximants which preserve asymptotic behavior. It is therefore convenient to write $\gamma(T)$ in future as $(c(T), d(T))$.

We now use coefficient function approximants in the family (18) for the components of $c(T)$ and $d(T)$. First let us suppose that $\mathrm{Ed}_T(x)$ delivers sufficient accuracy as an approximant of $\mathrm{cdf}_T(x)$ for $T > T^0$. Next we develop a first stage ERA of the form (22) leading to the coefficient functions $(c(T), d(T))$ under the new parameterization. Then we utilize the family (18) to produce best coefficient approximants of the form:

$$(24) \quad c_{0T} = \frac{c_{00} + c_{01}\eta + \ldots + c_{0r}\eta^r}{1 + e_{01}\eta + \ldots + e_{0s}\eta^s}, \quad \eta = T^{-1/2}$$

$$(25) \quad c_{jkT} = \frac{c_{jk0} + c_{jk1}\eta + \ldots + c_{jkr}\eta^r}{1 + e_{jk1}\eta + \ldots + e_{jks}\eta^s}, \quad \eta = T^{-1/2} \quad (j = 1, \ldots, m;\ k = 1,2)$$

$$(26) \quad d_{jkT} = \frac{d_{jk0} + d_{jk1}\eta + \ldots + d_{jkr}\eta^r}{1 + f_{jk1}\eta + \ldots + f_{jks}\eta^s}, \quad \eta = T^{-1/2} \quad (j = 1, \ldots, m;\ k = 1,2)$$

over the interval of discrete points $\{T_0 \leq T \leq T^0\}$; and

$$(27) \quad c_{0T} = 1\ , \quad c_{jkT} = 0 \quad (j = 1, \ldots, m;\ k = 1,2)$$

for all $T > T^0$.

When this second stage is completed we have the composite approximant

$$Era_T(x) = Era(x;\ c_T,\ d_T) = Ed_T(x)[n/n](x;\ c_T,\ d_T)$$

which preserves the asymptotic behavior of $Ed_T(x)$ as $T \to \infty$ and which modifies the Edgeworth series approximant in the domain $\{T_0 \leq T \leq T^0\}$.

3.5. Application

Research on a major application of the above procedure to tests for serial correlation and for unit roots is currently underway in conjunction with the author's research associate P. C. Reiss. This work is concentrating on the distribution of the first order non circular serial correlation discussed earlier in Section 2.2 and builds on the primitive analytic form of the Edgeworth approximation given by (11) in this case. At present our research suggests that substantial improvements can be achieved with low order rational functions. This is exemplified by the error curves shown in Figure 7 which refer to the distribution of $\hat{\alpha}$ when $\alpha = 0$, $T = 5$ and the composite approximant has the form $Era_T(x) = Ed_T(x)[2/2](x; \gamma(T))$ with $\gamma(T) = [4/2](\eta)$ and $\eta = T^{-1/2}$. Details of this work will appear elsewhere.

Figure 7 about here

4. FINAL REMARKS

Composite functional approximants of the type discussed in the last section can be directly written into regression software. With the use of some additional minor subroutines they can be employed to revise asymptotic critical regions or significance levels on a simple "user call" basis.

Many additional applications of this machinery are now possible. One that is presently within reach and probably well worth the computational cost involves an extension to the distribution of higher order serial correlation coefficients. The final composite approximants in this case could then be directly incorporated into the first stage of the Box Jenkins modeling procedure.

APPENDIX

We derive the Edgeworth expansion to $O(T^{-1})$ of the Wald statistic

(A1) $$W = (D \text{ vec } A^* - d)'\left\{D(N \otimes \Sigma^*)D'\right\}^{-1}(D \text{ vec } A^* - d)$$

given by equation (3) in the paper. We start with the characteristic function of W :

$$cf_W(s) = \int e^{isW} pdf(A^*, \Sigma^*) dA^* d\Sigma^*$$

(A2) $$= \frac{1}{(2\pi)^{q/2+nT_m/2} \Gamma_n\left(\frac{T_m}{2}\right)(\det \Sigma)^{T_m/2}(\det Q)^{1/2}}$$

$$\int \exp\left\{-\frac{1}{2}(a-d)'[Q^{-1} - 2isT_m K^{-1}](a-d)\right\}$$

$$\text{etr}\left(-\frac{1}{2}\Sigma^{-1}S\right)(\det S)^{T_m/2-(n+1)/2} da dS$$

where

$$Q = D(N \otimes \Sigma)D' \ , \quad K = D(N \otimes S)D' \quad \text{and} \quad a = D \text{ vec } A^* \ .$$

We transform $S \rightarrow \Sigma^{-1/2} S \Sigma^{-1/2} = B$ and upon evaluation of the normal integral in (A2) we find:

(A3) $$cf_W(s) = \frac{1}{2^{nT_m/2}\Gamma_n\left(\frac{T_m}{2}\right)} \int_{B>0} \text{etr}\left(-\frac{1}{2}B\right)(\det B)^{T_m/2-(n+1)/2}$$

$$\left(\det\left[I - 2isT_m Q^{1/2}\{\overline{D}(N \otimes B)\overline{D}'\}^{-1}Q^{1/2}\right]\right)^{-1/2} dB$$

$$= E_B\left[\left(\det[I - 2isT_m Q^{1/2}\{\overline{D}(N \otimes B)\overline{D}'\}^{-1}Q^{1/2}]\right)^{-1/2}\right]$$

where $\bar{D} = D(I \otimes \Sigma^{1/2})$ and the expectation E_B is taken with respect to a central Wishart variate with covariance matrix I_n and degrees of freedom T_m.

We define

$$G(B) = \bar{D}(N \otimes B)\bar{D}'$$

$$g(B) = \ell n(\det R) = \ell n\left(\det[I - 2isT_m Q^{1/2} G^{-1} Q^{1/2}]\right)$$

and

$$h(B) = \exp\left\{-\frac{1}{2}g(B)\right\} .$$

Evaluated at $B_0 = E_B(B) = T_m I_n$ these functions become:

$$G_0 = T_m^{-1}Q \ , \quad R_0 = (1 - 2is)I_q \ , \quad h(B_0) = (1 - 2is)^{-q/2} .$$

The Taylor expansion of $g(B)$ at B_0 to the second order is:

$$\text{(A4)} \qquad q\ \ell n(1 - 2is) + \frac{2is}{T_m(1 - 2is)}\mathrm{tr}(Q^{-1}dG) - \frac{2is}{T_m^2(1 - 2is)}\mathrm{tr}(dGQ^{-1}dGQ^{-1})$$

where $dG = \bar{D}(N \otimes dB)\bar{D}'$ and where we interpret the differential as $dB = B - T_m I_q$. We may also expand $h(B)$ using (A4) as follows:

$$h(B) = \exp\left\{-\frac{1}{2}g(B)\right\}$$

$$\text{(A5)} \quad \sim (1-2is)^{-q/2}\left[1 + \frac{is}{T_m(1-2is)}\mathrm{tr}(Q^{-1}dG) - \frac{1}{T_m^2}\left\{\frac{s^2}{(1-2is)^2} - \frac{is}{1-2is}\right\}\mathrm{tr}(dGQ^{-1}dGQ^{-1})\right.$$

$$\left. - \frac{s^2}{2T_m^2(1-2is)^2}\left(\mathrm{tr}(Q^{-1}dG)\right)^2\right] .$$

Since $E_B(dB) = 0$ and $E[\text{vec}(dB)\text{vec}(dB)'] = T_m(I_{n^2} + K_n)$, where K_n is the commutation matrix of order n^2, we deduce that:

(A6) $$E\left\{tr(Q^{-1}dG)\right\}^2 = T_m n_{ij} n_{k\ell}(\alpha_{ijk\ell} + \alpha_{ij\ell k}) = 2T_m n_{ij} n_{k\ell}\alpha_{ijk\ell}$$

(A7) $$E\left\{tr(Q^{-1}dGQ^{-1}dG\right\} = T_m n_{ij} n_{k\ell}(\alpha_{jk}\alpha_{i\ell} + \alpha_{jki\ell})$$

where

$$\alpha_{rs} = tr(\Sigma D_r' Q^{-1} D_s)$$

$$\alpha_{rstu} = tr(\Sigma D_r' Q^{-1} D_s \Sigma D_t' Q^{-1} D_u)$$

$$N = (n_{ij})$$

$$D = [D_1 \vdots D_2 \vdots \cdots \vdots D_m]$$

and D_i is the $q \times n$ submatrix of restriction coefficients relating to the i^{th} exogenous variable in (1) and (2). The usual summation convention of a repeated index applies in (A6) and (A7). Using these coefficients in the expectation of (A5), we find the following expansion of the characteristic function of W to $O(T^{-1})$:

(A8) $$cf_W(s) = (1 - 2is)^{-q/2}\left[1 + \frac{a_0 is}{2T_m(1 - 2is)} + \frac{a_1 is}{2T_m(1 - 2is)^2}\right] + o(T^{-1})$$

where

(A9) $$a_0 = n_{ij} n_{k\ell}(\alpha_{jk}\alpha_{i\ell} + \alpha_{jki\ell} - \alpha_{ijk\ell})$$

(A10) $$a_1 = n_{ij} n_{k\ell}(\alpha_{jk}\alpha_{i\ell} + \alpha_{jki\ell} + \alpha_{ijk\ell}) .$$

Upon inversion (A8) yields the following approximation to the cumulative distribution of W :

$$\text{(A11)} \quad P(W \leq x) = F_q\left(x\left[1 - \frac{a_0 + a_1 x/(q+2)}{2T_m q}\right]\right) + o(T^{-1})$$

where $F_q(\ )$ is the cdf of a χ_q^2 variate.

REFERENCES

[1] Deaton, A. and J. Muellbauer (1980). "An almost ideal demand system," American Economic Review, 70, 312-326.

[2] Muirhead, R. J. (1982). Aspects of Multivariate Statistical Theory, Wiley: New York.

[3] Phillips, P. C. B. (1977). "Approximations to some finite sample distributions associated with a first order stochastic difference equation," Econometrica, 45, 463-485.

[4] _________ (1978). "Edgeworth and saddlepoint approximations in a first order non-circular autoregression," Biometrika, 65, 91-98.

[5] _________ (1982). "Best uniform and modified Pade approximation of probability densities in econometrics," in W. Hildenbrand (ed.), Advances in Econometrics, Cambridge University Press, 1982, 123-167.

[6] _________ (1983). "ERA's: A new approach to small sample theory," Econometrica, 51, 1505-1526.

[7] _________ (1984). "The exact distribution of LIML: I," International Economic Review (forthcoming).

[8] Rice, J. R. (1964). The Approximation of Functions, Vol. 1, Addison Wesley: Reading.

[9] Rothenberg, T. J. (1977). "Edgeworth expansions for multivariate test statistics," U.C. Berkeley Discussion Paper (unpublished).

[10] Theil, H. (1975). Theory and Measurement of Consumer Demand, North-Holland: Amsterdam.

Figure 1: Distribution of Serial Correlation Coefficient; $\alpha=0, T=5$

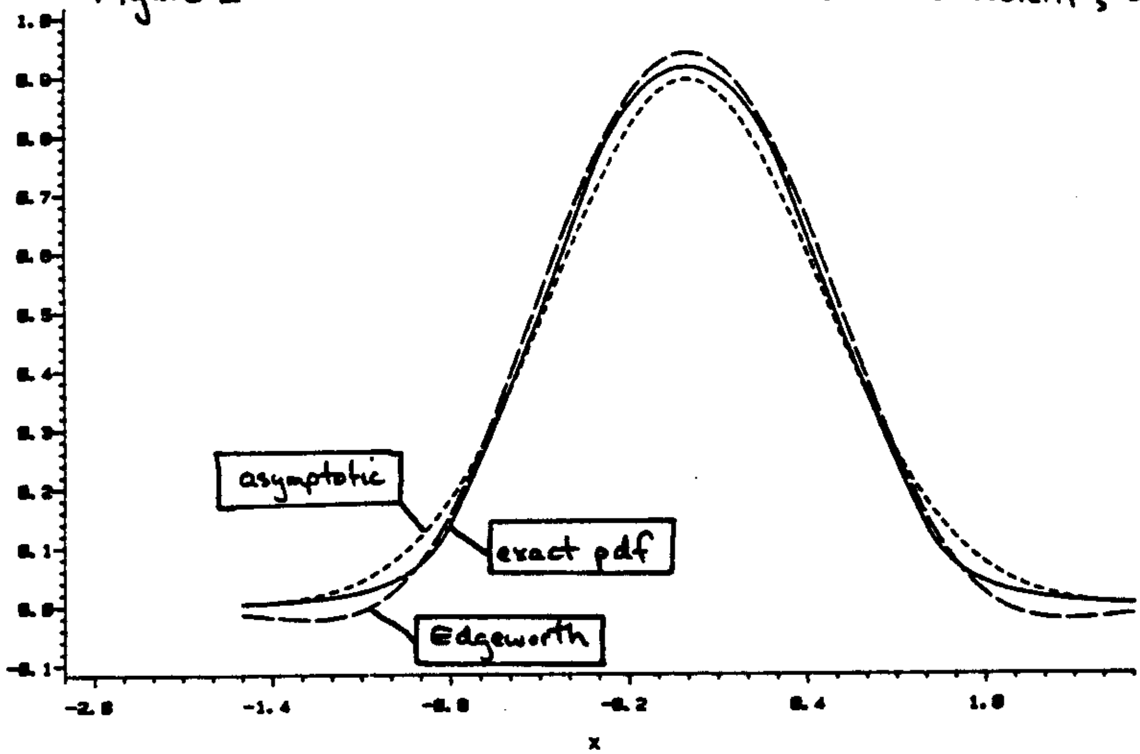

Figure 2: Distribution of Serial Correlation Coefficient; $\alpha=0, T=10$

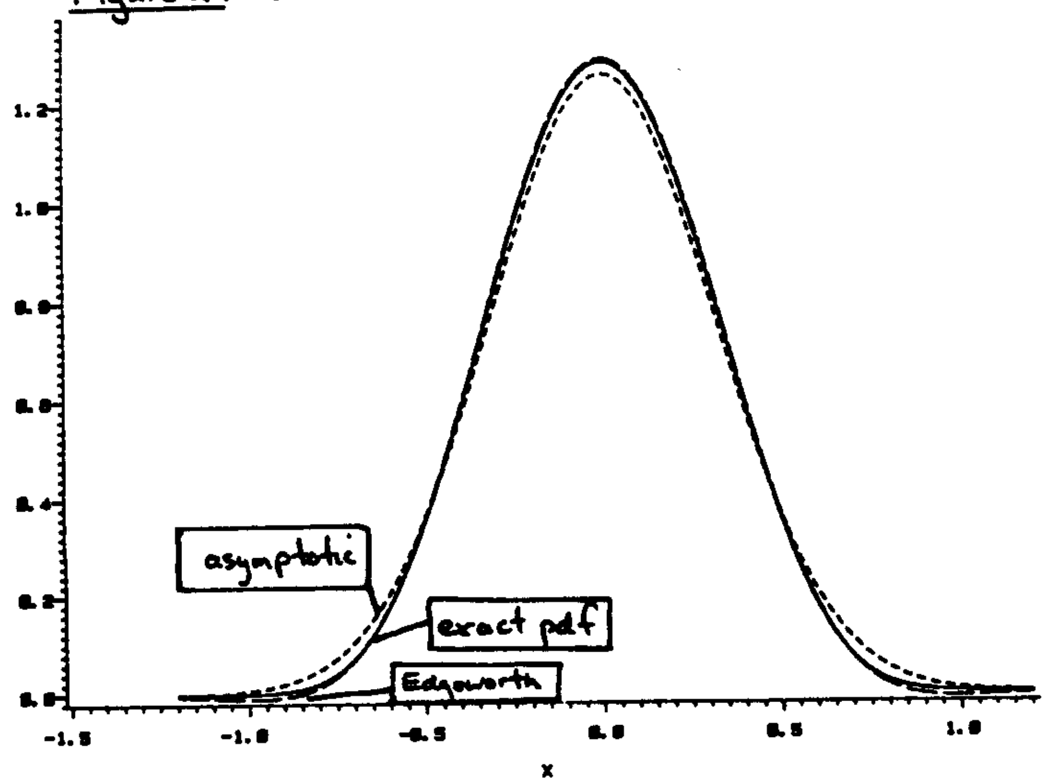

Figure 3: Distribution of Serial Correlation Coefficient; $\alpha=0$, $T=20$

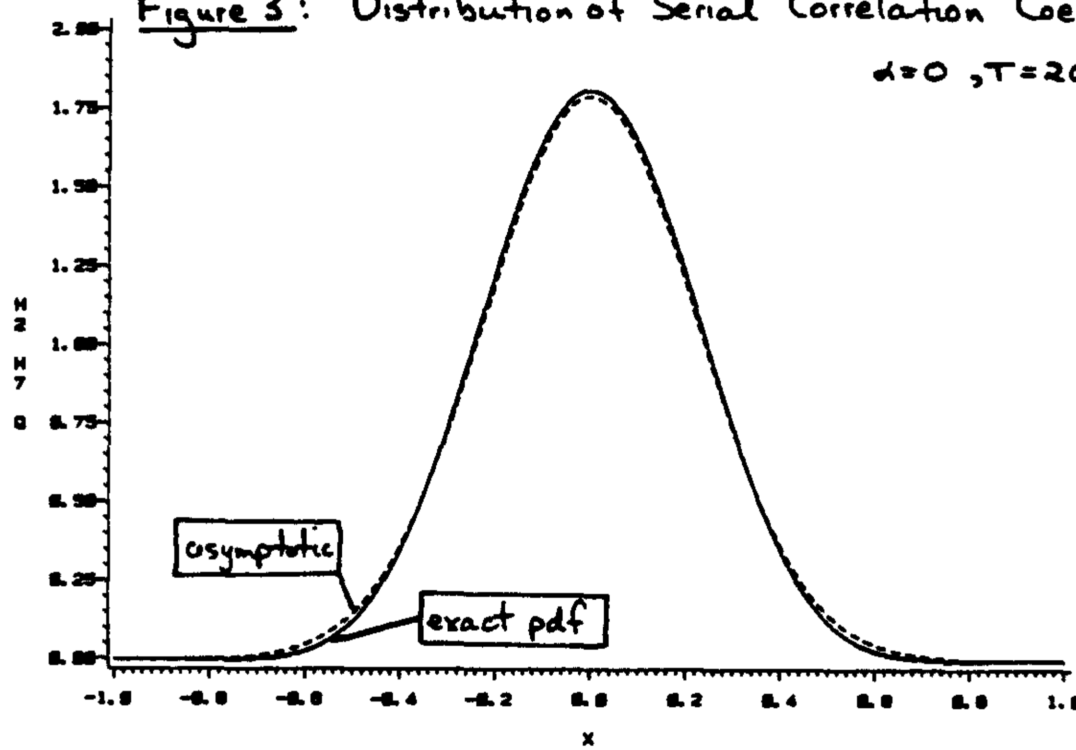

Figure 4: Distribution of Serial Correlation Coefficient; $\alpha=0.9$, $T=10$

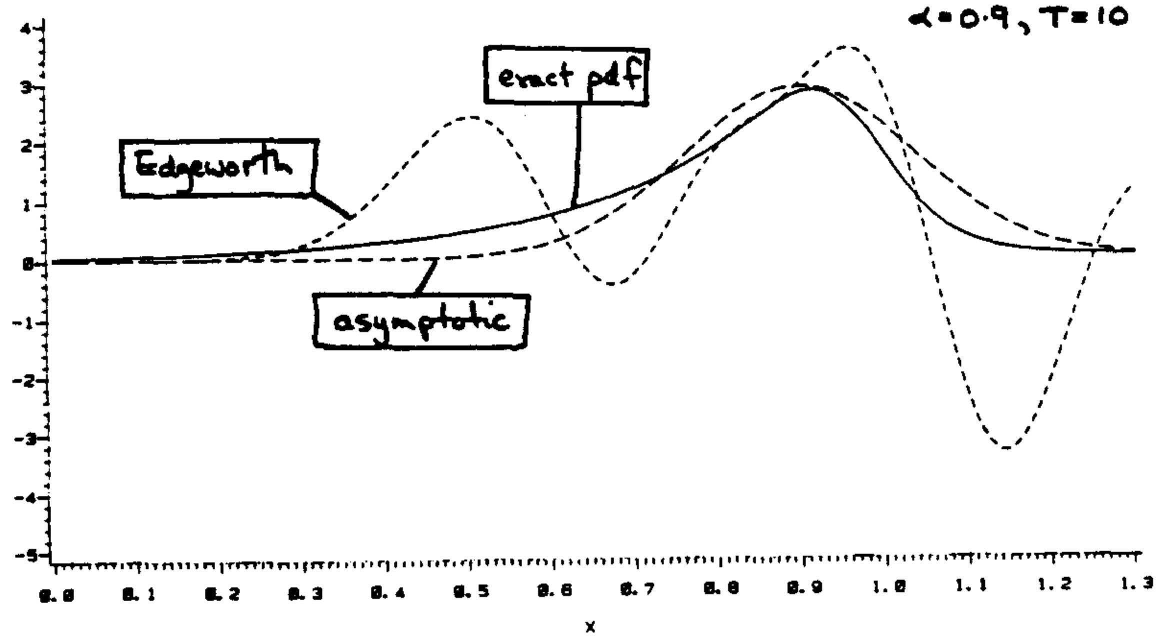

Figure 5: Distribution of Serial Correlation Coefficient; α=0.9, T=20

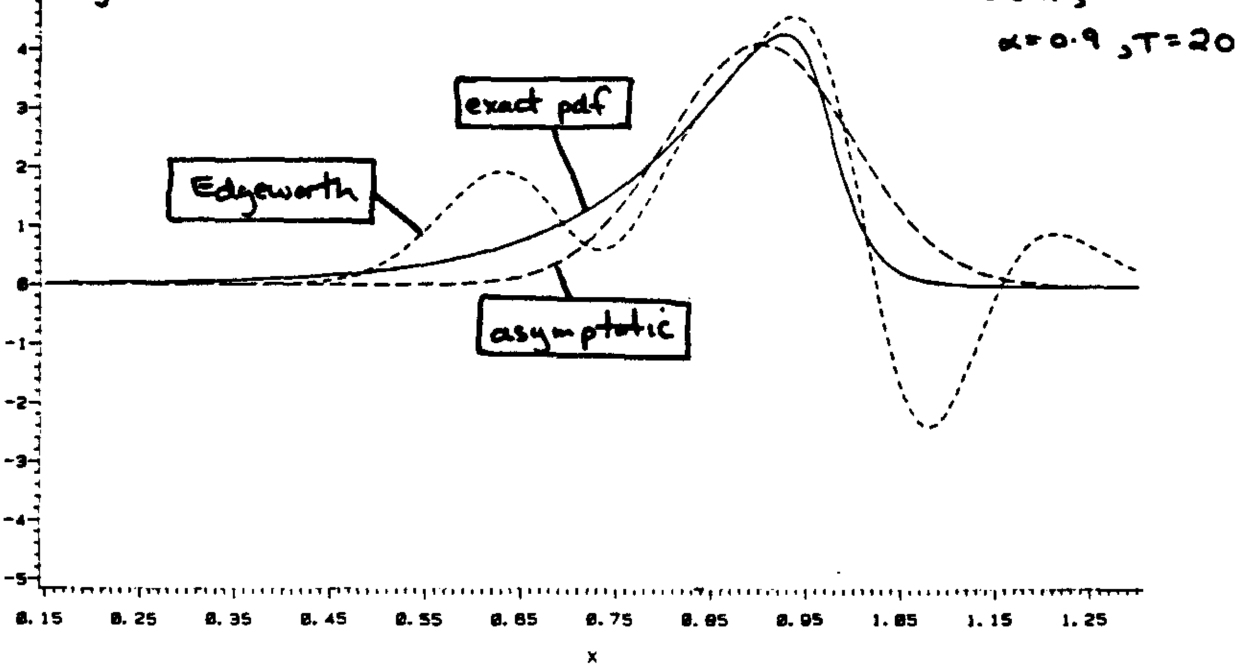

Figure 6: Distribution of Serial Correlation Coefficient; α=0.9, T=50