Cowles Foundation Paper #642

TESTING FOR SERIAL CORRELATION AND UNIT ROOTS USING A COMPUTER FUNCTION ROUTINE BASED ON EXTENDED RATIONAL APPROXIMANTS

P. C. B. Phillips and P. C. Reiss

ABSTRACT

This paper initiates a research program to provide computer function routines that can be used to deliver critical values or significance levels for statistical tests. These routines are easily integrated into existing econometric software and can be made available on a user call basis. The mathematical formulas underlying these approximants belong to the family of extended rational approximants (ERAs) introduced in [15]. The first part of this paper extends the algebraic theory of ERAs to distribution function approximation. Composite functional approximants are also

Advances in Statistical Analysis and Statistical Computing, Vol. 1, pgs. 1–50.

ISBN: 0-89232-467-8

developed to treat the parameter multidimensionality that is common in practical applications. The second part of the paper reports a detailed application of the approach to the distribution of the serial correlation coefficient under spherical Gaussian errors. The formulas we extract are error-corrected Edgeworth approximants that yield at least three decimal place accuracy over the entire distribution for all sample sizes ($T \geq 4$). These approximants can be used to mount a variety of tests, including tests for serial correlation and unit roots. Further extensions of this work to higher order serial correlation coefficients that are used in the Box–Jenkins model identification process are discussed in the conclusion.

1. INTRODUCTION

Recent advances in exact sampling and asymptotic expansion theories have produced general formulas for the probability density functions (p.d.f.'s) and cumulative distribution functions (c.d.f.'s) of many commonly employed econometric test statistics and estimators. These formulas are auspicious developments for applied researchers who wish to report the small sample properties of their statistics and estimators. Despite the promise of these theoretical developments for the conduct of applied research, however, there have not been many explicit attempts to use formulas from the finite sample literature in empirical work. There appear to be two major reasons why these formulas have not been used more frequently in applied research. First, the exact finite sample formulas often involve very complicated mathematical expressions.[1] Second, the reliability and accuracy of numerical results from asymptotic expansion formulas (and even exact p.d.f. and c.d.f. computations) is open to question (see [13] and [16]).

Because of the difficulties involved with the computation of exact distributions and unease over the reliability of approximations derived from asymptotic expansions, many econometricians have resorted to alternative numerical algorithms (for example Imhof's [9] routine) in order to calculate density and distribution functions. These algorithms are generally regarded as highly accurate. Unfortunately, the complexity and expense of these routines has for the most part limited their integration into econometric packages. Moreover, another level of computational complexity enters into the use of these routines for testing in that they require additional solution algorithms in order to deliver the appropriate critical values for Neyman–Pearson tests.

In the absence of simple, cheap, and accessible means for computing exact p.d.f. and c.d.f. values, econometricians and statisticians have resorted to longhand ways of summarizing distribution functions. The conventional means for presenting this information is a series of tables that report grids of probability values indexed against a set of parameters that control the shape of the distribution function (for example, sample size, degrees of freedom, and the concentration parameters). In situations where the distribution function depends upon only one or two of these shape parameters, these tables are easily developed. The t, F, and χ^2 tables in the back of econometric texts are familiar examples. In situations where there are several parameters that determine the shape and location of the distribution (for example, the Durbin–Watson statistic) or where some of the parameters vary continuously, clever ways of summarizing tabular information have been devised. Inevitably, these alternative methods sacrifice precision in the grid of reported values because of space limitations. The resulting sacrifice of information can be substantial, particularly if certain parameters take on a continuum of values and parts of the distribution (such as the tail areas) are sensitive to small changes in these parameters.

There is some hope that with the continuing technological advances in mainframe and personal computing and the increased availability of sophisticated numerical software, applied researchers will be able to overcome many of these computational problems and space constraints. Currently, we have the means with which we can carry out moderately complicated c.d.f. and p.d.f. calculations as well as the ability to store large volumes of statistical tables in computer banks. As the advance in computing technology continues to expand our capacities to store, calculate, and retrieve information on p.d.f.'s and c.d.f.'s, a question arises as to which means will be the most efficient for introducing finite sample results into applied work. The suggestion that we should begin incorporating numerical routines (such as Imhof's routine) directly into econometric packages has been advocated by some researchers (for example, Sargan and Bhargava [22]). This approach has distinct advantages over other options such as storing statistical tables on disk. On the other hand, the integration of numerical routines directly into econometric packages also has its drawbacks, as indicated earlier. Not the least of these is the substantial start-up

cost entailed with programming and storing the routines in one easy-to-use package. In addition to these start-up costs, there can be non-trivial variable costs associated with each use of these routines.

In this paper we develop and illustrate an alternative approach to computing probability values for c.d.f.'s. This approach is based upon Phillips's ([15], [18] and [19]) earlier work on functional approximation of p.d.f.'s. Phillips's functional approximation approach seeks to strike a middle ground between direct computation methods and the tabular form for p.d.f.'s and c.d.f.'s. Its main attractions are that it requires very little storage space, it is cheap to use, it is flexible enough to consider a variety of mathematical and Monte Carlo information on the c.d.f., and it is easy to program. The formulas delivered by this approach provide high accuracy and can be carried in a function routine designed to yield outputs corresponding to a variety of possible user choices. Thus the user may call for the probability level that is associated with a calculated statistic, the critical region for a given test size, or even power function evaluations for well-specified alternatives.

The present paper draws its motivation from this general research strategy. Its immediate purpose is quite specific: to illustrate the usefulness of these approximation techniques by providing programmable formulas that can be used to mount exact tests of serial correlation or unit roots for *any* sample size. The formula we extract belongs to a family of extended rational approximants (called ERAs). This family is developed for general purpose c.d.f. approximation and is related to the family of p.d.f. approximants introduced in [15]. This formula is ready to be hard-wired into existing econometric software packages. Its implementation requires only a few lines of computer code and its CPU (central processing unit) requirements are negligible. It thus liberates the applied investigator from the detailed tables or numerical calculations that must now be used in order to perform these tests.

There are of course practical questions and drawbacks raised by this alternative. These are addressed and extensively discussed in Sections 2 and 3, and the Appendix of this paper. Among the issues we will consider are: What form should a numerical approximant take? Is it possible to produce highly accurate significance level values over wide families of c.d.f.'s? And, are there systematic methods available for straightforwardly computing these approximants?

The plan for the remainder of this paper is as follows. Section 2 develops the algebra and theory behind our ERAs. Its main purpose is to extend the theory of ERAs to cumulative distribution function approximants and to composite function approximants. These latter approximants are designed to treat the multi-dimensionality of the shape parameters that index c.d.f.'s. A detailed application of these ERAs to the distribution of serial correlation coefficients is given in Section 3. Our final formulas are reported in Section 3.4. The Appendix to the paper deals with computational issues.

2. CDF AND COMPOSITE FUNCTION APPROXIMATION

2.1. Foregoing Research

The use of hard-wired computer function routines to efficiently compute statistical significance levels is not new. Statisticians (such as Fisher) have for years sought simple approximations to distributions that cannot be expressed in closed form. Hastings [7] was one of the first in a series of investigators who systematically devised polynomial and rational function approximants to some of the more commonly used distributions (such as the standard normal distribution). These approximations are continually being refined and improved. Many are now so simple that they can be programmed on hand calculators. Zelen and Severo [26] provide collections of some of the most commonly employed approximants. Indeed, a number of these are currently hard-wired into econometric packages enabling the user to automatically check for the significance of F and chi-squared statistics. There are two main limitations to these simple distribution function approximants. First, they are confined to relatively simple distribution functions that are indexed by one or two parameters. Second, they are constructed in a largely ad hoc fashion and there is no guarantee that they will have a uniform degree of accuracy over their entire range.

These limitations have been overcome in a series of recent papers by Phillips, who introduced a family of extended rational approximants and provided a theoretical framework to justify their use in distribution theory. In particular, [15] introduced mechanisms for approximating p.d.f.'s within $C_0^+[-\infty, +\infty]$, the class of

continuous, positive-valued functions that vanish at $\pm\infty$. That article suggested the following family of (nonreducible) ERAs of maximal degrees m and n for p.d.f. approximation:

$$era(x;\gamma) = s(x)[m,n](x;\gamma) = s(x)\frac{\gamma_0 + \gamma_1 x + \cdots + \gamma_m x^m}{1 + \gamma_{m+1}x + \cdots + \gamma_{m+n}x^n}, \tag{1}$$

where γ is a vector of rational coefficients (the γ_i), and $s(x)$ is a coefficient function chosen to embody relevant analytic (or experimental) information about the true p.d.f. This family of ERAs works well for density approximation when $s \in C_0^+[-\infty, +\infty]$ and no poles occur in the denominator of (1). The main role of the rational coefficients γ in $[m, n]$ is to build on the strengths of $s(x)$ as a primitive approximant. As an arbitrary real function, $s(x)$ is flexible enough to permit the direct use of leading terms from small sample theory, numerical information on the p.d.f., and asymptotic expansion information such as that provided by an Edgeworth expansion.

Phillips [15] studied the properties of these ERAs in the uniform error norm and established existence, uniqueness, denseness, and characterization theorems for a best rational approximant. The critical result of that work is the proof of the following alternation (oscillation) property of the best approximant in normal (that is, nondegenerate) situations:

THEOREM 1 (ERROR ALTERNATION OF THE BEST APPROXIMANT). *If the number of alternations of the error curve* $e(x) = pdf(x) - era(x)$ *(that is, 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* (1).

The importance of this theorem is that it extends the classical Tchebycheff approximation theorem to density approximation over the entire real line. In so doing, it provides a simple means for identifying the best approximant in applications.

Practical procedures for implementing ERAs in the family (1) have been considered in both [15] and [18]. The practical techniques used to construct the ERAs in those papers rely largely on

modifications of the multiple point Padé approximation that remove unwanted poles (and zeros) in conventional Padé approximants. These methods produced impressive precision in the several examples that were attempted. In particular, they demonstrated how ERAs can yield substantial improvements over alternative procedures such as Edgeworth approximations even when low-order rational functions are used (usually [3/3]'s and [4/4]'s).

2.2. C.D.F. Approximation

We now introduce a new family of approximants that facilitate the direct approximation of cumulative distribution functions and tail probabilities. This family has the form:

$$\begin{aligned} Era(x;\gamma) &= S(x)[n/n](x;\gamma) \\ &= S(x)\frac{\gamma_0 + \gamma_1 x + \cdots + \gamma_n x^n}{\gamma_{2n+1} + \gamma_{n+1}x + \cdots + \gamma_{2n}x^n}, \end{aligned} \tag{2}$$

where n is even, $S(x)$ is a primitive c.d.f. (perhaps of the form $\int_{-\infty}^{x} s(t)\, d(t)$) with the properties that $S \in C$, $S \uparrow 1$ as $x \uparrow \infty$, $S \downarrow 0$ as $x \downarrow -\infty$, and $S(x) > 0$ for all x. In the proofs and numerical work that follows, two different normalizations will be used. The first is

$$\sum_{j=n+1}^{2n+1} \gamma_j^2 = 1, \tag{3}$$

which is useful in the theoretical development; the second is

$$\gamma_{2n+1} = 1, \tag{4}$$

which is more convenient in practical work. The parameter space of these rational coefficients, based upon the first normalization and pole elimination condition, is

$$\Gamma = \left\{ \gamma : \sum_{j=n+1}^{2n+1} \gamma_j^2 = 1 \text{ and the denominator of (2)} > 0 \text{ for all } x \in (-\infty, \infty) \right\} \tag{5}$$

and is a subset of $\mathcal{R}^{2n+1}$. If we wanted to insist on the correct asymptotic behavior in (2) as $x \uparrow \infty$, we could impose an additional restriction in (5), namely, that $\gamma_n = \gamma_{2n}$.

We now proceed to develop some useful notation and definitions. We denote by ${}_0C_1^+[-\infty, \infty]$ the class of continuous functions F over $(-\infty, \infty)$ for which $F \uparrow 1$ as $x \uparrow \infty$, $F \downarrow 0$ as $x \downarrow -\infty$, and $F(x) \geq 0$ for all x. We denote the uniform error norm by $\|e(x)\| = sup_x|e(x)|$. Following the developments in [17], we may now develop a theory of best approximation to c.d.f.'s in ${}_0C_1^+[-\infty, \infty]$. The following results form the basis for this theory.

THEOREM 2 (EXISTENCE AND UNIQUENESS). *If* $cdf(x) \in {}_0C_1^+[-\infty, \infty]$ *then there exists a unique best approximant to* $cdf(x)$ *in the class of ERAs defined by* (2) *and* (5).

PROOF. The proof follows the existence proof given in [18]. Uniqueness follows from the arguments in [1], pp. 56-57. ∎

THEOREM 3 (DENSENESS). *Suppose* $cdf(x)$, $S(x) \in {}_0C_1^+[-\infty, \infty]$, $S(x) > 0$ *for all* $x \in (-\infty, \infty)$, *and let* $\varepsilon > 0$ *be given. Then there exists an* ERA *in the class defined by* (2) *and* (5) *for which* $\|cdf(x) - Era(x)\| < \varepsilon$.

PROOF. We start by considering the function defined by the difference between the c.d.f. and the primitive, $cdf(x) - S(x)$. This function is contained in $C_0[-\infty, \infty]$, the space of continuous real functions defined over $(-\infty, \infty)$ that vanish at plus and minus infinity. The following three functions are also contained in this space:

$$S(x)[1 + x^4]^{-1}, S(x)\frac{[1 + x^2]}{[1 + x^4]}, \quad \text{and} \quad S(x)\frac{[1 + (x-1)^2]}{[1 + x^4]}. \tag{6}$$

Let $\mathcal{B}$ be the set of all functions generated from these primitive members by pointwise addition, pointwise multiplication, and multiplication by real numbers. Note that $\mathcal{B}$ is an algebra of real-valued functions that vanish at infinity. It is therefore a sub-algebra of $C_0[-\infty, \infty]$. In fact, $\mathcal{B}$ is dense in $C_0[-\infty, \infty]$. To prove denseness we note that if $\{x_1, x_2\} \in (-\infty, \infty)$ with $x_1 \neq x_2$, then either $S(x_1)[1 + x_1^4] \neq S(x_2)[1 + x_2^4]$ or $S(x_1)[1 + x_1^4]^{-1} = S(x_2)[1 + x_2^4]^{-1}$. In the latter case we deduce that either

$$S(x_1)\frac{[1 + x_1^2]}{[1 + x_1^4]} \neq S(x_2)\frac{[1 + x_2^2]}{[1 + x_2^4]}$$

or

$$S(x_1)\frac{[1+(x_1-1)^2]}{[1+x_1^4]} \neq S(x_2)\frac{[1+(x_2-1)^2]}{[1+x_2^4]}.$$

Thus, $\mathcal{B}$ separates points of $(-\infty, \infty)$. Moreover, because $S(x)[1+x^4]^{-1} > 0$ for all $x \in (-\infty, \infty)$, it follows that at each point of $(-\infty, \infty)$ there is a function in $\mathcal{B}$ which does not vanish. Thus, using a generalization of the Stone-Weierstrass theorem ([23], pp. 166-167), $\mathcal{B}$ is dense in $C_0[-\infty, \infty]$. We also note that $\mathcal{B}$ is an algebra with respect to pointwise S-multiplication, where S-multiplication is defined by the operation: $b_1 b_2(x) = S(x)\bar{b}_1(x)\bar{b}_2(x)$, for $b_1 = S(x)\bar{b}_1(x) \in \mathcal{B}$ and for $b_2 = S(x)\bar{b}_2(x) \in \mathcal{B}$.

Now let $\mathcal{A}$ be the set of functions generated from the primitive members (6) by pointwise addition, multiplication by real numbers and pointwise S-multiplication. Here $\mathcal{A}$ is an algebra of real-valued functions on $(-\infty, \infty)$ that vanishes at infinity; $\mathcal{A}$ is also a subalgebra of $\mathcal{B}$. But, because $\mathcal{A}$ separates points of $(-\infty, \infty)$ and contains a function that does not vanish there, $\mathcal{A}$ is dense in $\mathcal{B}$. It therefore follows that $\mathcal{A}$ is also dense in $C_0[-\infty, \infty]$.

To prove the theorem we note that given $\varepsilon > 0$, there exists an $a \in \mathcal{A}$ for which $\|cdf(x) - S(x) - a(x)\| < \varepsilon$. Because $S(x) - a(x)$ is an extended rational function of the form (2) the theorem now follows. ∎

THEOREM 4 (ERROR ALTERNATION). *Suppose* $cdf(x)$, $S(x) \in {}_0C_1^+[-\infty, \infty]$, *with* $S(x) > 0$ *for all* $x \in (-\infty, \infty)$. *Let*

$$Era(x; \gamma) = S(x)\frac{\gamma_0 + \gamma_1 x + \cdots + \gamma_{n-\mu}x^{n-\mu}}{1 + \gamma_{n+1}x + \cdots + \gamma_{2n-\nu}x^{n-\nu}} \tag{7}$$

be the best approximant in the family of $[n/n]$ *defined by* (2). *Then* (7) *is completely characterized by the property that the number of alternations of the error curve*

$$e(x) = cdf(x) - Era(x) \tag{8}$$

is at least $N = 2n - \nu + 2$.

PROOF. By the number of alternations of $e(x)$ we mean the number of consecutive points of the interval $(-\infty, \infty)$ at which $e(x)$ takes on its maximum absolute value with alternate changes in sign. Because $S(x) \uparrow 1$ as x tends to infinity, we note that any

degeneracy in the rational fraction $[n/n](x)$ must follow the pattern $\mu \geq \nu$ if $e(x)$ is to be bounded at infinity. Hence, when degeneracy occurs in the best approximant the numerator will be at least as degenerate (in degree) as the denominator. Under these conditions on degeneracy, the classical proof of error alternation applies as in [1]; and a necessary and sufficient condition for best approximation is that the number of alternates of the error curve be at least $N = 2n - \nu + 2$ as stated in the theorem. ∎

2.3. *Composite Function Approximation*

The functional approximation theory developed above is essentially univariate in character. Most practical problems, on the other hand, are inherently multidimensional, either because we are dealing directly with multivariate distributions or because we wish to allow for parameter variation in a univariate distribution. The second source of multidimensionality will be our main source of concern, although the methods discussed below do apply in the former context. Since most influential procedures rely on univariate statistics whose distributions are parameter dependent, this choice of emphasis is thought to be more appropriate for the practical application of our methods.

We begin by considering a general approximation problem in which a given c.d.f. in ${}_0C_1^+[-\infty, \infty]$ is parameterized by a scalar T, which we will regard as the sample size in what follows. We denote the parameterized distribution by $cdf_T(x)$ and define T over the integer point set $\mathcal{T}_0 = \{T | T \geq T_0\}$ for some (usually small) value of T_0. Approximation of $cdf_T(x)$ is now a bivariate problem over the product space $\mathcal{R} \times \mathcal{T}_0$ and the family (2) might be extended to allow for bivariate rational functions in $\langle x, T \rangle$. But this form of extension of the theory is not the most promising. In fact, best uniform norm approximants to multivariate continuous functions do not always exist and when they do they are usually not unique. Moreover, the characterization theorems of univariate problems (such as Theorem 4 above) do not generalize to multivariate situations. The main difficulty is the lack of Tchebycheff sets of multivariate functions. These issues are discussed in detail by Rice [20], chap. 12.

Fortunately, there is an alternative and more natural approach to handling the multidimensionality of the approximation problem.

This is to use the concept of a *composite functional approximant.* The idea of a composite functional approximant can best be explained in the context of the approximation of $cdf_T(x)$. First, let $S_T(x)$ be a primitive c.d.f. that is indexed by T and suitable for use in Eq. (2). Next, let

$$Era(x;\ \gamma(T)) = S_T(x)[n/n](x;\ \gamma(T)) \tag{9}$$

be the best uniform approximant to $cdf_T(x)$ for a given T, where $\gamma(T)$ denotes the relevant vector of rational coefficients that vary over the integer point set $\mathcal{T}_0$. We call Eq. (9) the *first-stage* ERA for $cdf_T(x)$.

We now construct a composite rational approximant to $cdf_T(x)$ by replacing $\gamma(T)$ in Eq. (9) with a suitable vector of coefficient function approximants defined in terms of T. In our application we choose to approximate $\gamma(T)$ over a finite point set such as $\mathcal{T}_\gamma = \{T \in Z \mid T_0 \leq T \leq T^0\}$. Note that $\mathcal{T}_\gamma$ has an upper bound. Such a bound is appropriate when it is known that the primitive approximant is adequate for $T > T^0$. We will treat this case explicitly in the following section under asymptotic refinements.

Let γ_T be a vector of suitable approximants to the coefficient function $\gamma(T)$ over the point set $\mathcal{T}_\gamma$. Then

$$Era_T(x) = Era(x;\ \gamma_T) \tag{10}$$

will be called the *second-stage* ERA for $cdf_T(x)$. Determination of γ_T will involve a choice of coefficient function form and numerical computation based upon the first-stage results. The most suitable choice of form of γ_T will be problem specific. Several possibilities, all of them simple to employ, will be considered below. Our application in Section 3 will illustrate the operational features of these different choices.

The simplest way to reproduce first-stage ERA accuracy in Eq. (10) is to use spline functions for the elements of γ_T, each constructed with a sufficient number of knots to achieve the desired accuracy. The process can be demonstrated by taking an arbitrary element, say $\gamma_i(T)$, of the first-stage coefficient vector. Let us assume that the first-stage ERA, and hence $\gamma_i(T)$, are known on the mesh of integers $\mathcal{T}_\gamma = \{T_0 < T_1 < T_2 < \cdots < T_N = T^0\}$. Define ordinates $y_j = \gamma_i(T_j)(j = 0, \ldots, N)$ on this mesh and let $d_j = T_{j+1} - T_j$ denote the grid spacing. We may now solve the following system of equations that ensure continuity in the first

derivative of the spline and yield coefficients m_j $(j = 2, \ldots, N-1)$ that are needed in formula (13) below:

$$Am = z \tag{11}$$

where A is the symmetric tridiagonal matrix

$$\begin{bmatrix} 2(d_1 + d_2) & d_2 & & & \\ d_2 & 2(d_1 + d_2) & d_3 & & \\ & \cdots & \cdots & \cdots & \\ & & & d_{N-2} & 2(d_{N-2} + d_{N-1}) \end{bmatrix}$$

and

$$m' = (m_1, m_2, \ldots, m_{N-1})$$
$$z' = (z_1, z_2, \ldots, z_{N-1}),$$

where

$$z_j = 6\frac{y_{j+1} - y_j}{d_j} - 6\frac{y_j - y_{j-1}}{d_{j-1}}. \tag{12}$$

We set $m_0 = m_N = 0$ and then the natural cubic spline function that interpolates the values $\{y_j\}$ at the points $\{T_j\}(j = 0, \ldots, N)$ is given by

$$s\ell_i(T) = \left\{(T - T_j)y_{j+1} + \frac{(T_{j+1} - T)y_j}{d_j} - (T - T_j)(T_{j+1} - T) \cdot \frac{(d_j + T_{j+1} - T)m_j + (d_j + T - T_j)m_{j+1}}{6d_j}\right\} \tag{13}$$

over the interval $[T_j, T_{j+1}]$. We use the index i in Eq. (13) to indicate that $s\ell_i(T)$ is the spline approximant to $\gamma_i(T)$. Replacing γ_T by $s\ell(T)$ in Eq. (10), we now have a second-stage ERA that reproduces first-stage accuracy.

The representation (13) for the cubic spline is the one most suited for computational purposes [6]. In cases where there are many knots it is worth using an efficient algorithm (relying on the tridiagonal nature of the matrix A) to solve (11) (see [2], pp. 14–15). The choice of the natural spline function (13) with $m_0 = m_N = 0$ allows for linear extensions of (13) to the regions $\{T < T_0, T > T_N\}$ outside the mesh. This choice may be conveniently changed to accommodate other end point information by appropriate modification of the equation system (11).

Choice of the mesh, $\{T_j\}$, will also play an important role in the performance of the spline function approximant. Intuition and experience from the practical application we will report later suggest that use of a fine equispaced mesh in regions where the coefficient functions $\gamma(T)$ display most variation (typically small T) will perform best. This may be combined with the use of a coarse mesh in those parameter regions where the behavior of $\gamma(T)$ shows little variation. When $\gamma(T)$ is defined so as to ensure compatibility with asymptotic theory, our experience suggests that $\gamma(T)$ displays a slow monotonic approach to its natural asymptote as T becomes large. In this region a coarse grid yields an economical and effective spline which can be readily spliced into the asymptote for suitably large values of γ.

An alternative method of generating coefficient function approximants γ_T for (10) is to use polynomial or rational functions. More specifically, we introduce the general family

$$[r/s]_i(\eta) = \frac{\alpha_{i0} + \alpha_{i1}\eta + \cdots + \alpha_{ir}\eta^r}{1 + \beta_{i1}\eta + \cdots + \beta_{is}\eta^s} \tag{14}$$

of rational coefficient function approximants for $\gamma_i(T)$. In Eq. (14), $\eta = T^\nu$ is a suitably chosen power of T. Note that when $s = 0$, Eq. (14) reduces to a simple polynomial approximant. Typically ν can be chosen to accommodate the asymptotic behavior of the approximant and the spacing of the coefficients. We have found $\eta = T^{-1/2}$ to be a good choice in our applications.

As before, let the integer point set $\mathcal{T}$ be the domain of approximation. On such a finite point set we may consider the possibility of extracting a best approximant to $\gamma_i(T)$ in the family defined by Eq. (14). When we restrict our attention to these polynomial approximants, the Tchebycheff theory of best uniform approximation applies directly (see [20]). Thus, given $\gamma_i(T)$ defined on $\mathcal{T}$, there exists a unique best uniform approximant in the family (14) with $s = 0$. Let $[r/0]_{it}$ denote this approximant. Then the usual characterization theory for polynomial approximants also applies. Specifically, the error function $\gamma_i(T) - [r/0]_{iT}$ will possess an alternating set of $r + 2$ points of $\mathcal{T}$ (see [20], p. 33 for the details). We now use the notation $[r/0]_T$ to denote the vector of best uniform approximants to the elements of $\gamma(T)$ over $\mathcal{T}$. Then the second-stage ERA given earlier by Eq. (10) has the explicit form:

$$Era_T(x) = Era(x; [r/0]_T) \tag{15}$$

and is called the best composite (or best product) approximation to $cdf_T(x)$ over $x \in (-\infty, \infty)$ and $T \in \mathcal{T}$. Best product Tchebycheff approximations were first introduced by Weinstein [24] in the context of polynomial families. Extensions to rational functions were made in Henry and Weinstein [8] and the concept of best composite approximation was introduced in [4] as a principle of unification.

A theoretical difficulty prevents the automatic extension of best composite approximation of the form (15) to general rational functions such as (14). The difficulty arises because of the absence of an existence theory for a best approximant to $\gamma_i(T)$ over $\mathcal{T}$ within the family (14). In fact, best uniform rational approximants on finite point sets do not necessarily exist and the alternation theory for rational approximants also fails. Some examples which illustrate these failures are given in [20], pp. 130–131. Fortunately, although the theoretical development is impaired by these difficulties, rational functions still provide excellent approximations on finite point sets. As with interval approximation, these often provide substantial improvements over polynomial approximants of the same degree. In this case some of the improvement is purchased by the presence of discontinuities in the regions which bridge the discrete points of the domain of approximation. In view of these problems, the search for an adequate approximant in the rational family (14) is less formalized than in the case of polynomials. Our approach, therefore, is to select an approximant because of its performance rather than because of characteristics which distinguish it as best in the Tchebycheff sense. Let $[r/s]_{iT}$ be such an approximant to $\gamma_i(T)$ which is well defined for all $T \in \mathcal{T}$. Then the second-stage ERA

$$Era_T(x) = Era(x; [r/s]_T) \tag{16}$$

is a composite rational approximant to $cdf_T(x)$.

2.4. *Preservation of Asymptotic Properties*

Composite functional approximants such as (10), (15), or (16) may be formulated in such a way as to preserve the validity of asymptotic approximations, including second, third, or higher-order approximations. The simplest procedure involves the refinement of Edgeworth asymptotic series to improve their perfor-

mance in regions of the distribution where they possess inherent weaknesses ([19] provides some recent illustrations). To fix ideas, suppose we have 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 then write

$$P(T^{1/2}(\alpha_T - \alpha) \leq z) = \Phi\left(\frac{z}{\sigma} + T^{-1/2}p_1\left(\frac{z}{\sigma}\right) + T^{-1}p_2\left(\frac{z}{\sigma}\right)\right) + o(T^{-1}), \tag{17}$$

where σ represents the asymptotic standard error, Φ is the c.d.f. of a $N(0, 1)$ random variable, and p_1 and p_2 denote the appropriate Edgeworth polynomials. Then

$$Ed_T(x) = \Phi\left(T^{1/2}\frac{(x-\alpha)}{\sigma} + T^{-1/2}p_1\left(T^{1/2}\frac{(x-\alpha)}{\sigma}\right) + T^{-1}p_2\left(T^{1/2}\frac{(x-\alpha)}{\sigma}\right)\right) \tag{18}$$

is the associated approximation to the distribution of α_T.

Now let $S_T(x) = Ed_T(x)$ be the primitive approximant in Eq. (9). The resulting first-stage ERA for the distribution of α_T has the general form:

$$Era_T(x; \gamma(T)) = Ed_T(x)[n/n](x; \gamma(T)). \tag{19}$$

As usual, n is an even integer (say $n = 2m$). We note that since expression (19) is a best approximant, $[n/n]$ is a bounded rational function over $\mathcal{R}$ and has no real poles. Hence, we may decompose $[n/n]$ into partial fractions as follows:

$$[n/n](x; \gamma(T)) = c_0(T) + \sum_{j=1}^{m}\left\{\frac{c_j(T)}{x - d_j(T)} + \frac{\bar{c}_j(T)}{x - \bar{d}_j(T)}\right\} \tag{20}$$

where c_0 is a real scalar, $c_j = c_{j1} + ic_{j2}$, $d_j = d_{j1} + id_{j2}$, and the bars over letters are used to denote complex conjugates.

The decomposition of (20) implies a new parameterization of the approximant in terms of c_0, c_{jk}, and d_{jk} ($j = 1, \ldots, m$; $k = 1, 2$). This reparameterization is useful in the development of composite approximants that preserve asymptotic behavior. We will therefore have occasion in what follows to replace the coefficient vector $\gamma(T)$ with the alternative coefficients $\{c(T), d(T)\}$. Coefficient

function approximants for the components of $c(T)$ and $d(T)$ may now be developed in any of the families considered in the previous section.

Let us suppose that $Ed_T(x)$ delivers sufficient accuracy as an approximant of $cdf_T(x)$ for $T > T^0$ and that the first-stage ERA (19) applies over $x \in \mathcal{R}$ and $\mathcal{T} = \{T_0 \leq T \leq T^0\}$. Spline function coefficient approximants would now take the form

$$c_{0T} = s\ell_0(T), \qquad c_{jkT} = s\ell_{jk}(T), \qquad d_{jkT} = s\ell_{m+j,k}(T) \quad (21)$$

(where $j = 1, \ldots, m$; $k = 1, 2$), over $T \in \mathcal{T}$ and would be hooked into the following set of linear extensions:

$$c_{0T} = 1, \qquad c_{jkT} = 0, \qquad d_{j1T} = 0, \qquad d_{j2T} = \varepsilon \quad (22)$$

(where $j = 1, \ldots, m$; $k = 1, 2$), over $T \in (T^0, \infty)$ and where ε is a small positive scalar. With this second stage, the composite approximant

$$Era_T(x; \gamma(T)) = Era(x; c_T, d_T) = Ed_T(x)[n/n](x; c_T, d_T) \quad (23)$$

preserves the asymptotic behavior of $Ed_T(x)$ as $T \uparrow \infty$ and modifies the primitive Edgeworth series (approximant) in the domain $\{T_0 \leq T \leq T^0\}$. Similarly behaved composite approximants may be constructed using polynomial or rational coefficient approximants.

2.5. *Higher-Order Composite Approximation*

The process of constructing composite functional approximants that was outlined in Section 2.3 may be extended to problems of higher dimension. Consider for example a distribution that is indexed by two variables, $cdf_T(x; \alpha)$, where α is a scalar parameter in $\mathcal{R}$. Given α and T, we let

$$Era(x; \gamma(\alpha, T)) = S_T(x; \alpha)[n/n](x; \gamma(\alpha, T)) \quad (24)$$

be the best uniform approximant to $cdf_T(x; \alpha)$. Equation (24) extends Eq. (9) by allowing the primitive approximant, $S_T(x; \alpha)$, to be parameter dependent (as, for instance, an Edgeworth series approximant would be) and by functionalizing the rational coefficient vector $\gamma = \gamma(\alpha, T)$ on both α and T.

We now develop a sequential composite approximant for γ with approximants at each level. Thus, when α and T are fixed we have $\gamma = \gamma(\alpha, T)$ as given in Eq. (24). Then, as we allow α to

vary, we have a conditional mapping from α to γ for a given T, which we may write as $\gamma = \gamma_T(\alpha)$. This mapping will be continuous at α_0 provided that $cdf_T(x; \alpha_0)$ is normal for $[n/n]$ approximation. By this we mean that the best approximant to $cdf_T(x; \alpha_0)$ is not of degenerate degree. The continuity of $\gamma_T(\alpha)$ at $\alpha = \alpha_0$ then follows from Theorem 1 of [8]. In fact, if the degree of $[n/n]$ $(x; \gamma(\alpha, T))$ is constant over the domain of α, then $\gamma_T(\alpha)$ is continuous throughout this domain (Theorem 2 of [8]). In view of the continuity of $\gamma_T(\alpha)$, we may now develop best uniform approximants to $\gamma_T(\alpha)$ in the following family of rational functions of α:

$$[p/q](\alpha) = \frac{g_0 + g_1\alpha + \cdots + g_p\alpha^p}{1 + g_{p+1}\alpha + \cdots + g_{p+q}\alpha^q}. \tag{25}$$

When the degree of $[p/q]$ is sufficiently large, the error on the approximant will be sufficiently small to ensure that the composite approximant $S_T(x; \alpha)[n/n](x; \gamma(\alpha, T))$ has a non-vanishing denominator. Each of the rational coefficients g_i in (25) is conditional on T, say $g_i = g_i(T)$. Now if T varied continuously and $\gamma_T(\alpha)$ was nondegenerate and of order $[p/q]$, we could extend the above argument to the next dimension. Because, however, T is defined on a finite point set in our problem, we may construct a suitable approximant to $g_i(T)$ using the spline function formulas given in Section 2.3. Call the resulting approximant g_{iT}. When a sufficiently fine mesh is chosen for this construction the composite coefficient function

$$[p/q]_T(\alpha) = \frac{g_{0T} + g_{1T}\alpha + \cdots + g_{pT}\alpha^p}{1 + g_{p+1,T}\alpha + \cdots + g_{p+q,T}\alpha^q} \tag{26}$$

will be continuous in α, say, for all $T \in \mathcal{T}$. The third-order composite functional approximant to $cdf_T(x; \alpha)$ is then given by

$$Era_T(x; \gamma(\alpha, T)) = S_T(x; \alpha)[n/n](x; \gamma(x; [p/q]_T(\alpha))). \tag{27}$$

Higher-order approximants of this type clearly involve a substantial increase in computational burden. Yet the extraction of third-order approximants such as (27) is well within the reach of present computational equipment; and, once found, these approximants can be employed in applications at a negligible computational cost.

3. DISTRIBUTION OF THE SERIAL CORRELATION COEFFICIENT

3.1. *Objectives and Potential Applications*

As in [17], we use the autoregressive model

$$y_t = \alpha y_{t-1} + u_t \qquad (t = \cdots, -1, 0, 1, \ldots) \tag{28}$$

in which the u_t are i.i.d. (identical and independently distributed) $N(0, \sigma^2)$. Our focus will be upon the c.d.f. of the noncircular serial correlation coefficient

$$\hat{\alpha} = \left(\sum_{t=1}^{T} y_{t-1}^2 \right)^{-1} \left(\sum_{t=1}^{T} y_t y_{t-1} \right) \tag{29}$$

obtained from the application of least squares to (28). Our objective will be to develop an approximant to the c.d.f. of $\hat{\alpha}$ which delivers sufficient accuracy for all conceivable sample sizes of relevance in empirical work (we have chosen $T \geq 4$), which is computationally inexpensive, and which is computer-ready for incorporation in regression software.

The approximant we report below in Section 3.4 meets these criteria. It belongs to the composite function family considered in Section 2.3 and yields an error-corrected Edgeworth approximation which preserves the latter's asymptotic behavior. Our approximant has been developed for the case of $\alpha = 0$ in (28), so that it has the following direct applications:

(i) Tests for the absence of correlation ($\alpha = 0$) in consecutive observations based on the statistic $\hat{\alpha}$ defined in (29).

(ii) Tests for a unit root ($\alpha = 1$) in the autoregression (28) based on the statistic $\hat{\alpha} = (\sum_{t=1}^{T} x_{t-1}^2)^{-1}(\sum_{t=1}^{T} x_t x_{t-1})$ with $x_t = y_t - y_{t-1}$; other specific hypotheses such as $\alpha = \alpha_0$ in (28) may be tested with the same statistic but with x_t defined as $y_t - \alpha_0 y_{t-1}$.

(iii) Tests for the absence of serial correlation in regression disturbances through the use of BLUS residuals.

(iv) Tests for the presence of a unit root in regression disturbances through the use of the BLUS vector for regression error first differences.

Power considerations suggest that some of the above tests [in particular the unit root tests (ii) and (iv)] will be dominated by other tests that are more powerful for certain alternative hypotheses and sample sizes. The tabulations of power recently reported by Sargan and Bhargava [22] indicate that the Berenblut-Webb statistic with exact critical values computed by numerical integration provides such a test (of a unit root). In other contexts these test rankings are less definitive. For example, in unit root tests against alternatives in the neighborhood of unity, say $0.9 \leq \alpha < 1$, the computations in [22] show very little difference in power among a variety of different tests. Power against such alternatives is also very low because the distributions of the associated serial correlation coefficients are so close in the neighborhood of unity. Moreover, Anderson [3] showed that (28) is a model for which uniformly most powerful (UMP) tests concerning α do not exist. He also showed that end point (at $t = 1$ and $t = T$) modifications to the density of data generated by (28) do allow such tests to be developed. Durbin and Watson [5] utilized these end point modifications in deriving their theory of the UMP invariant properties of the Durbin-Watson test. These modifications were also used by Sargan and Bhargava [22] in their development of UMP invariant tests of the random walk hypothesis for least squares regression errors. For moderate-to-large sample sizes these modifications will have negligible effects. But for small sample sizes (say $T \leq 10$) the effect of the modifications on power rankings may be more important [as, indeed, are circular modifications to the statistic (29)]. These effects have not as yet been properly explored. Finally, we may remark that the exact tests (iii) and (iv) above continue to apply in regression models where the regressors are stochastic and independent of the errors. In this respect they differ from and are more general than the conventional bounds and exact Durbin-Watson tests as well as the Berenblut-Webb test used in [22].

Notwithstanding the above remarks, functional approximants similar to the one we develop below for the distribution of $\hat{\alpha}$ in (29) may, of course, also be developed for other statistics of importance in models such as (28). Thus, our example may be taken to be illustrative of potential applications as well as operational with regard to tests such as (i)-(iv). We note also that the approximant may be further extended to allow for variable α

by using the method of Section 2.5. These extensions are not needed, however, for the tests discussed above.

3.2. *Formation of the First Stage ERAs*

Earlier experience from [18] suggested that an Edgeworth approximation to $O(T^{-1})$ might provide a suitable leading coefficient function in this problem. For the c.d.f. of $\hat{\alpha}$ the Edgeworth approximation to $O(T^{-1})$ is given by [14]:

$$Ed_T(x) = \Phi\left(T^{1/2}x + \frac{1}{4T^{1/2}}(x + Tx^3)\right). \tag{30}$$

This particular representation of the Edgeworth approximation is chosen because $Ed_T(x)$ satisfies the requirements stipulated for $S(x)$ in Eq. (2), namely, that $Ed_T(x) \in C, \uparrow 1$ as $x \uparrow \infty, \downarrow 0$ as $x \downarrow -\infty$, and > 0 for all $x \in (-\infty, \infty)$. Compared to similar applications where the Edgeworth can be a poor initial approximant (see [13]), in this particular situation the Edgeworth yields a good initial approximant for modest sample sizes (see Figures 2 and 4). For small sample sizes between 3 and 14, however, the accuracy of the Edgeworth approximation may not be suitable for empirical work. In this range the Edgeworth typically understates the true size of the test. For instance, when $\alpha = 0$, $T = 5$, and the true size is 0.05, the Edgeworth is off by more than 20%. Although this is not nearly as large a percentage error as that which results from using the asymptotic distribution, we shall show below that substantial improvements in accuracy and the uniformity of errors are possible if ERAs are used to refine the Edgeworth approximation.

In Figure 1 we have plotted the c.d.f. of $\hat{\alpha}$ for various sample sizes. This figure shows that the c.d.f. changes smoothly with T and that for large values of T there is a good deal of curvature that needs to be captured in the approximant. Also, we note that because the distribution of $\hat{\alpha}$ is symmetric, we need only consider constructing the ERA so that it approximates the c.d.f. over the positive real line.[2] Following the approach to composite function approximation outlined in Section 2.3, our first task is to construct a series of first stage ERAs that are indexed by the sample size T. As a practical matter, we must first decide on the mesh of integers, $\mathcal{T}_\gamma$, over which we will construct our first-stage rational

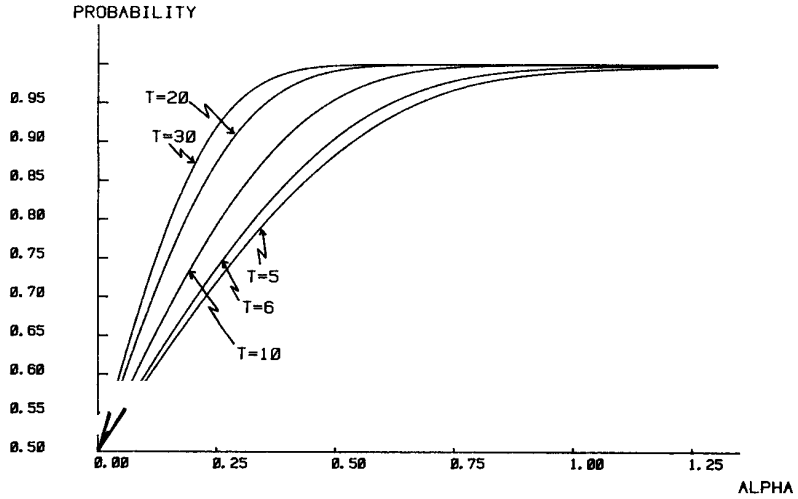

Figure 1. Serial Correlation Coefficient CDF.

approximants. Three practical considerations entered into our selection of $\mathcal{T}_\gamma$. First, it is expensive and computationally burdensome to compute c.d.f. values for all feasible sample sizes. On the other hand, because the second stage ERAs are being fit over a mesh, the grid must be fine enough to pick up any rapid changes that occur in the composite function. Second, after a sample size of 40, the Edgeworth yields an approximant that is accurate to four decimal places. Third, the Edgeworth approximant is very inaccurate for sample sizes less than four. These considerations led us to set $T_0 = 4$ and $T^0 = 36$. We then selected sixteen intermediate sample sizes for our mesh (T = 5, 6, 7, 8, 9, 10, 12, 14, 16, 18, 20, 22, 26, 28, 30, and 32). The mesh is finer for small sample sizes because the c.d.f. changes most rapidly for sample sizes between 5 and 10.

The next practical consideration in constructing the first-stage ERA is the question of how to choose the degree of the rational approximant. As a general rule, parsimony is important in specifying the first-stage coefficients because the number of second-stage computations rapidly increases with the number of first-stage coefficients that are estimated. We elected to fit a [2/2] rational function because of the smoothly changing nature of the c.d.f. Our experience suggests that in general the best way of proceeding is

to select the degree so as to minimize pole problems in the denominator (that is, choose n even and not too large) and yet allow sufficient flexibility in the denominator in order to trace out sharp changes of direction and preserve (near linear) asymptotic behavior as $x \uparrow \infty$ (see Figure 1).

The final practical issue in the development of the first stage ERAs centers on the estimation strategy for the rational function coefficients (the γ's). In Phillips [15], and [18], multiple-point Padé methods were used to estimate the coefficients. These approximants were constructed by selecting $2n + 1$ points of interpolation to estimate the $2n + 1$ rational coefficients. The main computational drawback in this approach to estimating the rational coefficients is that the points of interpolation must be chosen by trial and error (see [18]) so as to smooth out the error curve. In other words, the multiple-point Padé technique does not take into account how the error function is behaving at points intermediate to the points of interpolation. Although this may not be a serious problem in the first-stage approximation of functions that are smoothly changing over their domain, when we come to estimate second-stage ERAs based upon our first-stage values of the coefficient function (that is, $\gamma = \gamma(T)$), the selection of points of interpolation at the first stage turns out to have important consequences for the shape and regularity of the coefficient functions. In particular, the arbitrariness of the formal Padé procedure leads to coefficient functions that can have rapid changes. This led us to explore an alternative first-stage procedure that systematically produces a near-best approximant and yields coefficient functions that have slowly changing behavior. This latter property is what facilitates second-stage estimation.

In the spirit of trying to incorporate more information about the behavior of the error curve into the estimation of the rational coefficients, our procedure relaxes the Padé requirement that we only interpolate the approximant at $2n + 1$ points. Instead, we chose to fit the function over a much finer grid of points. This refinement obviously no longer guarantees that there is just one set of rational coefficients associated with the points of interpolation. We are therefore forced to specify a criterion for selecting rational coefficients that will produce an oscillating error curve similar to that possessed by a best approximant. Our preferred procedure works as follows. First, we compute the exact c.d.f.

values over a 0.01 grid between 0 and 1.5 using an extension of the Imhof density algorithm.[3] Next, we fit the rational coefficients using a number of alternative estimation techniques. Finally, we select a set of first-stage coefficient estimates based upon two criteria: (i) whether they had a smoothly, equioscillating error curve; and (ii) whether the coefficients changed smoothly enough so that a second-stage ERA could easily be fit to the first-stage coefficients.

The main computational differences among these alternatives are based upon the norm in which they minimize the error deviations of the rational approximant. The methods that we have extensively tested depend upon the following representations of the basic ERA:

$$\begin{aligned} cdf_T(x) &= Ed_T(x)[n/n](x) + e(x) \\ &= Ed_T(x)\frac{P(x)}{1+Q(x)} + e(x), \end{aligned} \tag{31}$$

$$y_T(x) = \frac{cdf_T(x)}{Ed_T(x)} = \frac{P(x)}{1+Q(x)} + \frac{1}{Ed_T(x)}e(x), \tag{32}$$

and

$$y_T(x) = -y_T(x)Q(x) + P(x) + \frac{1+Q(x)}{Ed_T(x)}e(x). \tag{33}$$

Given at least $2n+1$ observations on the c.d.f., the rational coefficients in (31) and (32) can be estimated by nonlinear least squares or generalized least squares, and the coefficients in (33) by ordinary least squares, generalized least squares, or least absolute error regression. As a practical matter, the main differences among these alternative techniques arise from the properties of their error terms. Alternatively, the differences can be viewed as arising from the different functions that the right-hand side must approximate. Our experience has been that no matter which equation is used as a basis for fitting the first stage ERAs, all of these techniques generate error curves, $e(x; \hat{y})$, that are smoothly oscillating about zero. Occasionally, poles will occur at points that may be within or outside the domain of approximation. For a pole within the domain, we suggest oversampling the grid near and at the pole so as to force the minimum error techniques to improve the ERA in the neighborhood of this point and eliminate the pole.

For poles that occur outside the domain of approximation, the domain of the approximation can be extended. In our application, these techniques enabled us to construct first stage ERAs without poles.

In experimenting with fitting our first-stage ERAs via these alternative methods, we also found that their error curves generally possess at least $2n + 1$ point of alternation (the requisite number for an error curve to qualify as a best approximant). The techniques differed, however, in the magnitudes of their error curves at their points of alternation. This is not surprising in view of the fact that the estimation methods used to estimate the rational coefficients in (32) and (33) are minimizing criteria that have heteroscedastic error terms.

Following our constructive approach to fitting the first stage ERAs, we experimented with a variety of alternative heteroscedasticity corrections to (31), (32), and (33) in an attempt to equilibrate the magnitudes of the error alternants. In the end, the most effective means by which we flattened the error curve was to oversample on the grid of ordinates at points where the error curve was highest (in absolute value). Oversampling tends to penalize these minimum norm techniques for not fitting these regions of the c.d.f. as well as to effectively "pull" down the error curve at these points. Although this discussion suggests that a major amount of tinkering may be necessary to generate a first-stage error curve with the desired equioscillation property, in practice we found that starting from an equispaced grid, it took only one or two trials in order to produce a nearly optimal approximant.

For computational and practical reasons we chose to develop and report first-stage ERAs that are based upon nonlinear least squares applied to Eq. (32).[4] Figures 2 and 3 graph the basic function that the nonlinear least squares routine is trying to fit. For small values of T, this ratio has an undulating character that cannot easily be captured by a low-degree polynomial. It appears, however, that as T increases the ratio becomes better behaved and that there is less work for nonlinear least squares to do in fitting a rational (or polynomial) approximant. Figure 3 shows, however, that this smoothing does not mean that the ERA has no curvature to capture.

In Figures 4(a) and (b) we have plotted examples of our first-stage ERA error curves. These error curves have alternating

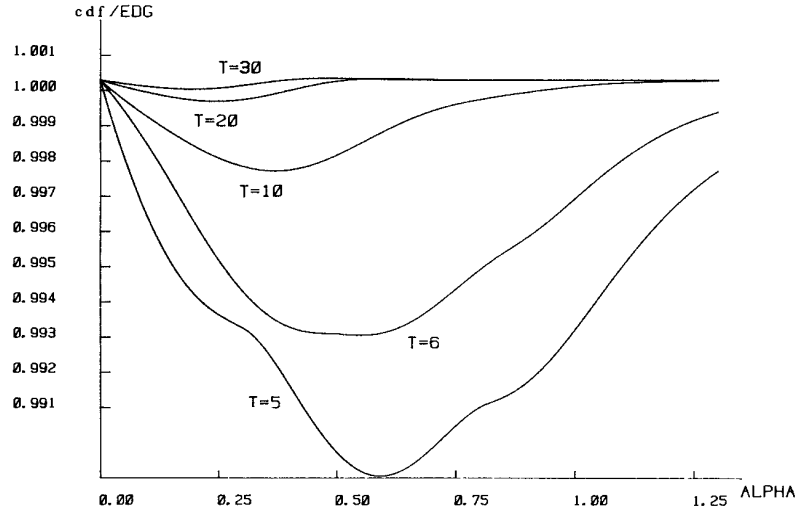

Figure 2. Ratio of the c.d.f. to the Edgeworth Approximant.

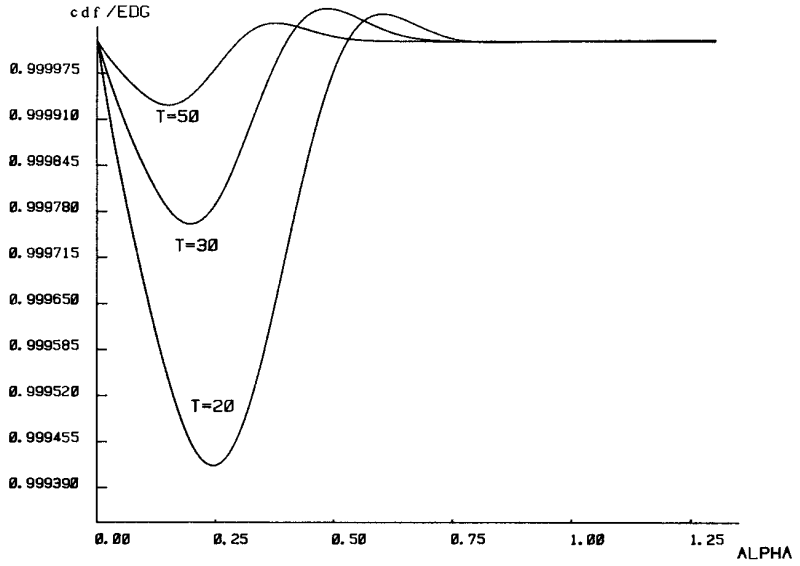

Figure 3. Ratio of the c.d.f. to the Edgeworth Approximant.

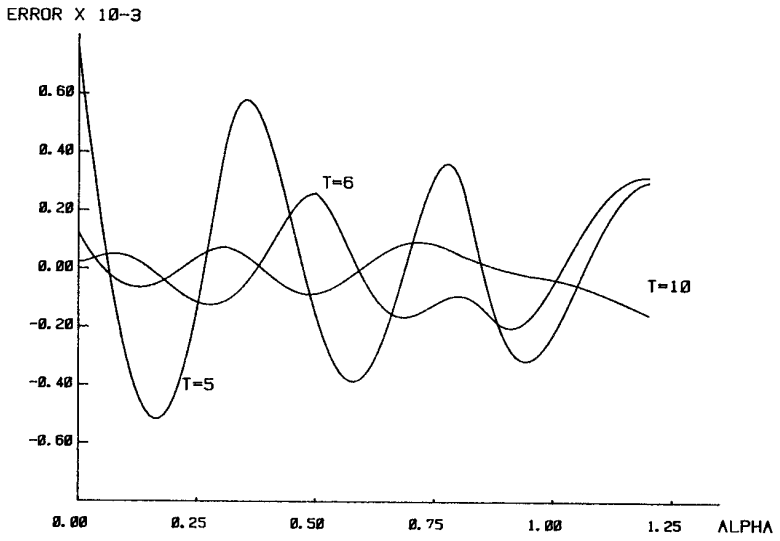

Figure 4(a). First-Stage Error Curves.

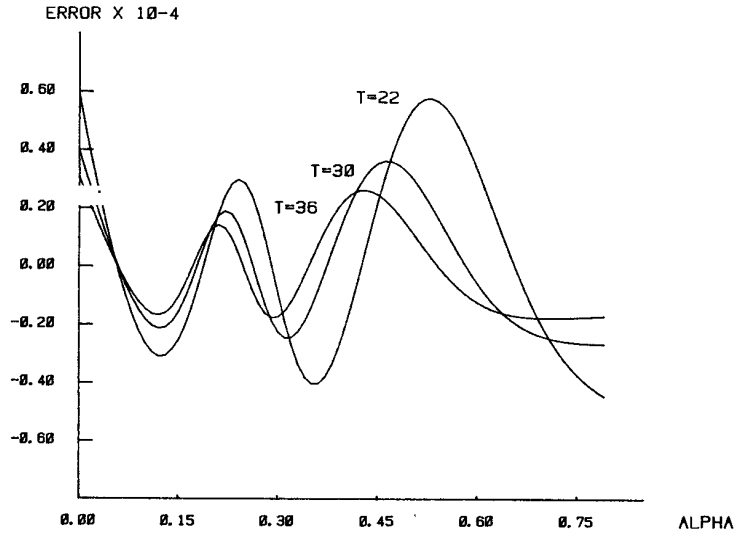

Figure 4(b). First-Stage Error Curves.

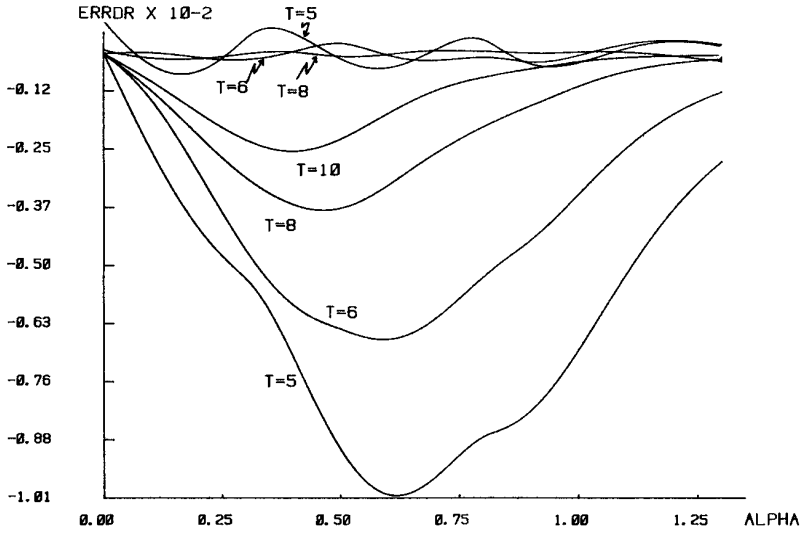

Figure 5. ERA and Edgeworth Errors.

properties that are much closer to those of the best approximant than those that have been obtained using Padé or other conventional uniform approximant methods (see for example [20] and [10]). In Figure 5 we have plotted for comparison the error curve on the primitive approximant against that of the first-stage ERA. Note that the original downward bias in the Edgeworth has been corrected and the absolute accuracy of the Edgeworth approximant has been considerably improved.

3.3. Formation of the Second Stage ERAs

When all of the first-stage ERAs have been found, work on the second-stage ERA can begin using the techniques discussed in Section 2.4. Our first step is to transform the first-stage rational function to the following form:

$$[2/2](x;\ \gamma(T)) = c_0(T) + \frac{\alpha_1(T)x + \alpha_2(T)}{1 + \beta_1(T)x + \beta_2(T)x^2}, \qquad (34)$$

which is more amenable to the preservation of asymptotic behavior as $T \uparrow \infty$. However, the form of (34) does not preclude the occurrence of poles in the second-stage ERAs. In our trials we encountered several (noninteger) poles using the above formulation. To

overcome this difficulty, we therefore rewrote Eq. (34) in partial fraction form (cf. (20)) as

$$[2/2](x;\, c(T), d(T)) = c_0(T) + \frac{c_1(T)}{x - d_1(T)} + \frac{\bar{c}_1(T)}{x - \bar{d}_1(T)} \tag{35}$$

$$(c_1 = c_{11} + ic_{12}, \qquad d_1 = d_{11} + id_{12})$$

$$= c_0(T) + \frac{2f_1(T)x + f_2(T)}{x^2 - 2d_{11}(T)x + (d_{11}(T)^2 + d_{12}(T)^2)} \tag{36}$$

$$\left(f_1(T) = \frac{\alpha_1(T)}{2\beta_2(T)} = c_{11}, \qquad f_2(T) = \frac{\alpha_2(T)}{\beta_2(T)}\right)$$

and we used these representations to develop our second-stage ERAs. Both the spline and the composite rational function techniques discussed in Section 2.3 were used to construct coefficient function approximants. The representation (35) was found to be the most adequate for the composite rational functions and the representation (36) was used for the splines.

To find the spline function approximants to the coefficients in (36) we computed the first-stage ERAs at $T = \{4, 5, 6, 7, 8, 9, 10, 12, 14, 16, 18, 20, 22, 26, 28, 30, 32, 36, 50\}$. These computations provided a native data set for the coefficient functions $\{c_0(T), f_1(T), f_2(T), d_{11}(T), d_{12}(T)\}$. For each of these coefficients a cubic spline was calculated from the data using Eq. (13). The resulting functions are displayed in Figures 6(a) to (e). The native data points are represented in these graphs by a cross. In each case, the functions stabilize after $T \geq 10$ and the approach to the asymptote as T grows large is smooth and well-behaved. The second-stage ERA based upon these splines as coefficient functions reproduces first-stage accuracy on the native data set and comes extremely close to achieving first-stage accuracy at intermediate values of T.

In fitting the second-stage ERAs to the coefficients in Eq. (35), we chose to estimate [5/2] and [6/2] as rational functions that employ powers of $T^{-1/2}$. Our choice of the order of the second-stage ERAs was dictated by a desire to reproduce as well as possible first-stage accuracy. Once again, nonlinear least squares was used to produce estimates of the rational coefficients. Plots of second-stage fits that are comparable to those in Figures 4(a) and (b) are provided in Figures 7(a) and (b). In Figure 7(c) we have plotted

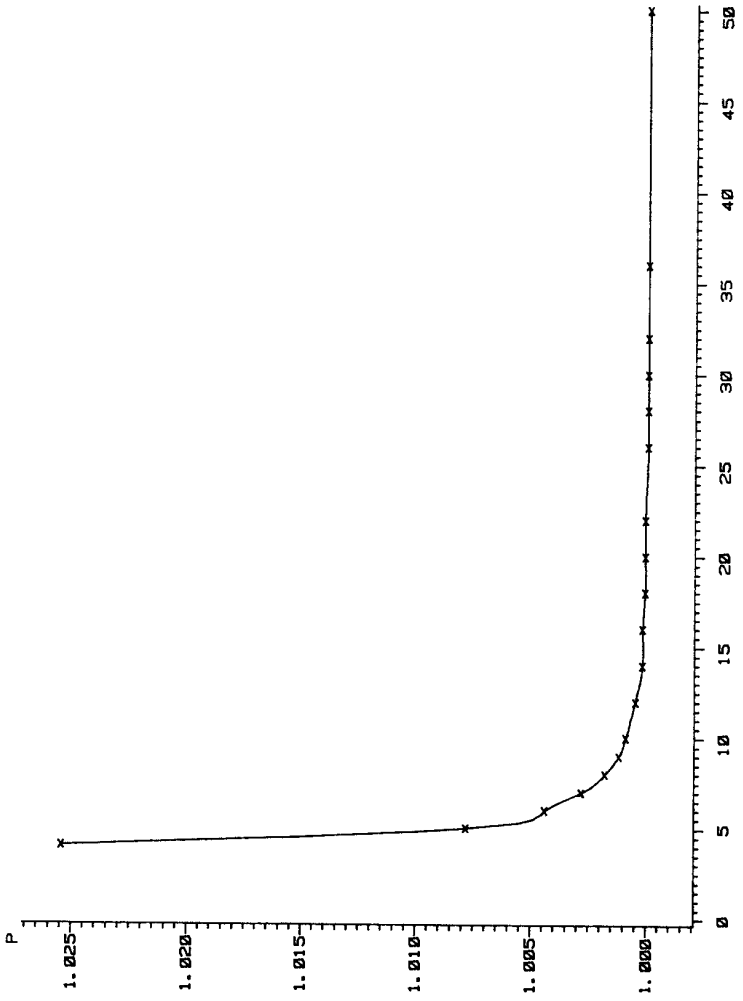

Figure 6(a). Spline Function for $c_0 = c_0(T)$.

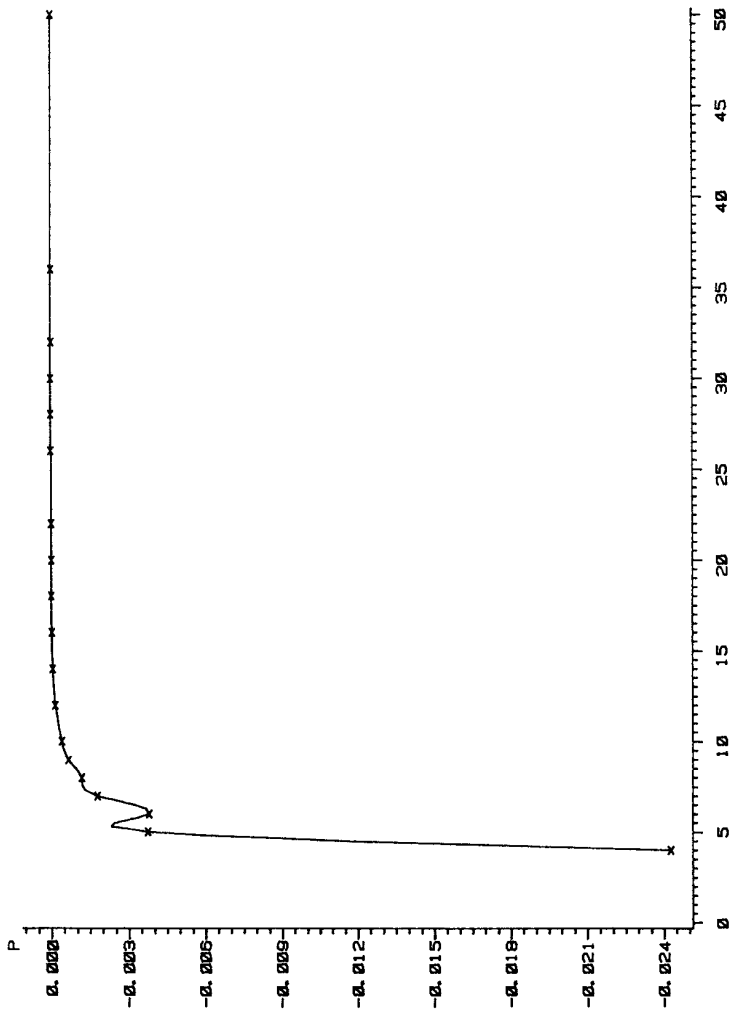

Figure 6(b). Spline Function for $f_1 = f_1(T)$.

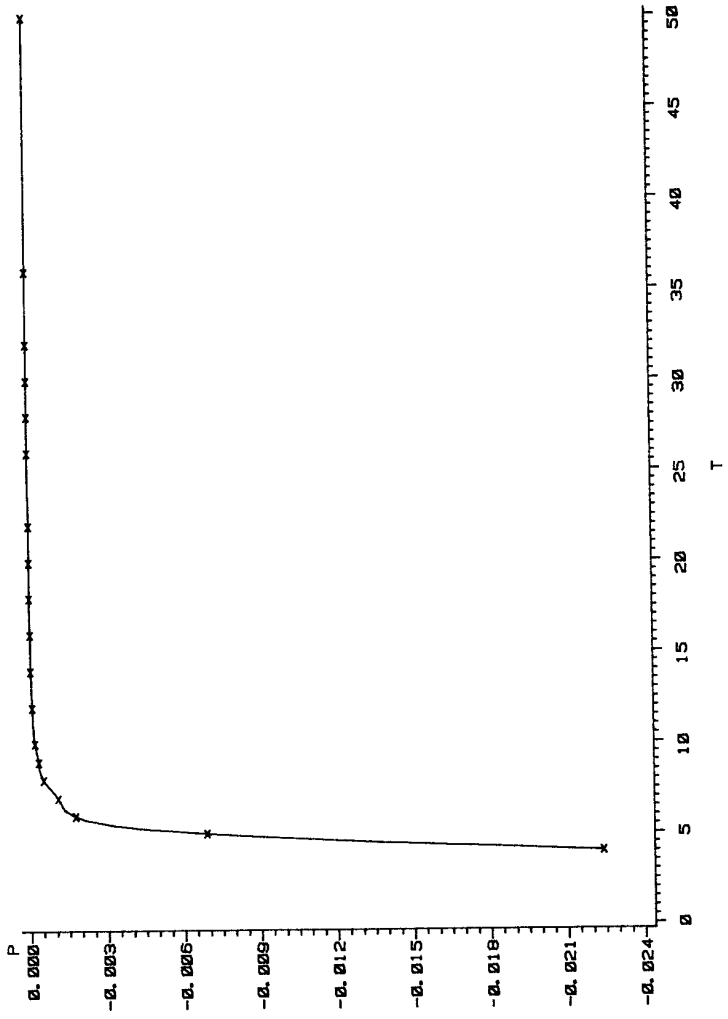

Figure 6(c). Spline Function for $f_2 = f_2(T)$.

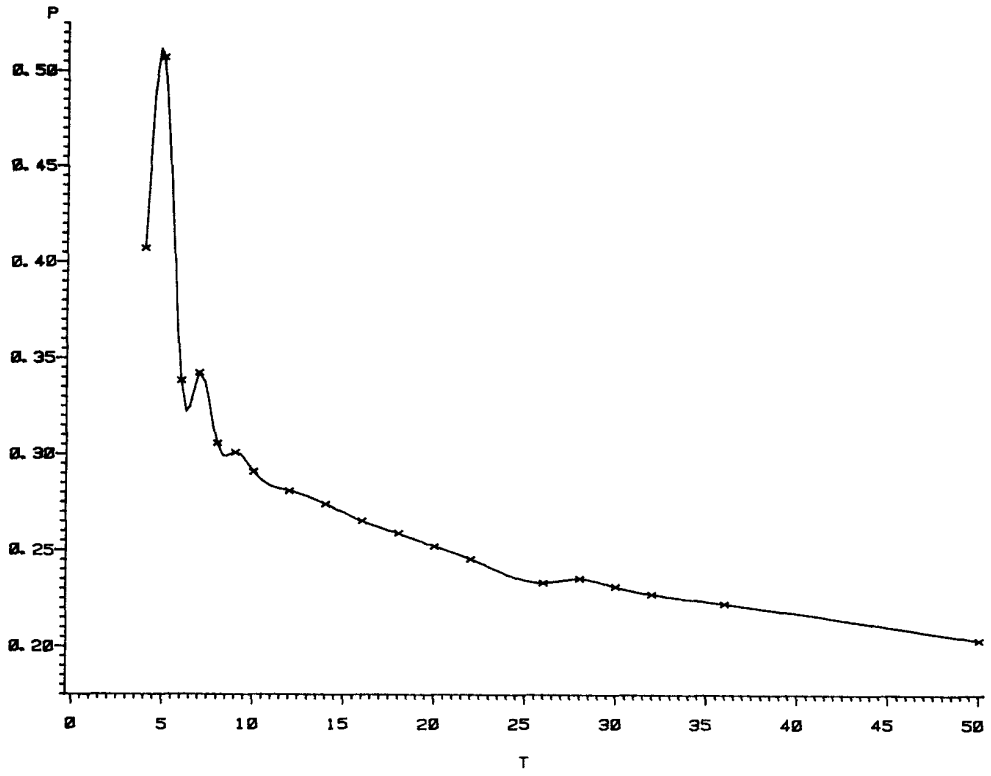

Figure 6(d). Spline Function for $d_{11} = d_{11}(T)$.

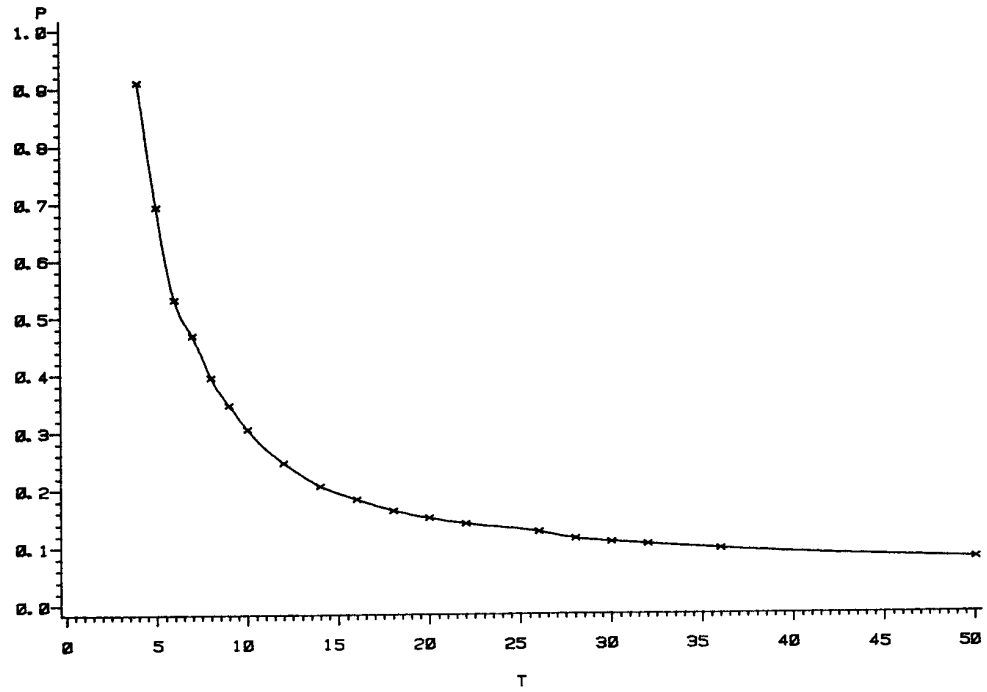

Figure 6(e). Spline Function for $d_{12} = d_{12}(T)$.

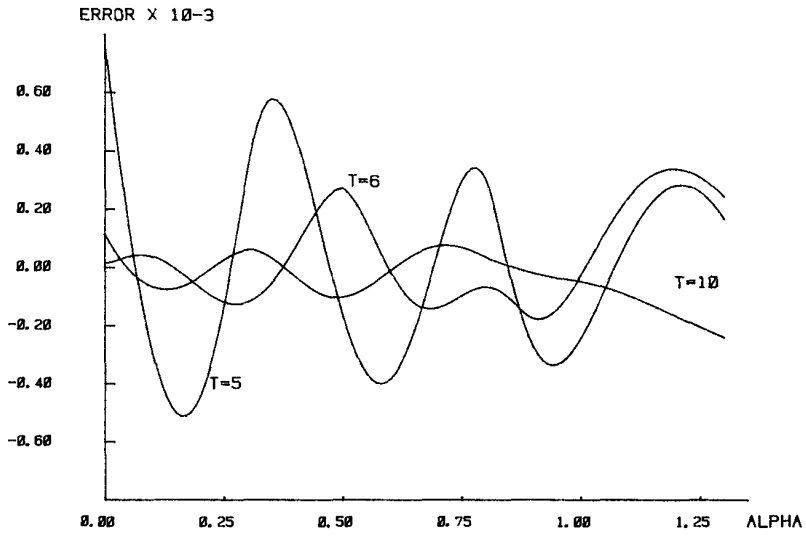

Figure 7(a). Second-Stage Error Curves.

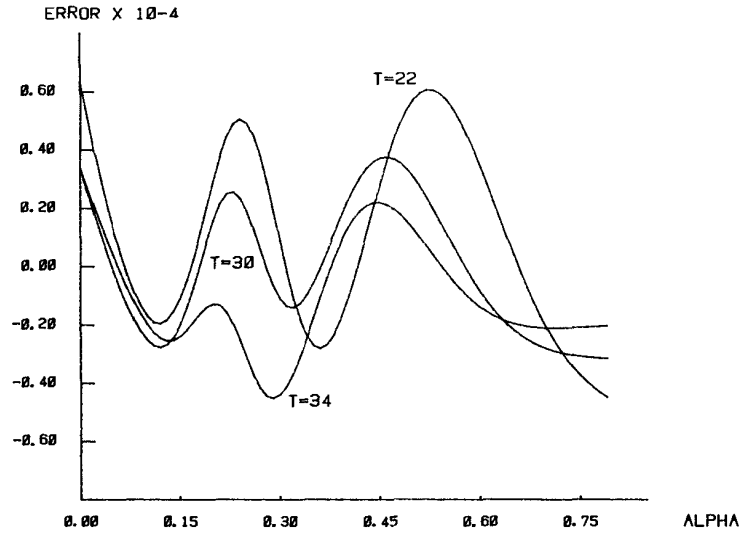

Figure 7(b). Second-Stage Error Curves.

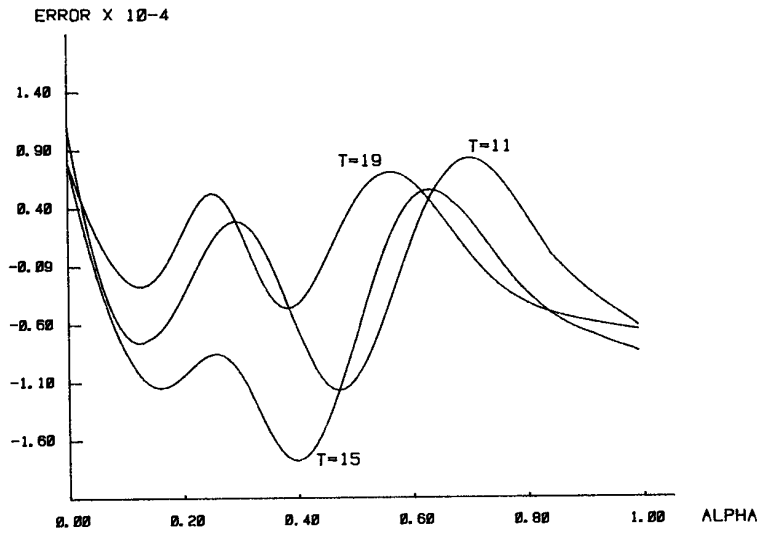

Figure 7(c). Second-Stage Error Curve (Predictions).

for comparative purposes the error curves for several sample sizes that were not part of the mesh over which we fit the first-stage ERAs.

3.4. Final Formulas

Based upon our extensive trials we have selected the following two rational approximants (detailed in full in Table 1 on page 36). These rational approximants satisfactorily reproduce first-stage accuracy and are ready to be programmed into regression software.

Note that these formulas are in a form that delivers probability values for an estimated value of α. It is also interesting to consider the opposite problem, which is to find a critical value of α given a desired size of the test. As it stands, it is impossible to solve Eq. (2) explicitly for the critical value of α. However, a numerical routine can perform this calculation quite simply and would require only a few extra lines of computer code for its implementation.

3.5. Density Approximation

Phillips [15] developed the necessary formulas and procedures for directly approximating probability density functions. These

Table 1. Final Formulas:

$$Era_T(x) = Ed_T(x)\left\{c_0 + \frac{2f_1x + f_2}{x^2 - 2d_{11}x + (d_{11}^2 + d_{12}^2)}\right\}$$

$$Ed_T(x) = \Phi\left(T^{1/2}x + \frac{1}{4T^{1/2}}(x + Tx^3)\right)$$

$$\Phi(\cdot) = \text{cdf}(N(0, 1))$$

A. The Spline Coefficient Function

The spline coefficient functions are defined (for $4 \le T \le 50$) by

$$s\ell_i(T) = \left\{(T - T_j)y_{j+1} + \frac{(T_{j+1} - T)y_j}{d_j} - (T - T_j)(T_{j+1} - T)\frac{(d_j + T_{j+1} - T)m_j + (d_j + T - T_j)m_{j+1}}{6d_j}\right\}$$

over the interval $T_j \le T \le T_{j+1}$. The constants for each coefficient function are given by the following:

Spline for $c_0 = c_0(T)$

T	Y	D	M
4	0.1025405D + 01	0 1000000D + 01	0.0
5	0.1007800D + 01	0.1000000D + 01	0.2212108D − 01
6	0.1004400D + 01	0.1000000D + 01	−0.3254305D − 02
7	0.1002800D + 01	0.1000000D + 01	0.1696145D − 02
8	0.1001800D + 01	0.1000000D + 01	0.6972678D − 04
9	0.1001200D + 01	0.1000000D + 01	0.4249483D − 03
10	0.1000900D + 01	0.2000000D + 01	0.3048016D − 04
12	0.1000500D + 01	0.2000000D + 01	−0.3914613D − 05
14	0.1000200D + 01	0.2000000D + 01	0.1351783D − 03
16	0.1000200D + 01	0.2000000D + 01	−0.8679855D − 04
18	0.1000100D + 01	0.2000000D + 01	0.6201592D − 04
20	0.1000100D + 01	0.2000000D + 01	−0.1126514D − 04
22	0.1000100D + 01	0.4000000D + 01	−0.1695536D − 04
26	0.1000000D + 01	0.2000000D + 01	0.1899865D − 04
28	0.1000000D + 01	0.2000000D + 01	−0.5081197D − 05
30	0.1000000D + 01	0.2000000D + 01	0.1326135D − 05
32	0.1000000D + 01	0.4000000D + 01	−0.2233420D − 06
36	0.1000000D + 01	0.1400000D + 02	0.6958635D − 08
50	0.9999985D + 00	0.0	0.0

Table 1 (cont.)

Spline for $f_1 = f_1(T)$

T	Y	D	M
4	$-0.2426631D-01$	$0.1000000D+01$	0.0
5	$-0.3746400D-02$	$0.1000000D+01$	$-0.3409020D-01$
6	$-0.3787600D-02$	$0.1000000D+01$	$0.1299414D-01$
7	$-0.1790900D-02$	$0.1000000D+01$	$-0.5658947D-02$
8	$-0.1164300D-02$	$0.1000000D+01$	$0.1421050D-02$
9	$-0.6563100D-03$	$0.1000000D+01$	$-0.7369129D-03$
10	$-0.4005600D-03$	$0.2000000D+01$	$0.1316174D-04$
12	$-0.1480000D-03$	$0.2000000D+01$	$-0.5943876D-04$
14	$-0.5030600D-04$	$0.2000000D+01$	$-0.7705692D-05$
16	$-0.1544600D-04$	$0.2000000D+01$	$-0.3989470D-05$
18	$0.1818000D-05$	$0.2000000D+01$	$-0\ 2730429D-05$
20	$0.8654800D-05$	$0\ 2000000D+01$	$-0.7296141D-06$
22	$0.1125800D-04$	$0.4000000D+01$	$-0.7015145D-06$
26	$0.1180300D-04$	$0.2000000D+01$	$0.7213256D-06$
28	$0.1325600D-04$	$0.2000000D+01$	$-0.1154174D-05$
30	$0.1232400D-04$	$0.2000000D+01$	$0.3178721D-06$
32	$0.1145500D-04$	$0.4000000D+01$	$-0.2281411D-07$
36	$0.9990400D-05$	$0\ 1400000D+02$	$0.1203126D-07$
50	$0.5661994D-05$	0.0	0.0

Spline for $f_2 = f_2(T)$

T	Y	D	M
4	$-0.2237575D-01$	$0.1000000D+01$	0.0
5	$-0.6880200D-02$	$0.1000000D+01$	$-0.1476514D-01$
6	$-0.1757200D-02$	$0.1000000D+01$	$-0.3174731D-02$
7	$-0.1076000D-02$	$0.1000000D+01$	$0.8132645D-03$
8	$-0.5128900D-03$	$0.1000000D+01$	$-0.7868675D-03$
9	$-0.3194200D-03$	$0.1000000D+01$	$0.1163656D-03$
10	$-0.1952500D-03$	$0.2000000D+01$	$-0.9439481D-04$
12	$-0.9449900D-04$	$0.2000000D+01$	$0.3618143D-05$
14	$-0.5347800D-04$	$0.2000000D+01$	$-0.9672759D-05$
16	$-0.3573100D-04$	$0.2000000D+01$	$0.1618934D-06$
18	$-0.2505000D-04$	$0.2000000D+01$	$-0.1573814D-05$
20	$-0.1867600D-04$	$0\ 2000000D+01$	$-0.3271356D-06$
22	$-0.1445000D-04$	$0.4000000D+01$	$-0.3396431D-06$
26	$-0.9288800D-05$	$0.2000000D+01$	$-0.5155303D-07$
28	$-0.7485200D-05$	$0.2000000D+01$	$-0.1768957D-06$
30	$-0.6224200D-05$	$0.2000000D+01$	$-0.5476412D-07$
32	$-0.5260700D-05$	$0.4000000D+01$	$-0\ 5029780D-07$
36	$-0.3876700D-05$	$0.1400000D+02$	$-0.2534953D-07$
50	$-0.1631507D-05$	0.0	0.0

Table 1 (cont.)

Spline for $d_{11} = d_{11}(T)$

T	Y	D	M
4	$0.4069132D+00$	$0\ 1000000D+01$	0 0
5	$0\ 5067900D+00$	$0.1000000D+01$	$-0.5119480D+00$
6	$0.3380800D+00$	$0.1000000D+01$	$0\ 4362711D+00$
7	$0.3421300D+00$	$0.1000000D+01$	$-0.1965766D+00$
8	$0\ 3056000D+00$	$0.1000000D+01$	$0.1065551D+00$
9	$0.3010500D+00$	$0.1000000D+01$	$-0\ 3776378D-01$
10	$0.2910600D+00$	$0.2000000D+01$	$0.1186003D-01$
12	$0.2810900D+00$	$0.2000000D+01$	$-0.1683196D-02$
14	$0.2741300D+00$	$0.2000000D+01$	$-0.6122438D-03$
16	$0.2652600D+00$	$0.2000000D+01$	$0.1267171D-02$
18	$0.2589600D+00$	$0\ 2000000D+01$	$-0\ 6014412D-03$
20	$0.2522000D+00$	$0.2000000D+01$	$0.4485934D-03$
22	$0.2456300D+00$	$0.4000000D+01$	$-0.9079323D-03$
26	$0.2335400D+00$	$0.2000000D+01$	$0.2893250D-02$
28	$0.2355900D+00$	$0.2000000D+01$	$-0.3401137D-02$
30	$0.2312100D+00$	$0\ 2000000D+01$	$0\ 1066299D-02$
32	$0.2275000D+00$	$0.4000000D+01$	$0.1409421D-03$
36	$0.2225400D+00$	$0.1400000D+02$	$-0\ 3347571D-04$
50	$0.2036835D+00$	0.0	0.0

Spline for $d_{12} = d_{12}(T)$

T	Y	D	M
4	$0.9093072D+00$	$0.1000000D+01$	0 0
5	$0.6928200D+00$	$0.1000000D+01$	$0.4110065D-01$
6	$0\ 5299400D+00$	$0.1000000D+01$	$0.1572406D+00$
7	$0.4673700D+00$	$0\ 1000000D+01$	$-0.6820309D-01$
8	$0.3949800D+00$	$0.1000000D+01$	$0.5665174D-01$
9	$0.3473300D+00$	$0.1000000D+01$	$-0.9963870D-02$
10	$0.3049200D+00$	$0.2000000D+01$	$0.1464374D-01$
12	$0.2469900D+00$	$0.2000000D+01$	$0.1385714D-02$
14	$0.2066300D+00$	$0.2000000D+01$	$0.6168405D-02$
16	$0.1832600D+00$	$0.2000000D+01$	$-0.5743319D-03$
18	$0.1639800D+00$	$0.2000000D+01$	$0.2263923D-02$
20	$0.1505800D+00$	$0.2000000D+01$	$0.3386396D-03$
22	$0.1404300D+00$	$0.4000000D+01$	$0.1256519D-02$
26	$0.1256200D+00$	$0.2000000D+01$	$-0.1880125D-02$
28	$0.1142500D+00$	$0.2000000D+01$	$0.2820215D-02$
30	$0.1088400D+00$	$0.2000000D+01$	$-0\ 4607346D-03$
32	$0.1041600D+00$	$0.4000000D+01$	$0.1177235D-03$
36	$0.9549900D-01$	$0.1400000D+02$	$0.1393219D-03$
50	$0.7798729D-01$	0.0	0.0

For $T > 50$,

$$c_0 = 1, \quad f_1 = f_2 = d_{11} = 0, \quad \text{and} \quad d_{12} = \varepsilon.$$

Table 1 (cont.)

B. The Second Stage ERA Coefficient Function

The second stage ERAs were fitted to

$$[2/2](x; c(T), d(T)) = c_0(T) + \frac{c_{11}(T) + ic_{12}(T)}{x - d_{11}(T) - id_{12}(T)} + \frac{c_{11}(T) - ic_{12}(T)}{x - d_{11}(T) + id_{12}(T)}$$

according to

$$c_0 = [5/2]$$

$$c_{11} = \frac{1}{\sqrt{T}}[5/2]$$

$$c_{12} = \frac{1}{\sqrt{T}}[6/2]$$

$$d_{11} = \frac{1}{\sqrt{T}}[5/2]$$

$$d_{12} = \frac{1}{\sqrt{T}}[5/2]$$

Numerator:

	c_0	c_{11}	c_{12}	d_{11}	d_{12}
0	−.00375346	00303693	−.000220875	1.84872	.841728
1	.0823311	−.0610821	−.0109917	−14.9849	−10.3950
2	−.711465	.482070	.274271	23.1109	60.1172
3	3.01684	−1 85280	−2.32525	84.8834	−199.104
4	−6.22756	3.45125	9.52452	−315.163	309.302
5	4.91787	−2.48854	−19.1841	297.993	−144 937
6			15.1337		

Denominator:

	c_0	c_{11}	c_{12}	d_{11}	d_{12}
1	−7.38542	−5.13001	−6.21730	−7.02527	−7.12108
2	11.1815	6.53295	9.11970	11.0877	11.5276

The following expressions convert these coefficients into those in Eq. (34):

$$c_0 = c_0$$

$$\beta_1 = -2\beta_2 d_{11} \qquad \beta_2 = \frac{1}{d_{11}^2 + d_{12}^2}$$

$$\alpha_1 = 2\beta_2 c_{11} \qquad \alpha_2 = -(4\beta_2 - \beta_1^2)^{1/2} c_{12} + \frac{\alpha_1 \beta_1}{2\beta_2}.$$

techniques are similar in principle to those developed here for the c.d.f. and they have been shown to yield very accurate approximants to the exact p.d.f. of the serial correlation coefficient. One natural question that is raised here by our choice of approximating the c.d.f. is whether the derivative of our ERA c.d.f. approximant yields an accurate implicit approximant to the p.d.f. (The parallel question could be posed for p.d.f. approximants: Does the integral of the p.d.f. ERA discussed in [15] yield a good approximant to the c.d.f.?) In view of the denseness property (Theorem 3 above), the ERA can be made arbitrarily close to the c.d.f. by increasing the degree of the ERA. But since the c.d.f. is in this case smooth as well as continuous, and since the ERA is also smooth, the derivative of the ERA should be close to the actual p.d.f. if the c.d.f. approximant is close enough. Further, we note that the derivative of the c.d.f. error curve is the error curve for what we will call the implicit p.d.f. approximant and that the best c.d.f. approximant has at least N points at which the derivative of the error curve vanishes. Thus, the error curve of the implicit p.d.f. approximant will have at least $N + 1$ points of alternation. Our main interest is in the performance of the implicit p.d.f.

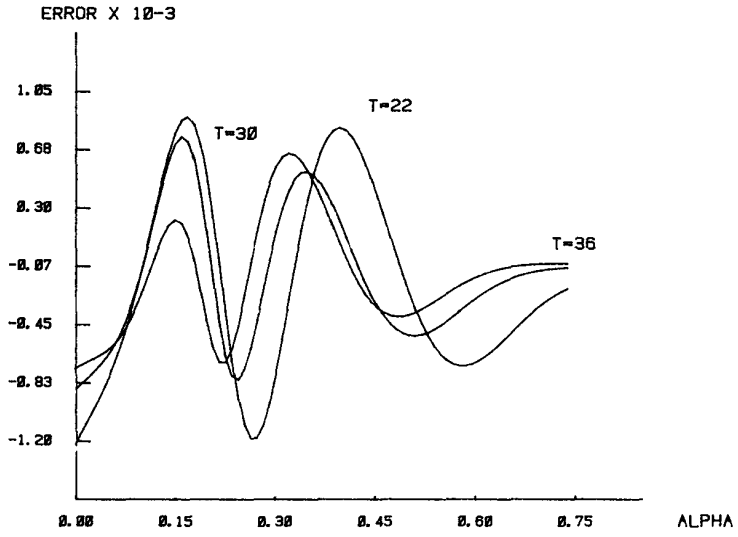

Figure 8. Implicit p.d.f. Error Curves.

approximant in the cases of low-degree rational approximation that are most relevant for practical applications. Here we have some systematic evidence from our own applications. In particular, Figure 8 displays the error curves of several of our implicit p.d.f. approximants. Despite the low order of the c.d.f. approximant, these graphs show that the implicit p.d.f. approximant does very well in approximating the p.d.f. (for comparison, see the p.d.f. error curves in [18]).

4. CONCLUSIONS AND EXTENSIONS

This paper has extended Phillips's ERA techniques in two important directions. First, it has extended the algebraic theory of ERAs to cover distribution function approximation. Second, it has introduced composite functional approximants that are capable of handling the parameter multidimensionality frequently encountered in applied problems. In Section 3 we used these techniques to approximate the distribution of the serial correlation coefficient. We found that we could satisfactorily approximate the c.d.f. of the autoregressive parameter using ERAs that employed either cubic splines or composite rational functions in the formulation of the second-stage ERA.

In conclusion, we want to underscore the constructive nature of the ERAs discussed in this paper. Their flexible form provides substantial leeway in the selection of the theoretical, numerical, and Monte Carlo information that can be embodied in the approximant. In general, it appears that the more work the leading coefficient function can do in fitting the c.d.f., the less the work that is required of the rational coefficients. There are also several other trade-offs that exist in the construction of ERAs. First, the degree of the approximant must be selected so as to balance the desired second-stage precision against the computational burden of fitting composite coefficient functions that do not have poles. Our trials suggest that for most practical applications, even very low-degree ERAs will provide the accuracy necessary for statistical tests (often [2/2]'s and at most [4/4]'s). A second practical consideration that arises in the construction of these ERAs is the selection of the discrete parameter grid over which the ERAs will be fitted. Although the precise selection criteria will depend upon the particular application, our experience has been that it is most

important to select a more concentrated grid over those parameter regions where the ratio of the c.d.f. to $S(x)$ changes rapidly.

The c.d.f. approximation techniques developed in this paper have a wide number of uses and many extensions of our own application are now possible. The most straightforward extension would be to functionalize the serial correlation coefficients on α as well as T. This higher-order composite function would allow the investigator to conduct power computations directly for some of the tests described in Section 3.1. Another extension would be to treat the problem of serial correlation with a fitted mean and to functionalize the ERA on the intercept parameter. Similar ERAs may be constructed for higher-order serial correlation coefficients. These ERAs would enable exact tests to be performed in the Box–Jenkins model identification process. In particular, composite function ERAs for these serial correlation coefficients could deliver critical values for any sample size and chosen test size.

Finally, although the methods we have used in this paper are not the only techniques that are available for the constructive development of distribution approximants, they do have the advantage over many other methods because they readily integrate information about the true distribution function from diverse analytic and numerical sources. Our theoretical results also provide a rigorous basis for the development of approximants in the ERA family. Recently, Nankervis and Savin [12] have successfully used response surface regressions from experimental Monte Carlo data to correct the distribution of the "t" statistic in an AR(1). Their approach involves an empirical search for an appropriate response surface representation of the mean and standard deviation of their statistic as a function of the parameters of the model. The response surfaces are then used to correct conventional inferential procedures based upon the "t" distribution. The goal of this research is therefore closely related to our own: to provide accurate and easily implemented corrections to the distributions upon which we rely in econometric inference.

APPENDIX: COMPUTATION OF BEST APPROXIMANTS

This appendix describes a new method for constructing the best (uniform) extended rational approximant to a real-valued, univari-

ate, continuous function. The method is based upon least squares techniques that use information from both critical and collocation points of the error curve to update estimates of the approximant. Although we have not found it necessary to apply this technique to the distribution function of the serial correlation coefficient, in practice this technique could readily be used to further improve the performance of our first-stage ERAs.

1. Existing Algorithms. Existing rational approximation algorithms are either based upon the equioscillation alternate property of the best approximant or they directly minimize the maximum alternate error via nonlinear programming techniques. The relative performance of each of these algorithms has been studied by a number of investigators (most notably [10]). The general conclusion that can be drawn from these experiments is that for small grids of function values, the direct methods are more reliable because they continually minimize the maximum error of the approximation. One drawback to these direct algorithms, however, is their slow convergence to the best approximant. Indeed, they often require an average of between five to fifty times as much computer time as the methods that are based upon the equioscillation theorem.

The computational complexity of the direct algorithms has prompted a number of attempts to improve the reliability of the more simple methods. The principal reliability problems of the simpler algorithms, however, have yet to be adequately handled. These problems are generally twofold: the algorithms can produce approximants with poles or approximants with too few oscillations. Both of these problems appear to stem from the fact that these techniques do not use much information about how the function is changing relative to some reference set of critical or collocation points (see [11]). It would seem, therefore, that if any progress is to be made in improving the behavior of these simple algorithms, it must be done by introducing more information about the behavior of the error curve into the approximant. The following algorithm uses the interpolation techniques discussed in Section 3.2.

2. A Least Squares Algorithm. Consider an initial approximation to $cdf_T(x)$ that has (at least) the required number of alternations

for it to be a best approximant. For this initial rational approximant, there are a set of equations that define the collocation and critical points of the approximant's error curve. These points form a reference set, M_0, to the best approximant's $i = 1, \ldots, n + m + 2$ (at least) critical points and $j = 1, \ldots, n + m + 1$ collocation points. Using the equioscillation theorem, we know that the best approximant's reference set, M^*, is defined by the following equations:

$$\text{for critical points:} \qquad cdf_T(x_i^*) - Era_T(x_i^*) = \lambda(-1)^i \qquad \text{(A.1)}$$

$$\text{for collocation points:} \qquad cdf_T(x_j^*) - Era_T(x_j^*) = 0. \qquad \text{(A.2)}$$

When these equations are multiplied by the denominator of $Era_T(x)$ we obtain the following system of equations in terms of λ and the rational coefficients (for any reference set M_0):

$$cdf_T(x) = \begin{cases} -Q(x_i)cdf_T(x_i) + Ed_T(x_i)P(x_i) + \lambda(-1)^i Q(x_i) \\ \quad + \lambda(-1)^i + \varepsilon_i \\ -Q(x_j)cdf_T(x_j) + Ed_T(x_j)P(x_j) + \varepsilon_j, \end{cases}$$

where the terms $\{\varepsilon_i, \varepsilon_j\}$ represent errors in the approximation of the true set of critical and collocation points $M^* = \{x_1^*, \ldots, x_v^*\}$ by $M_0 = \{x_1, \ldots, x_v\}$. Writing these equations as a system of $v = 2(m + n + 1) + 1$ equations, we have

$$y = W\lambda + Zq + Xp + \varepsilon, \qquad \text{(A.3)}$$

where

$$Z = \begin{bmatrix} -x_1(y_1 - w_1\lambda) & -x_1^2(y_1 - w_1\lambda) & \ldots & -x_1^n(y_1 - w_1\lambda) \\ \vdots & \vdots & & \vdots \\ -x_v(y_v) & -x_v^2(y_v) & \ldots & -x_v^n(y_v) \end{bmatrix}$$

$$X = \begin{bmatrix} Ed_T(x_1) & x_1 Ed_T(x_1) & x_1^2 Ed_T(x_1) & \ldots & x_1^m Ed_T(x_1) \\ \vdots & \vdots & \vdots & & \vdots \\ Ed_T(x_v) & x_v Ed_T(x_v) & x_v^2 Ed_T(x_v) & \ldots & x_v^m Ed_T(x_v) \end{bmatrix}$$

$$y = \begin{bmatrix} cdf_T(x_1) \\ \vdots \\ cdf_T(x_v) \end{bmatrix} \qquad W = \begin{bmatrix} -1 \\ 1 \\ \vdots \\ 0 \end{bmatrix}.$$

Apart from the nonlinearity of the λ appearing in Z, Eq. (A.3) has the form of a linear regression equation. This suggests that if we start with an initial reference set M_0 that is close to M^*, we can produce a new rational approximant (and a new reference set) based upon the following procedure. Holding the λ in Z fixed, apply least squares to (A.3) so as to minimize the squared error on the old reference set. This produces new values of the rational coefficients and λ, as well as a new reference set, M_1, that can then be used as input into subsequent revisions of the approximant.

Two natural questions arise here: why is this a sensible procedure; and why should this algorithm converge? The answer to the first question rests in the way the algorithm includes information about the behavior of the rational approximant's error curve. To see this, consider the two approximation methods that are nested within this procedure. On the one hand, when only the critical points are used to solve (A.3), this algorithm is a version of the generalized Remes algorithms and the solution to (A.3) is exact. On the other hand, when only the collocation points of the rational approximant are included in Eq. (A.3), we have a multiple-point Padé approximant as a special case.

The second question posed above has yet to be evaluated formally. It does appear, however, that some of the local convergence results for generalized Remes algorithms can be applied to this algorithm. In practical applications of the least squares method, the algorithm usually converges very rapidly to the best approximant (provided the algorithm has a sufficiently good initial approximant). As is the case with all algorithms based upon the linearizations contained in Eq. (A.3), however, the algorithm is not entirely reliable in that given a poor initial approximant, the algorithm can produce an approximation with poles or an approximation that has too few alternants. One of the attractive features of this algorithm, however, is that the least squares approach suggests ways in which these degeneracies can be overcome.[5]

3. Numerical Examples. This section illustrates how to apply the algorithm and reports some preliminary results on its performance. The results described here are, however, by no means a complete evaluation of the algorithm. There are many factors that control the performance of the algorithm (such as the size of the initial reference grid, interpolation formulas, and the choice of $Ed_T(x)$)

and a great deal of work remains to be done before the significance of each factor is known.

One final issue that must be resolved before the algorithm can be implemented is the question of how to treat the nonlinearity of λ in Z. A wide range of strategies for coping with the nonlinearity are possible. Only three such possibilities are considered here. The first way of proceeding is the naive way of simply setting λ equal to some constant. The second way is to set λ equal to the maximum error on the preceding iteration's rational approximant. The final way is to use the least squares estimate from the previous iteration. These schemes will be denoted A1, A2, and A3.

EXAMPLE 1. In this first example we approximate e^x over the unit interval by a [2/2] rational function. $Ed_T(x)$ has been set equal to 1 for all x, $N = 101$ and $\{x_1, x_2, \ldots, x_{101}\} = \{.00, .01, \ldots, 1.00\}$. The initial approximant is obtained by applying least squares to (A.3). This yields the following rational approximant (whose error curve is plotted in Figure 1):

$$a_0 = 1.00000635 \qquad a_1 = .542098 \qquad a_2 = .108358$$
$$b_0 = 1.0 \qquad b_1 = -.457717 \qquad b_2 = .0648901$$

with critical points

$$\{-.635, .317, -.342, .275, -.660, .865\} \times 10^5.$$

This least squares approximant is extremely close to the best approximant. Using the iterative least squares procedures described above, we obtain the following results:

Method A1						
Iteration	*Maximum Error × 100000*					
1	−.519	395	−.414	.371	−.536	.560
5	−.456	.446	−.447	.437	−.454	.447
10	− 477	477	−.477	.477	−.477	.477

Method A2						
Iteration	*Maximum Error × 100000*					
1	− 582	.429	−.430	366	− 484	.472
5	− 477	477	−.477	.477	−.477	.477

Method A3						
Iteration	*Maximum Error × 100000*					
1	−.571	.423	− 426	.368	−.493	.487
5	− 451	.448	−.447	.442	−.449	.447
10	− 477	477	−.477	.477	−.477	477

In each case, the algorithm converges quickly to the best approximant.[6] Although the method that uses the upper bound estimate of λ appears to be the most rapid method, this is not always true. For example, if the fixed-lambda method uses a more accurate estimate of the best approximant's lambda, then the fixed-lambda method generally will converge at a faster rate.

EXAMPLE 2. The second application of this method is to the probability density function (p.d.f.) considered by Phillips [18]. Although Phillips did find a very accurate approximant to the p.d.f. using Padé techniques with a [4/4] ERA, the final approximant left room for improvement. In particular, its error curve oscillated 14 times (four more than necessary for a best approximant) and had a maximum error of $.125 \times 10^{-3}$. Although the error oscillations were of roughly similar magnitude, there was enough variation to indicate that there was room for further improvement. The application of our least squares technique to his problem yielded the following results. The critical points of the initial, equispaced grid approximant are

$$\{.116, -.106, .094, -.066, .045, .061, -.052,$$
$$.040, -.031, .014, -.007\} \times 10^{3}.$$

Note that this initial approximant alternates two more times than is necessary for the best approximant. By applying the fixed-lambda method ($\lambda = .00009$), we obtain the following results:

Method A1						
Iteration	*Maximum Error × 1000*					
1	.095	−.107	.082	−.067	.072	− 081
	.050	−.029	.046	− 050		
5	.084	−.083	.082	−.087	.088	−.090
	.103	− 050	094	−.088		
10	.085	−.084	.082	−.085	.087	−.086
	.083	−.083	.086	−.087		

After one iteration, the approximant reverts to the minimum number of oscillations for the best approximant. After only 11 iterations, we obtain an approximant that is very close to the best approximant.

ACKNOWLEDGMENTS

We are extremely grateful to Alison Gold for performing part of the calculations found in Section 3. Phillips's research was supported by the NSF under Grant No. SES 8218792 and by the John Simon Guggenheim Foundation. Reiss's research was supported by Stanford Graduate School of Business research funds.

NOTES

1. For example, they typically have leading analytic terms that are multiplied by multiple infinite series of invariant polynomials. Although the leading terms are easily computed (and recognizable among the most commonly encountered multivariate statistical distributions), the infinite series portion of the function normally presents computational difficulties in all but the simplest cases.

2. The reader may wish to prove that the distribution is indeed symmetric in this noncircular case.

3. See [18]. Although this algorithm makes efficient use of symmetric matrix eigenvalue routines, we have found that repetitive use of this program can be extremely expensive. Fortunately, our computations here involve fixed and not variable costs.

4. Nonlinear least squares applied to equation (32) was favored over nonlinear least squares applied to Eq. (31) because the multiplicative scaling factor of the primitive

approximant adversely affected the convergence of the nonlinear least squares algorithm We also found that estimating the ratio of the c.d.f. to the primitive approximant produced more satisfactory error curves.

5. This may explain why in practice the Remes algorithms fail to converge for very simple functions. For example, when a Tchebycheff reference set is used for $\sqrt{x}$ on the unit interval, Remes algorithms invariably have trouble converging because most of the critical points for the best approximant are bunched near the origin. Other methods of generating an initial approximant (such as multiple-point Padé) have similar defects in that they also tend to overlook information about how the function is changing.

6. The average amount of CPU time on a DEC 20 for each iteration is about 0 25 seconds.

REFERENCES

[1] Achieser, N. I. (1956), *Theory of Approximation.* New York: Frederick Ungar.

[2] Ahlberg, J. H., E. N. Nilson, and J. L. Walsh (1967), *The Theory of Splines and Their Applications.* New York: Academic Press.

[3] Anderson, T. W. (1948), On the theory of testing for serial correlation. *Skandinavisk Aktuarietidskrift, 31,* 88–116.

[4] Brown, J. A. and M. S. Henry (1975), Best Tchebyshev composite approximation. *SIAM Journal of Numerical Analysis, 12,* 336–344.

[5] Durbin, J. and G. Watson (1950), Testing for serial correlation in least squares regression, I. *Biometrika, 37,* 409–428.

[6] Handscomb, D. C. (1966), Spline functions. Chapter 19 in D. C. Handscomb (ed.), *Methods of Numerical Approximation.* Oxford: Pergamon Press.

[7] Hastings, C. (1955), *Approximations for Digital Computers.* Princeton: Princeton University Press.

[8] Henry, M. S. and S. E. Weinstein (1974), Best rational approximation of functions: II. *Journal of Approximation Theory, 12,* 6–22.

[9] Imhof, P. J. (1961), Computing the distribution of quadratic forms in normal variables. *Biometrika, 48,* 419–426.

[10] Lee, C. M. and F. D. K. Roberts (1973), A comparison of algorithms for rational approximation. *Mathematics of Computation, 27,* 111–127.

[11] Meinardus, G. (1967), *Approximation of Functions: Theory and Numerical Methods.* Berlin: Springer-Verlag.

[12] Nankervis, J. C. and N. E. Savin (1984), *The Student's Approximation in a Stationary First-Order Autoregressive Model.* Mimeograph, Trinity College, Cambridge.

[13] Phillips, P. C. B. (1977), Approximations to some finite sample distributions associated with a first-order stochastic difference equation. *Econometrica, 45,* 463–486.

[14] Phillips, P. C. B. (1978), Edgeworth and saddlepoint approximations in a first-order non-circular autoregression. *Biometrika, 65,* 91–98.

[15] Phillips, P. C. B. (1982), Best uniform and modified Padé approximation of probability densities in econometrics. In W. Hildenbrand (ed.), *Advances in Econometrics.* Cambridge: Cambridge University Press, pp. 123–167.

[16] Phillips, P. C. B. (1983), Marginal densities of instrumental variable estimators in the general single equation case. In *Advances in Econometrics*, Volume 2. Cambridge: Cambridge University Press.

[17] Phillips, P. C. B. (1982), Failure of the alternation theorem in Tchebycheff approximation by rational functions over $C_0[-\infty, \infty]$. *Cowles Foundation Discussion Paper*, No. 638.

[18] Phillips, P. C. B. (1983), ERA's: A new approach to small sample theory. *Econometrica*, *51*, 1505-1525.

[19] Phillips, P. C. B. (1983), Finite sample econometrics using ERA's. *Text of Invited Lecture to the 51st annual meeting of the Japan Statistical Society.*

[20] Rivlin, T. J. (1969), *An Introduction to the Approximation of Functions.* New York: Dover.

[21] Rice, J. R. (1964/67), *The Approximation of Functions, Vol. I and II.* Reading, MA: Addison-Wesley.

[22] Sargan, J. D. and L. Bhargava (1983), Testing residuals from least squares for being generated by the Gaussian random walk. *Econometrica*, *51*, 153-174.

[23] Simmons, G. F. (1963), *An Introduction to Topology and Modern Analysis.* New York: McGraw-Hill.

[24] Weinstein, S. E. (1969), Approximation of functions of several variables: Product Chebyshev approximations, I. *Journal of Approximation Theory*, *2*, 433-447.

[25] White, John S. (1982), Edgeworth series expansions for the distribution of the serial correlation coefficient. *Applied Time Series Analysis*, 481-497.

[26] Zelen, M. and N. C. Severo (1965), Probability functions. In M. Abramowitz and I. A. Stegun (eds.), *Handbook of Mathematical Functions.* New York: Dover, pp. 927-995.