Cowles Foundation Paper 573

Econometrica, Vol. 51, No. 5 (September, 1983)

ERA'S: A NEW APPROACH TO SMALL SAMPLE THEORY

BY P. C. B. PHILLIPS[1]

This article proposes a new approach to small sample theory that achieves a meaningful integration of earlier directions of research in this field. The approach centers on the constructive technique of approximating distributions developed recently by the author in [**10**]. This technique utilizes extended rational approximants (ERA's) which build on the strengths of alternative, less flexible approximation methods (such as those based on asymptotic expansions) and which simultaneously blend information from diverse analytic, numerical and experimental sources. The first part of the article explores the general theory of approximation of continuous probability distributions by means of ERA's. Existence, characterization, error bound, and uniqueness theorems for these approximants are given and a new proof is provided for the convergence result obtained earlier in [**10**]. Some further aspects of finding ERA's by modifications to multiple-point Padé approximants are presented and the new approach is applied to the noncircular serial correlation coefficient. The results of this application demonstrate how ERA's provide systematic improvements over Edgeworth and saddlepoint techniques. These results, taken with those of the earlier article [**10**], suggest that the approach offers considerable potential for empirical application in terms of its reliability, convenience, and generality.

1. INTRODUCTION

ANALYTICAL RESEARCH on the small sample properties of econometric methods of estimation and testing has taken three main directions. The first of these has involved the mathematical task of extracting the form of the probability density function (pdf) or cumulative distribution function (cdf) of the relevant statistic under assumptions concerning the structure of the model and the stochastic properties of the errors that drive its equations. The second has been concerned with characterizing the distribution by the analysis and approximation of its moments. This has included work on the question of the existence of moments and conditions for asymptotic approximations to them to be valid. Finally, there has in recent years been a growing literature concerned with the derivation of direct approximations to the distributions themselves. These approximations have frequently been obtained by truncating asymptotic series expansions after a small number of terms. This literature has also investigated the validity of expansions as asymptotic series, explored questions of higher order optimality, and considered the actual numerical performance of the approximations in a variety of situations.

In addition, and largely in parallel to this analytical research, are the experimental and purely numerical investigations. The former have continued tradi-

[1] This article is a development of ideas which were first presented at the 1980 World Congress of the Econometric Society [**10**] and which were subsequently extended in [**11**]. Deborah Blood and Christopher Sims deserve my thanks for their comments on this work. Alison Gold deserves special mention and thanks for her help in performing the computations reported in Section 5. It is a pleasure also to thank Glena Ames and Lydia Zimmerman for their time and skill in preparing the typescript.

The research reported in this paper was supported by the Cowles Foundation and by the NSF under Grant Number SES 8007571.

tions established in the 1950's and 1960's with an attempt to improve certain features of the design and efficiency of the experiments, together with the means by which the results of the experiments are characterized. The latter have grown in usefulness with advances in computer technology which have, in particular, facilitated numerical integration on a much larger scale in recent years.

It is the purpose of the present article to suggest a new approach to small sample theory that allows for a convenient integration of analytical, experimental, and purely numerical directions of research. The approach centers on a flexible technique of approximating distributions which was first introduced by the author in **[10]**; and it brings into play a new family of extended rational approximants (ERA's) which are designed to accommodate information from sources as diverse as the following: (i) exact analytical knowledge concerning the distribution, its moments or its tail behavior; (ii) alternative approximations based on crude asymptotic theory or more refined asymptotic series; (iii) purely numerical data arising from numerical integrations of moments or numerical evaluations at certain isolated points in the distribution; and even (iv) soft quantitative information of the Monte Carlo variety. The technique itself is not based on an asymptotic series expansion in terms of the sample size or any other parameter. This means that accurate approximations can be obtained even in very small samples. As a result, the technique has its greatest advantage in cases where existing methods based on asymptotic approximations run into difficulty.

The idea that underlies the new technique is very simple. It is motivated by the observation that, in spite of the complex analytic forms of many of the exact pdf's presently known for econometric statistics, when we do turn around and obtain numerical tabulations or graphical plots of the densities we typically end up with well behaved, bounded, continuous functions that tend to zero at the limits of their domain of definition. The form of these pdf's strongly suggests that we should be able to get excellent approximations to them in the class of much simpler functions and certainly without the use of multiple infinite series. We need to deal with approximating functions (or approximants as they are called) that are capable of capturing the stylized form of a density; in particular, we want the approximant to be able to go straight for long periods in a direction almost parallel to the horizontal axis and yet still be able to bend, quite sharply if necessary, to trace out the body of the distribution wherever it is located. One class of functions that seems particularly promising in this respect, as well as being simple in form, are rational functions. Even low degree rational functions can go straight for long periods and then bend quite sharply. In this, of course, they are very different from low degree polynomials whose graphs typically display a distinct roly-poly character.

With these promising features rational functions are well suited to the task of approximating a wide class of continuous distributions. Their own intrinsic qualities as approximants are enhanced by the use of an extended family of rational functions which exploit analytic information about the true distribution in the formation of a suitable coefficient function. The coefficient function is a vehicle for importing a relevant but relative analytic structure to the approximant, whose performance is then fine tuned by the shape of the rational function itself. It is this extended family of approximants that we call ERA's.

In [**10**] the author addressed the basic theoretical issues of existence and convergence of best rational approximants to continuous pdf's and demonstrated a practical method of constructing a good rational approximant from limited information about the true distribution. The method in [**10**] is based on the idea of working from local Taylor series approximations at certain points of the distribution toward a global approximation which will perform well in the whole domain over which the distribution is defined, while retaining the good performance of the Taylor series approximations in the immediate locality of the points of expansion. This is, in part, achieved by the use of multiple-point Padé approximants which are rational functions constructed so as to preserve the local Taylor series behavior of the true pdf at certain points to as high an order as possible. The points selected for local expansion in the application reported in [**10**] were simply the origin and the tails. Such local expansions can, in fact, be obtained from information about the characteristic function of the distribution so that direct knowledge even of the local behavior of the true pdf is not necessary for the application of the technique. The final step in the method involves modifications to the crude multiple-point Padé approximant which are designed to improve its global hehavior. This may involve the removal of unwanted zeroes and poles which occur in the bridging region between the points of local expansion.

The present article extends the theory of rational approximation given in [**10**] and reports a numerical application of ERA's in a time series context which may be of some interest to practitioners. The theory of [**10**] is recast in terms of $[n/m]$ approximants where numerator and denominator polynomials are not necessarily of equal degree. A function analytic proof is provided of the convergence result in [**10**], showing that uniformly close approximation by ERA's is possible for a general class of continuous distributions. Error bound, characterization and uniqueness theorems are also established for best ERA's. As in the classical theory of approximation [**1**], these results help to identify best approximants in applications and indicate the extent and nature of possible improvements in preliminary trial cases. However, some interesting divergences with the classical mathematical theory arise in the alternation result which characterizes the best ERA. This and other theoretical developments are presented in Section 3 of the paper. Section 4 describes some useful modifications to preliminary approximants of the Padé class, which extend the suggestions made earlier in [**10**]. The application is reported in Section 5 and deals with the noncircular serial correlation coefficient. It is hoped that this example will illustrate how ERA's can build on the strengths of a variety of asymptotic methods and successfully temper with data from other sources what would otherwise be rigid analytic formulae. Future extensions of this work and additional applications of ERA's are outlined in Section 6.

2. A FAMILY OF ERA'S

In selecting a suitable class of approximants, the requirements we need to take into account are largely dictated by the shape of the true function and the interval over which the approximation is to be used. As in **[10]** our discussion will concentrate on probability densities that belong to $C_0[-\infty, \infty]$, the class of continuous, positive valued functions that vanish at $\pm\infty$. This covers a very wide class of continuous distributions and, although further generalizations of the methods we present are possible, they will not be pursued here. Rational functions are themselves promising candidates in their capacity to capture the various shapes of distributions within this class. But their capabilities are almost always enhanced by the use of a well chosen coefficient function which serves to provide a relevant analytic scaffolding for the rational function to do its work.

These considerations lead us to make explicit the following class of approximating functions:

DEFINITION (A Family of ERA's): We define *the class $G(m,n)$ of extended rational approximants of maximal degrees m and n* as the class of functions R_{mn} of the form

$$R_{mn}(x;s,\gamma) = s(x)\frac{P_m(x)}{Q_n(x)} = s(x)\frac{a_0 + a_1x + \cdots + a_mx^m}{b_0 + b_1x + \cdots + b_nx^n}, \qquad -\infty < x < \infty, \quad m \leq n, \tag{1}$$

where (i) $s \in C_0[-\infty, \infty]$; (ii) the numerator and denominator are reduced to their lowest degree by the cancellation of identical factors; (iii) m and n are even integers with $m \leq n$; (iv) $\gamma' = (a_0, a_1, \ldots, a_n, b_0, b_1, \ldots, b_n) \in \Gamma$, the parameter space, which is defined as the following subset of $n+m+2$ dimensional Euclidean space: $\Gamma = \{\gamma : \sum_{i=0}^{n} b_i^2 = 1,\ Q_n(x) > 0 \text{ for all } x \in (-\infty, \infty)\}$.

This definition of the family $G(m,n)$ of ERA's of the form (1) extends that of **[10]** and allows for fractions with numerator polynomials of any degree up to m and denominator polynomials of any degree up to n. The normal approximant[2] $R_{mn}(x)$ in this family will involve polynomials $P_m(x)$ and $Q_n(x)$ of degrees m and n, respectively. This terminology is to be distinguished from what is called an abnormal, degenerate, or defective approximant in which the numerator and denominator polynomials are of lower than prescribed degree. Since we will be dealing with probability densities in $C_0[-\infty, \infty]$, it is natural to set m and n as even integers: good global approximation over the entire interval $[-\infty, \infty]$ within $G(m,n)$ will always require n even to exclude poles and m even also if we

[2] In what follows and where there is no risk of ambiguity we will simplify the representation of the rational fraction (1) by using the notation $R_{mn}(x) = R_{mn}(x; s; \gamma)$.

wish to exclude negative probabilities. Similar conditions apply in the case of defective approximants.

The coefficient function $s(x)$ in the ERA (1) is a vehicle by which additional information about the true density can be readily embodied in the approximant. This can be soft or hard quantitative information. Simple examples of the former are: (i) the information already explicit in $s(x)$, namely that pdf > 0 and $\to 0$ as $|x| \to \infty$; or (ii) nonparametric probability density estimates of pdf(x) obtained from Monte Carlo experimental data. Examples of the latter are: (i) knowledge that pdf(x) has moments up to a certain order, suggesting a specification of $s(x)$ with the same tail behavior; or (ii) knowledge that pdf(x) takes a simple primitive form in an important leading case; or (iii) saddlepoint or Edgeworth approximations to pdf(x) suitably modified if necessary to ensure that these are everywhere positive and vanish at infinity; or (iv) the crude asymptotic approximation to pdf(x). Reference **[10]** provides a detailed illustration of the second type of hard quantitative information. An illustration of the use of the third and fourth types will be given later in the present paper in Section 5.

The normalization condition $\sum_{i=0}^{n} b_i^2 = 1$ on the parameter space Γ eliminates the redundancy that results from the multiplication of $P_m(x)$ and $Q_n(x)$ by an arbitrary constant. This form of the condition is useful for the theoretical development, as in **[10]**; but other normalizations such as $b_0 = 1$ are often more useful in applications, simplifying as they do the analytic equations needed to be solved to find the rational coefficients. Note that the alternative condition $b_0 = 1$ ensures that the ERA (1) is always well behaved as x passes through the origin. In distribution function rather than density approximation it is additionally convenient to set $m = n$, $a_n = b_n$, and replace $s(x)$ in (1) by a primitive cdf of the form $S(x) = \int_{-\infty}^{x} s(t)\,dt$ appropriately weighted. These modifications ensure desirable behavior in a cdf approximant at the limits of its domain.

The above definition of $G(m,n)$ does not exclude the possibility of approximants which possess zeroes and which become negative over part of the interval $[-\infty, \infty]$. This possibility can be eliminated, of course, by the further requirement that $P_m(x) > 0$, parallel to the pole elimination condition on $Q_n(x)$. But such a requirement seems unnaturally restrictive. For it excludes functions which become negative only on the extreme tails and which may, nonetheless, be excellent approximants over a wide domain. Since densities in $C_0[-\infty, \infty]$ vanish at infinity and the best ERA displays the error alternation property (Theorem 4 of Section 3), small negative probabilities in the distant tails of a good approximant are not uncommon. While these can be removed, some deterioration in the quality of the approximant is to be expected elsewhere in the distribution. In making such a choice an element of judgment is required concerning the importance of errors of different magnitude in different parts of the distribution. Similar consideration may also apply in practice to the occurrence of poles, provided these lie outside the relevant domain in which the approximant is to be used. However, for the theoretical development that follows, the latter are excluded in our definition of the class $G(m,n)$ by the condition $Q_n(x) > 0$.

3. BEST UNIFORM APPROXIMATION BY ERA'S

In its general form, we consider the problem of approximating a pdf in $C_0[-\infty, \infty]$ by an ERA of the form (1). We take the domain of approximation to be the infinite interval $(-\infty, \infty)$ and, as in **[10]**, we use the uniform norm $\|e(x)\| = \sup_x |e(x)|$. Other choices of norm are certainly possible and will generally lead to different best approximations, where they exist. However, for accurately approximating a given pdf over a wide interval the choice of the uniform norm seems most appropriate. Moreover, with the modifications that were discussed in Section 2, ERA's of the same general form as (1) comprise a suitable class of direct approximants to continuous cdf's as well as probability densities.

As defined above, the problem of approximation comes close to the framework of rational approximation in the space $C[-\infty, \infty]$ of all continuous functions f over $(-\infty, \infty)$ for which $\lim_{x\to\infty} f(x) = \lim_{x\to-\infty} f(x)$ and the limit is finite. Some aspects of rational approximation in $C[-\infty, \infty]$ have received attention in the classical mathematical literature of approximation theory,[3] notably by Walsh **[18, 19]**, Achieser **[1]**, and Timan **[17]**. Of these authors, Achieser's treatment comes closest to our own. He develops a theory of uniform approximation applicable to a restricted version of our class $G(m, n)$. His restriction operates on the class of admissible coefficient functions $s(x)$, which are required to satisfy

$$\lim_{x\to\infty} s(x)x^k = \lim_{x\to-\infty} s(x)x^k \neq 0 \tag{2}$$

where $k = m - n$ (our notation). In our context, (2) is too restrictive. For example, we may wish to impose rather different behavior on $s(x)$ at infinity. Thus, if we know that as $|x| \to \infty$ $\text{pdf}(x) = 0(|x|^{-\mu})$ with $\mu > 0$, or $0(e^{-\alpha|x|^\beta})$ with $\alpha, \beta > 0$, it makes sense to incorporate similar behavior in the coefficient function $s(x)$ and then set $m = n$ in the ERA (1). These possibilities are excluded by (2). Moreover, (2) provides an essential simplification of the theory in that infinity becomes an allowable extreme point in the approximation error. Under this simplification the theory of best uniform approximation in $C[-\infty, \infty]$ is essentially the same as it is for the space of continuous functions $C[a, b]$ over a finite interval $[a, b]$. In particular, all the major results apply as for a finite interval. When we relax (2) and require only that $s \in C_0[-\infty, \infty]$ as in our definition of (1), this is no longer the case. While the existence, uniqueness and convergence results apply as for finite intervals, the usual characterization theorem (Achieser [**1**, p. 55], Meinardus [**6**, pp. 161–162], Rice [**15**, p. 80]) fails. This theorem tells us that a necessary and sufficient condition for the best uniform approximant is that its error curve oscillate a specific number of times, which depends on the degree of the approximant. Theorem 3 below shows that for rational approximation of pdf(x) over $[-\infty, \infty]$ within the class (1), this condition is sufficient but not necessary.

[3] A general discussion of this literature and its bearing on the problem of approximation considered here was given by the author in an earlier version **[11]** of this paper.

Thus, while the classical work of Achieser in the mathematical literature is very relevant to our problem of best approximation to pdf(x) within the class of approximants (1), we cannot completely rely on this work in the development of our own theory. Moreover, with regard to the important problems of characterizing the best approximant, our theory involves an important departure from existing results. The following theorems form the basis of this theory. Proofs, together with some subsidiary technical material, are given in the Appendix in order to facilitate the reading of the paper by those interested in its main ideas.

THEOREM 1 (Existence): *If* pdf$(x) \in C_0[-\infty, \infty]$, *then there exists a best uniform approximant to* pdf(x) *in the class of ERA's defined by* (1).

This theorem ensures that, given pdf(x) and $s(x)$ in $C_0[-\infty, \infty]$, there exists a set of parameters γ^* and a corresponding ERA $R'_{nn}(x; s, \gamma^*)$ in the class (1) for which

$$\|\mathrm{pdf}(x) - R'_{mn}(x; s, \gamma^*)\| = \inf_{\gamma \in \Gamma} \|\mathrm{pdf}(x) - R_{mn}(x; s, \gamma)\|. \tag{3}$$

The prime in $R'_{mn}(x)$ is used to distinguish this rational fraction from a limiting function $R_{mn}(x; s, \gamma^*)$ which does not necessarily belong to the class (1) (it may not even be continuous) because the set Γ is not closed.[4]

THEOREM 2 (Uniformly Close Approximation by ERA's): *Suppose* pdf$(x) \in C_0[-\infty, \infty]$ *and let* $\epsilon > 0$ *be given. Then there exists an ERA* R_{mn} *of the form* (1) *for which* $\|\mathrm{pdf}(x) - R_{mn}(x)\| < \epsilon$.

COROLLARY (Convergence [**10**]): *If* pdf$(x) \in C_0[-\infty, \infty]$ *and if we define*

$$E(n; s) = \|\mathrm{pdf}(x) - R'_{nn}(x; s, \gamma^*)\| \tag{4}$$

where $R'_{nn}(x; s, \gamma^*)$ *is the best uniform approximant to* pdf(x) *in the class of ERA's defined by* (1) *with* $m = n$, *then*

$$\lim_{n \to \infty} E(n; s) = 0. \tag{5}$$

Theorem 2 shows that, for any choice of density function in $C_0[-\infty, \infty]$, there is an arbitrarily close rational approximant in the class (1). Moreover, the best approximant of the form R_{nn} converges to the given pdf uniformly over $(-\infty, \infty)$ as $n \to \infty$. Note that neither of these results gives an order of magnitude on the error of approximation as $n \to \infty$. Some theorems on error magnitudes and rates of convergence are available in the approximation theory literature but these

[4] This problem is discussed in detail with an example in [**10** and **11**].

apply, in the main, to special functions. To take two examples: (i) it is known [7] that whereas the error on an nth degree polynomial approximation to $|x|$ over $[-1, 1]$ has order n^{-1}, the error on a rational approximant of type $R_{nn}(x; s(x) = 1)$ has order at most $\exp(-n^{1/2})$; (ii) it is also known [**4**] that the error on a rational approximant of type $R_{nn}(x; s(x) = 1)$ to the function $\exp(-|x|)$ over $(-\infty, \infty)$ has order at most $\exp(-\alpha n^{1/2})$ for some constant $\alpha > 0$. Both of these rather specialized results on rational approximants suggest that, at least for certain classes of functions, the rate of convergence of polynomial approximants can be dramatically improved by the use of rational functions with numerator and denominator of equal degree. These results corroborate the numerical experience discussed in [**10**] which suggests that rational functions of this type tend on the whole to provide better approximations than those for which the numerator and denominator polynomials differ markedly in degree.

THEOREM 3 (Error Bound): *Let the* ERA *of the class* (1)

$$R_{mn}(x; s, \gamma) = s(x)\frac{P_m(x)}{Q_n(x)} = s(x)\frac{a_0 + a_1 x + \cdots + a_{m-\mu}x^{m-\mu}}{b_0 + b_1 x + \cdots + b_{n-\nu}x^{n-\nu}} \tag{6}$$

be in its lowest terms with no common factors in $P_m(x)$ *and* $Q_n(x)$, *with* $a_{m-\mu} \neq 0$, $b_{n-\nu} \neq 0$ *and* $0 \leq \mu \leq m$, $0 \leq \nu \leq n$. *We set* $d = \min(\mu, \nu)$ *and* $N = m + n - d + 2$. *If at the consecutive points* $x_1 < x_2 < \cdots < x_N$ *in the interval* $[-\infty, \infty]$ *we have*

$$\mathrm{pdf}(x_i) - R_{mn}(x_i) = (-1)^i \lambda_i \tag{7}$$

where all λ_i *have the same sign and are all different from zero, then*

$$\inf_{\gamma \in \Gamma} \|\mathrm{pdf}(x) - R_{mn}(x; s, \gamma)\| \geq \min\{|\lambda_1|, |\lambda_2|, \ldots, |\lambda_N|\}. \tag{8}$$

The same result is true of $R_{mn}(x; s, \gamma) \equiv 0$ *in* (6), *in which case we take* $N = m + 2$.

This theorem can be used to find a lower bound for the deviation of the best approximation. It suggests that when the error curve oscillates a sufficient number of times (which will usually be $N = n + m + 2$, with the "defect" in the rational fraction $d = 0$) and the extreme values of the oscillation are close in absolute value, the rational fraction is also close to the best approximant in that class. When the error curve oscillations differ markedly in magnitude or when there are insufficient oscillations, there will usually be considerable scope for improvement in the approximant. This idea is formalized in the following result.

THEOREM 4 (The Equioscillation or Error Alternation Theorem): *Suppose* $R_{mn}(x; s, \gamma)$ *is an ERA of the class* (1), *having the same form as* (6) *above. Let* $d = \min(\mu, \nu)$ *as in Theorem* 3 *and let*

$$h = \|\mathrm{pdf}(x) - R_{mn}(x; s, \gamma)\| > 0. \tag{9}$$

We define

$$(10) \qquad e(x) = \mathrm{pdf}(x) - R_{mn}(x; s, \gamma)$$

and we call the number of consecutive points of the interval $[-\infty, \infty]$ *at which* $e(x)$ *takes on its maximum value* h *with alternate changes of sign "the number of alternations of* $e(x)$*."*

Part A—Necessary Conditions: If $R_{mn}(x; s, \gamma)$ *is the best uniform approximant to* pdf(x), *then either* (i) $d = \nu$ *(numerator most degenerate) and the number of alternations of* $e(x)$ *is at least* $N = n + m - d + 2$; *or* (ii) $d = \mu$ *(denominator most degenerate) and the number of alternations of* $e(x)$ *is at least* $N = n + m - \mu - \nu + [\nu]_e + 2$; *where* $[\nu]_e$ *denotes the largest even integer less than or equal to* ν; *or* (iii) *if* $R_{mn}(x) \equiv 0$ *then the number of alternations of* $e(x)$ *is at least* $N = m + 2$.

Part B—Sufficient Conditions: If $R_{mn}(x; s, \gamma)$ *has the form* (6) *and the number of alternations of the error* $e(x)$ *is at least* $N = n + m - d + 2$, *then* $R_{mn}(x; s, \gamma)$ *is the best uniform approximant to* pdf(x) *in the class* (2).

This theorem characterizes the best approximant. Specifically, Part B tells us that if the error curve takes on its maximum value with alternate changes of sign at least $N = n + m - d + 2$ times then this is sufficient to ensure that the rational function (6) is a best approximant. In the classical characterization theorem [**1**, pp. 55–57; **6**, pp. 161–162], this condition is also necessary. It is an interesting feature of the present problem that the condition is no longer necessary. This divergence from the classical theory only occurs when the best approximant has a degenerate denominator as in Part A(ii). (Readers may refer to [**12**] for a detailed discussion and a counterexample to the classical theorem.) In the usual nondegenerate situation where the best ERA in the class (1) has numerator and denominator polynomials of full degree (with $\mu = \nu = d = 0$) the standard alternation theorem applies with the number of alternations being at least $N = n + m + 2$.

Finally, we have the following result which establishes the uniqueness of the best ERA.

THEOREM 5 (Uniqueness): *The best uniform approximant to* pdf(x) *in the class of ERA's defined by* (1) *is unique when reduced to its lowest terms.*

4. PRACTICAL ISSUES IN THE CONSTRUCTION OF ERA'S

The first steps in the practical implementation of our approach require important elements of judgment in the following three areas: (i) choice of the coefficient function $s(x)$; (ii) selection of the degree of the ERA $R_{mn}(x)$; and (iii) determination of the coefficients of the polynomials in $R_{mn}(x)$. The problem is one of constructive functional approximation within the general family of approximants defined in (1). 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 information and reliable numerical or experimental data about the distribution as possible. This will directly affect the choice of $s(x)$ and the prescribed degree of $R_{mn}(x)$. As argued in [**10** and **11**], important theoretical and practical considerations suggest a specialization of the family of ERA's to that in which numerator and denominator polynomials are of the same degree (that is, $m = n$). Leading case analyses such as those developed in [**13**] and applied in [**10**] will often lead to a suitable choice of $s(x)$; and a variety of other choices involving both soft and hard quantitative information were discussed earlier in Section 2. Knowledge of the local behavior of the true distribution in certain regions can be used to determine the polynomial coefficients in the ERA, which will then magnify or attenuate as appropriate the shape of the leading coefficient function $s(x)$. Local information about the distribution may take the form of Taylor expansions at certain points, evaluations by numerical integration or estimates of the function values obtained from Monte Carlo simulations. Some operational guidelines for this aspect of the constructive process were laid out by the author in the earlier article [**10**] and were based on modified multiple-point Padé approximants. The ideas given there will be extended in the discussion that follows to deal with nearly defective ERA's, a problem not considered in [**10**].

It is sufficient for our purposes to take a multiple-point Padé approximant whose coefficients have been determined by solving the general system of equations given by (43) in [**10**]. We use the following notation:

$$R_{mn}(x) = s(x)\frac{P_m(x)}{Q_n(x)} = s(x)[m/n](x) \tag{11}$$

for rational functions obtained in this way which have the same general form as members of the family (1) but which may possess unwanted poles to the real axis. To illustrate, we consider a case which arises later in the application of Section 5. Here, we attempt to approximate a distribution which is symmetric about the origin. We correspondingly select $s(x)$ and $[m/n](x)$ to be even functions of x, the latter defined explicitly in terms of x^2 rather than x. Now consider a case such as this in which $m = n = 4$ and $[4/4](x)$ takes the form

$$[4/4](x) = \frac{\sum_{i=0}^{4} a_i x^{2i}}{\sum_{i=0}^{4} b_i x^{2i}} = \frac{a_4(x^2-\gamma_1)(x^2-\gamma_2)(x^2-\gamma)(x^2-\bar{\gamma})}{b_4(x^2-\delta_1)(x^2-\delta_2)(x^2-\delta)(x^2-\bar{\delta})} \tag{12}$$

where (γ_1, γ_2) and (δ_1, δ_2) are the real zeroes of the numerator and denominator polynomials considered as functions of $z = x^2$. Zeroes and poles of $[4/4](x)$ arise for real values of x when $\gamma_i > 0$ and $\delta_i > 0$ for either $i = 1, 2$.

We recall from Section 3 that the best ERA is defective when it is the same as that for a lower maximal degree. In practice, nearly defective ERA's tend to show up in cases such as (12) when after many trials the numerator and denominator polynomials are found to have a nearly identical factor. This induces a difficulty in the approximant when the factors produce a real zero and pole, which will then be in close proximity to each other. The resulting discontinuity in the approximant can most simply be removed by cancellation of the troublesome factors. Thus, if $\gamma_1 \simeq \delta_1$ in (12) with both numbers positive, we would obtain in this way the new function:

$$(13) \qquad [3/3](x) = \frac{a_4(x^2-\gamma_2)(x^2-\gamma)(x^2-\bar{\gamma})}{b_4(x^2-\delta_2)(x^2-\delta)(x^2-\bar{\delta})}.$$

A more sophisticated modification would be obtained by the use of the following approximant:

$$(14) \qquad [4/4]^*(x) = \frac{a_4(\epsilon x^2+\gamma_1)(x^2-\gamma_2)(x^2-\gamma)(x^2-\bar{\gamma})}{a_4(\eta x^2+\delta_1)(x^2-\delta_2)(x^2-\delta)(x^2-\bar{\delta})}$$

in which ϵ and η are small positive quantities selected so that the additional factor will behave substantially like the constant function γ_1/δ_1 over a wide interval symmetric about $x = 0$. This additional modification will be most suitable if $x = 0$ is itself an interpolation point in the original multiple point Padé. Frequently, it will be sufficient to set $\epsilon = \eta = 10^{-6}$ (or thereabouts) but the separate parameterization of ϵ and η may be helpful in some cases, particularly if we wish to further improve the performance of the approximant in the tails.

The theoretical results of Section 3 can be used to sharpen our understanding of the behavioral characteristics of modified Padé approximants such as (13) and (14). Thus, if our objective is an ERA that has reliable global as well as local behavior, we will usually have to give up some degree of local performance in a crude Padé to ameliorate the global characteristics of the approximant. The existence result shows that this is possible. It is achieved in practice by trading off near perfect fits in certain regions of the distribution in exchange for much better fits in other regions than are attained by crudely constructed Padé's of the same maximal degree. The equioscillation theorem formalizes this process. As stated, it does not rigorously apply to always positive approximants but the idea of trading off error magnitudes in different regions to improve global behavior will still ordinarily apply. Moreover, as argued in Section 3, small negative probabilities (often around 10^{-6}) in the extreme tails may not be unacceptable if the gains elsewhere are sufficiently important. In practice, troublesome singularities rather than negative probabilities present the real difficulty with Padé approximants. The constructive modifications suggested above, together with those given earlier in [**10**], provide an effective (if *ad hoc*) procedure for tackling this problem and they bring the approximants within the general arena of ERA's in the family defined by (1) where the tradeoffs that underlie the strict equioscillation property apply.

5. AN APPLICATION TO THE SERIAL CORRELATION COEFFICIENT

The model we use is the autoregression

$$(15)\qquad y_t = \alpha y_{t-1} + u_t \qquad (t = \ldots, -1, 0, 1, \ldots)$$

in which the u_t are i.i.d. $N(0, \sigma^2)$. We look, in particular, at the distribution of the least squares estimator of α in (15) given by $\hat{\alpha} = (\sum_{t=1}^{T} y_{t-1}^2)^{-1}(\sum_{t=1}^{T} y_t y_{t-1})$. This statistic is a noncircular serial correlation coefficient, which can be used to estimate the correlation between consecutive observations in an ordered sample. However, its distribution is supported over the entire interval $(-\infty, \infty)$ rather than $(-1, 1)$ and approximations to the distribution which remain adequate in the tails, at least for moderate values of the sample size T, are difficult to obtain. Recent work in this and related contexts has been done by Reeves **[14]**, Goldsmith **[5]**, Phillips **[8, 9]**, and Evans and Savin **[3]**. Our application of the methods of this article concentrate on the case where $\alpha = 0$ and $T = 10$ but can easily be extended to deal with other cases of importance in econometrics (such as the random walk with $\alpha = 1$). Such extensions together with the treatment of a variable sample size are currently under way and will be reported in later work.

In the first stage of our procedure we need to select the coefficient function $s(x)$. Several interesting choices are presented by our analytic knowledge of the distribution of $\hat{\alpha}$. Of these, the most obvious are the following: (i) the asymptotic normal:

$$(16)\qquad \mathrm{an}(x) = (T/2\pi)^{1/2}\exp\{-Tx^2/2\}, \qquad -\infty < x < \infty;$$

(ii) the noncircular saddlepoint approximation **[9]**:

$$(17)\qquad \mathrm{sp}(x) = \frac{(T-3)\Gamma((T-1)/2)(1-x^2)^{(T+1)/2}}{2\pi^{1/2}\Gamma(T/2)(1-2x^2)^{1/2}}, \qquad -2^{-1/2} < x < 2^{-1/2};$$

(iii) the circular saddlepoint approximation **[2]**:

$$(18)\qquad \mathrm{lp}(x) = \left[B\left(\frac{1}{2}, \frac{T+1}{2}\right)\right]^{-1}(1-x^2)^{(T-1)/2}, \qquad -1 \le x \le 1;$$

(iv) the Edgeworth approximation to $0(T^{-1})$ **[9]**:

$$(19)\qquad \mathrm{ed}(x) = \mathrm{an}(x)\left[1 + \frac{1}{4T}(1 + 2Tx^2 - T^2x^4)\right]; \qquad -\infty < x < \infty.$$

As distinct from the examples given in **[8** and **10]**, the asymptotic normal approximation (16) provides a reasonable overall picture of the exact density (Figure 1). This performance is helped by the symmetry of the exact distribution and the fact that $\hat{\alpha}$ is unbiased in the special case $\alpha = 0$. Note, however, that tail area probabilities are distorted upwards as shown in Figure 2. Thus, for a one

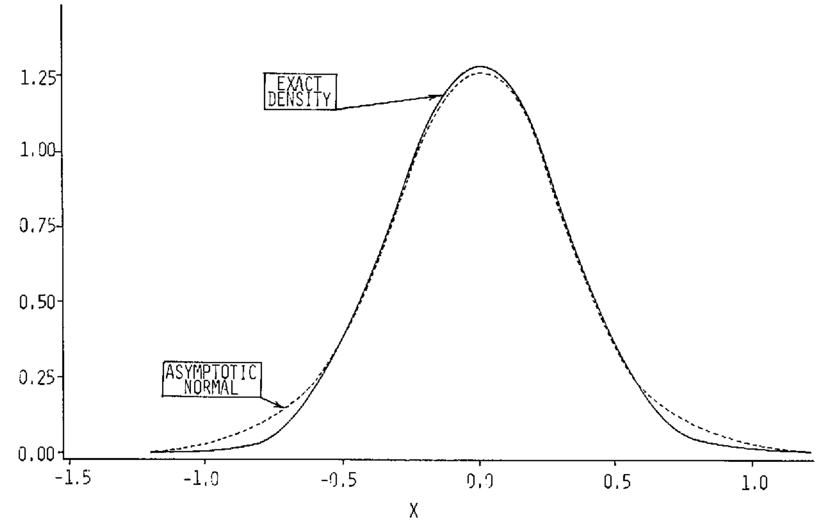

FIGURE 1—Distribution of the serial correlation coefficient in the null case.

sided test of $\alpha = 0$ (no serial correlation) at the nominal size of 5 per cent, the asymptotic distribution gives a critical value of 0.52, whereas the true size of the test with this critical value is 3.8 per cent. When the nominal size of the asymptotic test is 1 per cent (with a critical value of 0.66) the true size of the test is 0.4 per cent.

Before they can be used in the construction of ERA's, the saddlepoint (SP) approximations (17) and (18) need modification to ensure that they exist and are positive over the entire real line. This can be simply and satisfactorily achieved

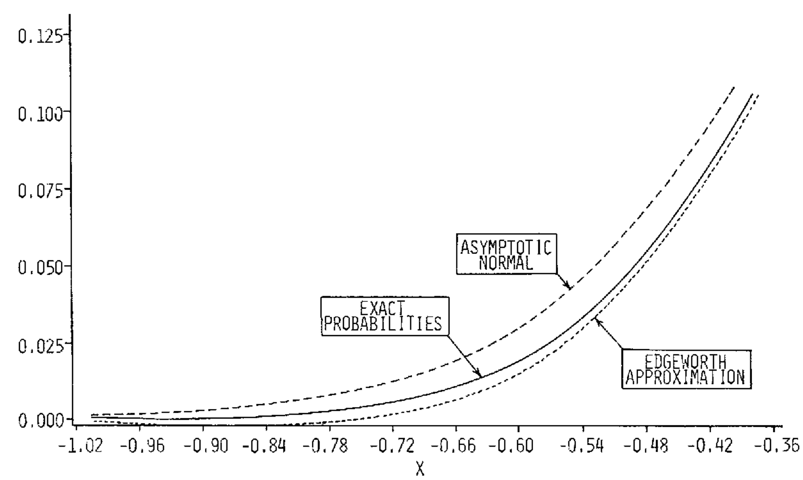

FIGURE 2—Tail probabilities of the serial correlation coefficient.

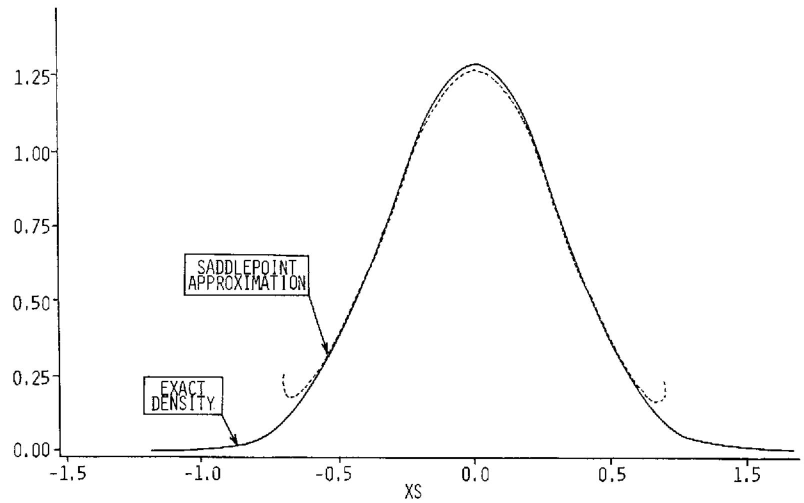

FIGURE 3—Distribution of the serial correlation coefficient.

by splicing the SP with the tails of a t distribution of the form $c(1 + x^2/d)^{-(T+1)/2}$ where the parameters are chosen to ensure an osculating interpolation at a suitable point. We choose to work with the noncircular SP approximant (17), largely because of the challenge of dealing with its singularity at $x = \pm 2^{-1/2}$ (see Figure 3).

The Edgeworth approximation (19) is shown against the exact density in Figure 4. This is the best of the crude approximants (16)–(19) and only encounters difficulties in the tail area, where the density is negative for $|x| > 0.86$. This

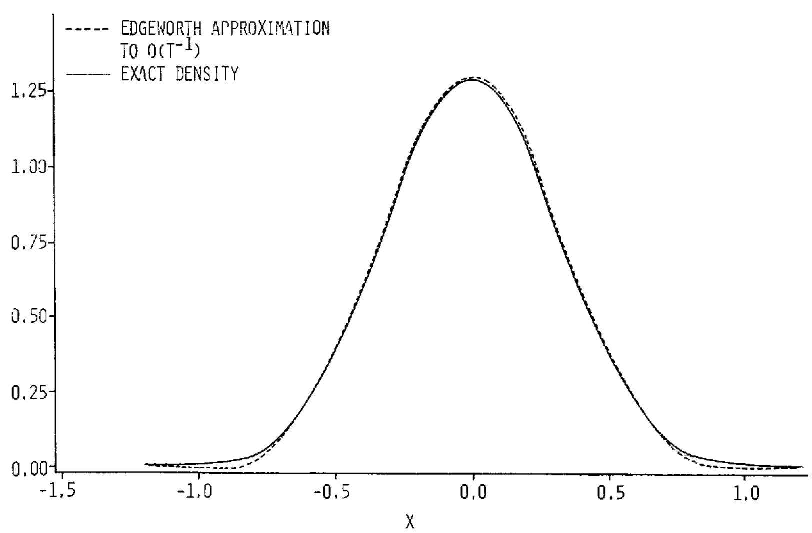

FIGURE 4—Distribution of the serial correlation coefficient.

leads to the underestimation of tail probabilities shown in Figure 2. In view of the negative tail area we decided to design an ERA which would correct for the absolute error in the Edgeworth approximation leading to an approximant of the form:

$$(20)\qquad R_{nn}(x) = \mathrm{ed}(x) + \mathrm{an}(x)[n/n](x) = \mathrm{an}(x)\left\{1 + \frac{1}{4T}(1 + 2Tx^2 - T^2x^4) + [n/n](x)\right\}.$$

Here the asymptotic distribution $\mathrm{an}(x)$ is used as the coefficient function in (1), so that the ERA is, in fact, an error corrected Edgeworth approximation.

The promising performance of multiple point $[n/n]$ Padé approximants in **[10** and **11]** encouraged us to use similar methods again. We set $n = 4$ and selected $2n + 1 = 9$ points of interpolation with the exact density to fit the rational coefficients. The exact density was computed by an extension of Imhof's procedure which yields the density ordinates directly and which achieves economies by the use of symmetric matrix eigenvalue routines.

In view of the symmetry of the exact distribution for $\alpha = 0$ we defined the polynomials of $[n/n](x)$ explicitly in terms of x^2 as in (12). Figure 5 shows the error curve of a [4/4] Padé approximant which interpolates at the points {0.02, 0.10, 0.24, 0.42, 0.49, 0.58, 0.70, 0.90, 1.30}. This has the form

$$(21)\qquad R^a_{4,4}(x) = \mathrm{mp}(x)\,\frac{a_4(x^2-\gamma_1)(x^2-\gamma_2)(x^2-\gamma)(x^2-\bar{\gamma})}{b_4(x^2-\delta_1)(x^2-\delta_2)(x^2-\delta)(x^2-\bar{\delta})}$$

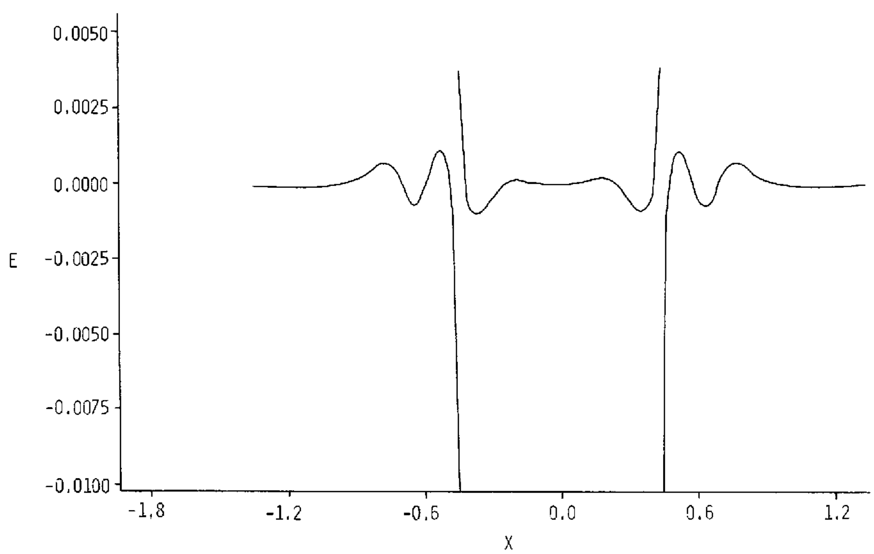

FIGURE 5—Error curve of $R^a_{4,4}(x)$.

where

$$a_4 = 3.053840, \qquad b_4 = -6.628521,$$
$$\gamma_1 = 0.199577, \qquad \delta_1 = 0.199636,$$
$$\gamma_2 = 2.201005, \qquad \delta_2 = -2.198206,$$
$$\gamma, \bar{\gamma} = 0.742363 \pm 0.479389i, \qquad \delta, \bar{\delta} = 0.461869 \pm 0.361182i,$$

and $\mathrm{mp}(x)$ is the SP approximation (17) modified with t distribution tails as earlier discussed. The singularities of (21) at $x = \pm 0.446806$ show up clearly as discontinuities in the graph of the error function $e(x) = \mathrm{pdf}(x) - R_{4,4}(x)$. Elsewhere the approximant performs well.

Examination of (21) reveals that $x^2 - \gamma_1$ and $x^2 - \delta_1$ are nearly identical factors to the fourth decimal place. Several attempts were made to remove the (real) singularities by selection of alternative points of interpolation. Each of these met with no success and gave rise to an approximant like (21) with nearly identical factors in the numerator and denominator. We, therefore, cancelled these factors as in (13) and computed the error curve for the new function

$$R_{3,3}(x) = \mathrm{mp}(x)\frac{a_4(x^2-\gamma_2)(x^2-\gamma)(x^2-\bar{\gamma})}{b_4(x^2-\delta_2)(x^2-\delta)(x^2-\bar{\delta})} \tag{22}$$

which is shown in Figure 6. $R_{3,3}(x)$ sucessfully bridges over the singularities in (21) without suffering an appreciable decrease in accuracy elsewhere. In fact, (22) delivers nearly three decimal place accuracy over the entire interval

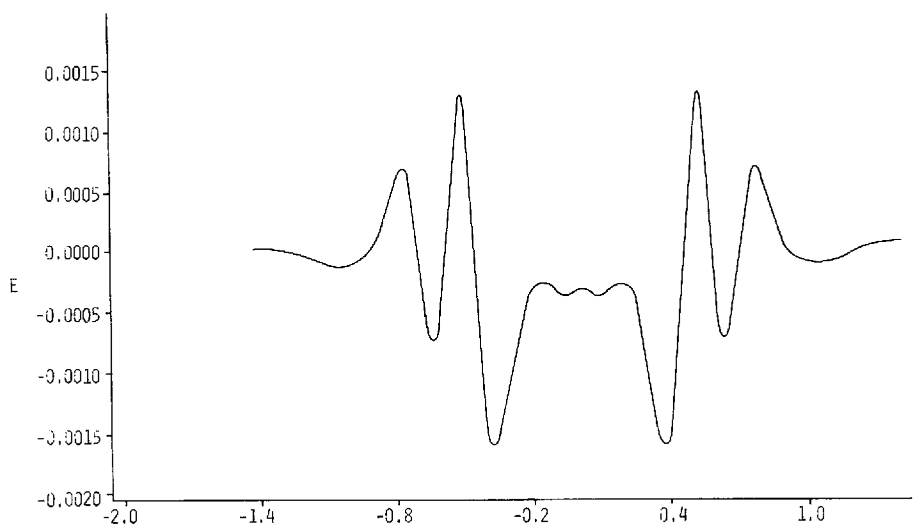

FIGURE 6—Error curve of $R_{3,3}(x)$.

$(-\infty, \infty)$; and its graph is indistinguishable from the exact density on the scale of Figure 1.

We note that the error curve in Figure 6 exhibits six unequal oscillations, two short of the number required for the direct application of the error bound and alternation theorems in Section 3. However, the inequality in the oscillations indicate that there is room for further improvement in this approximant by trading off errors of different magnitudes in different parts of the distribution. Some tinkering with the coefficients of (22) would be necessary to achieve this. Ideally (and if cost were no limitation) this tinkering can be formally embodied in a numerical algorithm of the type prescribed by Meinardus **[6]**. This algorithm relies on arbitrary function and derivative evaluations for the exact density within its iterations and, in large part for budgetary reasons, has not yet been investigated by the author within this context.

Instead, we followed the intuitive argument of Section 4 that led to (14). With a further modification along these lines we obtained

$$R^b_{4,4}(x) = \text{mp}(x)\frac{a_4(\epsilon x^2 + \gamma_1)(x^2 - \gamma_2)(x^2 - \gamma)(x^2 - \bar{\gamma})}{b_4(\epsilon x^2 + \delta_1)(x^2 - \delta_2)(x^2 - \delta)(x^2 - \bar{\delta})} \tag{23}$$

and set $\epsilon = 10^{-7}$. The error curve for (23) is shown in Figure 7. As expected (since $x = 0.02$ was an original point of interpolation in (21)) this modification reduces the error in the neighborhood of the origin with no apparent ill effects elsewhere. The error curve now displays nine oscillations (one short of the number required for the direct application of Theorems 3 and 4). Once again the unequal oscillations reveal the scope for further refinements.

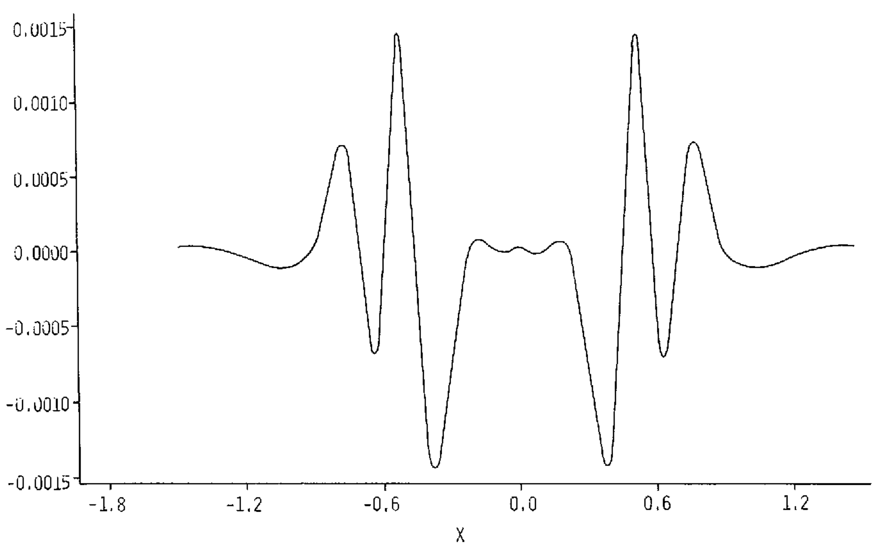

FIGURE 7—Error curve of $R^b_{4,4}(x)$.

In our final application we decided to use the error corrected Edgeworth approximation (20). On our fifth trial we found the following approximant which interpolates to the error on the crude Edgeworth approximation (19) at the points $\{0.04, 0.18, 0.30, 0.44, 0.65, 0.81, 0.91, 1.00, 1.30\}$:

$$R^c_{4,4}(x) = \mathrm{an}(x)\left\{1 + \frac{1}{4T}(1 + 2Tx^2 - T^2x^4) + [4/4](x)\right\} \tag{24}$$

with

$$[4/4](x) = \frac{a_4(x^2-\gamma_1)(x^2-\gamma_2)(x^2-\gamma)(x^2-\bar{\gamma})}{b_4(x^2-\delta_1)(x^2-\delta_2)(x^2-\delta)(x^2-\bar{\delta})} \tag{25}$$

where

$a_4 = 3.393565,$ $\quad b_4 = -1.043281,$

$\gamma_1 = 0.459718,$ $\quad \delta_1 = 2.689876,$

$\gamma_2 = -0.036547,$ $\quad \delta_2 = -0.409958,$

$\gamma, \bar{\gamma} = 0.158796 \pm 0.249000i,$ $\quad \delta, \bar{\delta} = 0.692623 \pm 0.624088i.$

The error curve for (24) is shown in Figure 8. The approximant $R^c_{4,4}(x)$ can be seen to deliver accuracy to four decimal places within one unit over the domain $-1.5 \leq x \leq 1.5$. Inspection of (25) reveals singularities in the extreme tails at $x = \pm 1.640084$. Since the exact density is of the order 10^{-5} in this locality, the singularities can be safely ignored as strictly beyond the domain of interest.

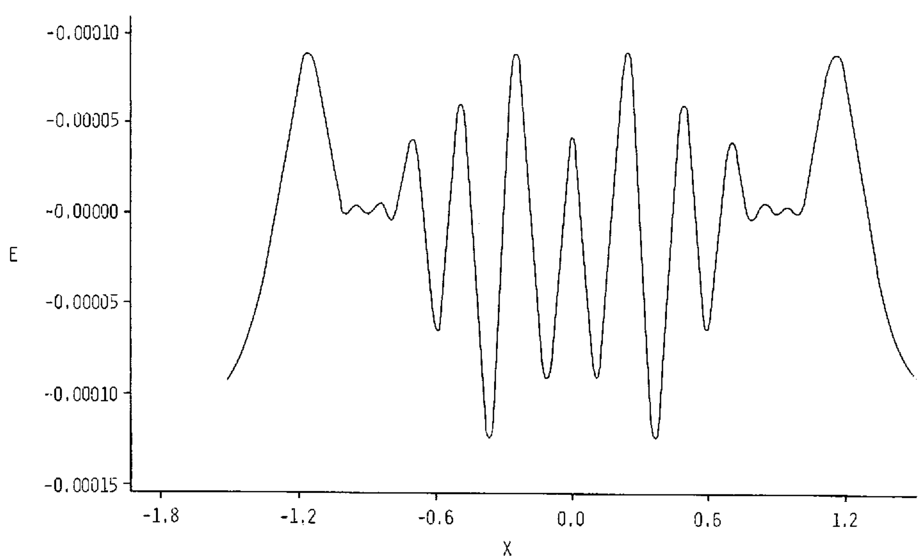

FIGURE 8—Error curve of $R^c_{4,4}(x)$.

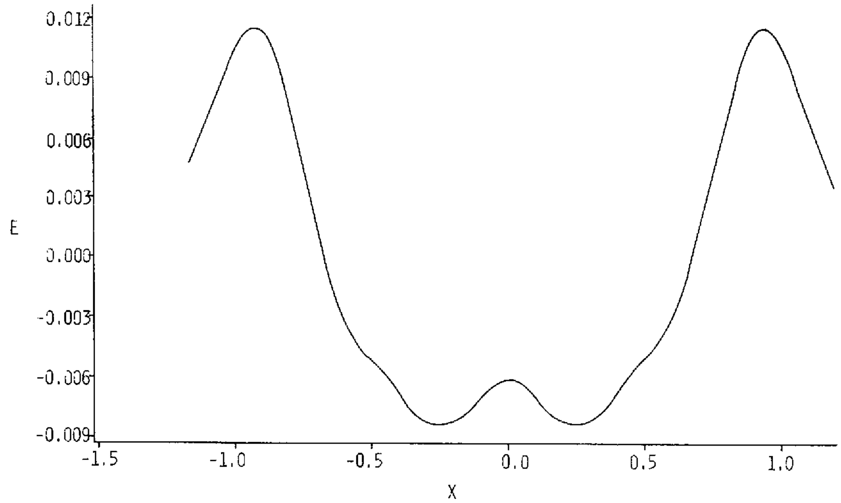

FIGURE 9—Error curve of the Edgeworth approximation to $O(T^{-1})$.

Within the domain $[-1.5, 0]$ the error curve oscillates 14 times. The oscillations are of similar magnitude with the exception of the minor ripples in the error curve around the region $(-1.00, -0.80)$. These were confirmed by calculation to be of order 10^{-6}, corresponding to the accuracy at which the Imhof algorithm truncation error had itself been set. Thus, while improvements in the performance of this approximant are still possible, its accuracy is already approaching that obtained by direct numerical integration. It is noteworthy that this accuracy is achieved at the greatly reduced cost of a few numerical integrations at the isolated points of interpolation. When we account also for the trials leading to the selection of (24) and (25) this still amounts to sizeable economies.

The extent of the improvements attained by $R_{4,4}^c(x)$ can be gauged by comparison of Figure 8 with the error curve of the crude Edgeworth approximation shown in Figure 9.

6. EXTENSIONS

Practical implementation on an appreciable scale of the approach suggested in this article will require general programmable formulae for the coefficients of the ERA's in terms of the relevant model and data set. This will certainly be possible in some of the simpler cases. Some statistical operations will call for direct approximants to the cdf and to tail area probabilities rather than the pdf, in which case the methods suggested in Section 2 will be appropriate. Finally, the technique will have interesting applications beyond those we have discussed here. To take an example from Bayesian inference, the technique offers a convenient mechanism for pooling the information available from analytic study and numerical (Monte Carlo) integration in the characterization of multidimensional posterior distributions. These are some of the many issues that are left for exploration in later work.

Yale University

Manuscript received August, 1980; final revision received August, 1982.

APPENDIX

PROOF OF THEOREM 1: [**10**] proves this theorem for the case $m = n$. The arguments presented there remain essentially unchanged in the more general case of $[m/n]$ approximants and are given in full in [**11**].

The following result which extends the Stone–Weierstrass Theorem to locally compact spaces is useful in the proof of Theorem 2.

LEMMA: *Let X be a locally compact Hausdorff space and let A be a subalgebra of $C_0(X, R)$, the space of all continuous real functions defined on X which vanish at infinity. If A separates points of X and for each point in X contains a function which does not vanish there, then A is dense in $C_0(X, R)$.*

PROOF: See [**16**, pp. 166–167].

PROOF OF THEOREM 2: Set $X = (-\infty, \infty)$ and, with the usual topology, this is a locally compact Hausdorff space. Let $s(x)$ be a positive valued continuous function on $(-\infty, \infty)$ which vanishes at infinity. Consider the functions $s(x)$, $s(x)[1 + x^2]^{-1}$, and $s(x)[1 + (x-1)^2]^{-1}$ defined on $(-\infty, \infty)$. Let A be the set of all functions generated from these primitive members by the following three operations: addition, multiplication by real numbers, and pointwise s-multiplication defined by $a_1 a_2(x) = s(x)\bar{a}_1(x)\bar{a}_2(x)$ for $a_1 = s(x)\bar{a}_1(x) \in A$ and $a_2 = s(x)\bar{a}_2(x) \in A$. We note that A is an algebra of real valued functions on $(-\infty, \infty)$ which vanish at infinity. It is, therefore, a subalgebra of $C_0[-\infty, \infty]$.[5] If $x_1, x_2 \in (-\infty, \infty)$ with $x_1 \neq x_2$, then either $s(x_1) \neq s(x_2)$ or $s(x_1) = s(x_2)$. In the latter case we deduce that either $[1 + x_1^2]^{-1} \neq [1 + x_2^2]^{-1}$ or $[1 + (x_1 - 1)^2]^{-1} \neq [1 + (x_2 - 1)^2]^{-1}$. Thus, A separates points of $(-\infty, \infty)$. Moreover, since $s(x) > 0$ for $x \in (-\infty, \infty)$ it follows that for each point of $(-\infty, \infty)$ there is a function in A which does not vanish there. Hence A is dense in $C_0[-\infty, \infty]$.[6] But $\text{pdf}(x) \in C_0[-\infty, \infty]$ so that there exists an $a \in A$ for which $\|\text{pdf}(x) - a(x)\| < \epsilon$. Since $a(x)$ is a rational function of the form (1) the theorem is proved.

PROOF OF COROLLARY:[7] Given $\epsilon > 0$ there exists an $a \in A$ and an integer n_0 for which $E(n; s) = \|\text{pdf}(x) - R'_{nn}(x; s, \gamma^*)\| \leq \|\text{pdf}(x) - a(x)\| < \epsilon$ for all $n \geq n_0$. It follows that $E(n, s) \to 0$ as $n \to \infty$, as required.

PROOF OF THEOREM 3: De la Vallée-Poussin first proved this theorem for polynomial and trigonometric function approximants. Achieser's proof in [**1**, pp. 52–53] for rational fractions holds also for our class of rational fractions (1).

PROOF OF THEOREM 4: The proofs of Part B (sufficiency), Part A(i) and Part A(iii) follow as in Achieser [**1**, pp. 55–56]. The difference with the classical theory in [**1**] arises in Part A(ii). In this case, when the denominator is most degenerate, we cannot use the construction that appears in the classical proof exactly as it stands. Instead (and working in the notation of [**1**]), we may find

[5] $C_0[-\infty, \infty]$ is not restricted to positive valued functions here as it is in the body of the article.

[6] Note that the algebra generated by A using pointwise multiplication is dense in $C_0[-\infty, \infty]$. An independent proof of this property has kindly been given to me by John Boyd III.

[7] [**11**] provides an earlier proof of this result by a longer but more direct method which overcomes a flaw in the original proof presented in [**10**].

polynomials $\phi(x)$ and $\psi(x)$ of degree $m - \mu + [\nu]_e$ and $n - \nu + [\nu]_e$ respectively where $[\nu]_e$ denotes the largest even integer less than or equal to ν and for which $\Phi(x) = A(x)\psi(x) - B(x)\phi(x)$. The remainder of the proof now holds with $\Omega(x)$ equal to any positive polynomial of degree $[\nu]_e$ and $N' \leqq n + m - \mu - \nu + [\nu]_e + 1$, leading to A(ii) as stated in the theorem.

PROOF OF THEOREM 5: [**1**, pp. 56–57].

REFERENCES

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

[2] DANIELS, H. E.: "The Approximate Distribution of Serial Correlation Coefficients," *Biometrika*, 43(1956), 169–185.

[3] EVANS, G. B. A., AND N. E. SAVIN: "Testing for Unit Roots: 1," *Econometrica*, 49(1980), 753–779.

[4] FREUD, G., D. J. NEWMAN, AND A. R. REDDY: "Chebyshev Rational Approximate to $e^{-|x|}$ on the Whole Real Line," *Quarterly Journal of Mathematics Oxford*, (1977).

[5] GOLDSMITH, H.: "The Exact Distributions of Serial Correlation Coefficients and an Evaluation of Some Approximate Distributions," *Journal of Statistical Computation and Simulation*, 5(1977), 115–134.

[6] MEINARDUS, G.: *Approximation of Functions: Theory and Numerical Methods*, Vol. 13. New York: Springer, 1967.

[7] NEWMAN, D. J.: "Rational Approximation to $|x|$," *Michigan Mathematical Journal*, 11(1964), 11–14.

[8] PHILLIPS, P. C. B.: "Approximations to Some Finite Sample Distributions Associated with a First Order Stochastic Difference Equation," *Econometrica*, 45(1977), 463–485.

[9] ———: "Edgeworth and Saddlepoint Approximations in a First Order Non-Circular Autoregression," *Biometrika*, 65(1978), 91–98.

[10] ———: "Best Uniform and Modified Padé Approximation of Probability Densities in Econometrics," pp. 123–167 in *Advances in Econometrics*, ed. by W. Hildenbrand. Cambridge: Cambridge University Press, 1982.

[11] ———: "A New Approach to Small Sample Theory," Cowles Foundation Discussion Paper No. 608, 1981.

[12] ———: "Failure of the Alternation Theorem in Chebyshev Approximation by Rational Functions over $C_0[-\infty, \infty]$," Cowles Foundation Discussion Paper No. 638, 1982.

[13] ———: "Exact Small Sample Theory in the Simultaneous Equations Model," in *Handbook of Econometrics*, ed. by M. D. Intriligator and Z. Griliches. New York: North Holland, forthcoming, 1983.

[14] REEVES, J. E.: "The Distribution of the Maximum Likelihood Estimator of the Parameter in the First Order Autoregressive Series," *Biometrika*, 59(1972), 387–394.

[15] RICE, J. R.: *The Approximation of Functions*, Vols. I and II. New York: Addison Wesley, 1964 and 1967.

[16] SIMMONS, G. F.: *An Introduction to Topology and Modern Analysis*. New York: McGraw-Hill, 1963.

[17] TIMAN, A. F.: *Theory of Approximation of Functions of a Real Variable*. New York: MacMillan, 1963.

[18] WALSH, J. L.: "The Existence of Rational Functions of Best Approximation," *Transactions of the American Mathematical Society*, 33(1931), 668–689.

[19] ———: "Interpolation and Approximation by Rational Functions in the Complex Domain," *American Mathematical Society*, 1965 (originally published, 1935).