COWLES FOUNDATION FOR RESEARCH IN ECONOMICS

AT YALE UNIVERSITY

Box 2125, Yale Station
New Haven, Connecticut 06520

COWLES FOUNDATION DISCUSSION PAPER NO. 562

BEST UNIFORM APPROXIMATION TO PROBABILITY DENSITIES IN ECONOMETRICS

P.C.B. PHILLIPS

September 1980

BEST UNIFORM APPROXIMATION TO PROBABILITY DENSITIES IN ECONOMETRICS*

by

P. C. B. PHILLIPS

YALE UNIVERSITY

0. ABSTRACT

A new method of approximating the probability density functions (pdf's) of econometric estimators and test statistics is developed. It is shown that best uniform approximants to a general class of pdf's exist in the form of rational functions. A procedure for extracting the approximants is devised and is based on modifying multiple-point Padé approximants to the distribution. The new approximation technique is very general and should be widely applicable in mathematical statistics and econometrics. It has the advantage, unlike the Edgeworth and saddlepoint approximations, of readily incorporating extraneous information on the distribution, even qualitative information. The new procedure is applied to a simple simultaneous equation estimator and gives exceptionally accurate results even for tiny values of the concentration parameter.

---

*Paper presented to the 1980 World Congress of the Econometric Society. I wish to acknowledge with many thanks the substantial contribution of Sidnie Feit in programming the computational work reported in section 8 of the paper. Ralph Bailey also helped me with some of the original computational work on this section and to him my thanks. It is a pleasure also to thank my secretary, Karen Marini, for her time and skill in preparing the typescript.

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

1. INTRODUCTION

The idea of approximating small sample distributions rather than extracting their exact mathematical forms has a long history in statistics and a number of different techniques have been explored (Kendall and Stuart [21], give an introductory survey of some of these techniques in Chapters 6, 12 and 13). Approximations are clearly of importance in those cases where mathematical difficulties have prevented the development of an exact theory. Regression models with lagged endogenous variables as regressors provide such an example, which is of particular relevance in econometrics. Approximations to distributions are also useful in those cases where the exact mathematical expressions are too complicated for numerical computations. Some examples of the latter are discussed by Phillips [31], [32].

Several authors have recently obtained approximations to the distribution of econometric estimators and test statistics based on asymptotic series. The approximations used in most of these studies have been based on the first few terms of Edgeworth-type asymptotic expansions of the distribution function (df) or probability density function (pdf) of the statistic under consideration. An alternative approach which can, when it is available, provide significant improvements on the Edgeworth approximation, particularly in tail areas, is based on the method of steepest descents in contour integration. This method leads to the saddlepoint approximation. Its use was systematically explored for the first time in statistics by Daniels [9,10] and has recently been the subject of renewed interest ([11], [12], [13], [19], [29], [30].

Both these methods of approximation are capable of representing the

exact distribution to an acceptable degree of accuracy in certain parameter environments. This is confirmed by the numerical evaluations in Anderson and Sawa [1,2], Phillips [27,29], and Holly and Phillips [19]. Moreover, the approximate distributions that have been obtained in the literature have already given valuable information concerning the small sample behavior of competing estimators and the adequacy of asymptotic theory in simple simultaneous equations and dynamic models. However, given the current state of our knowledge, the use of either of these methods in practical econometric work to advise on the choice of estimator and improve inferential accuracy is bound to encounter difficulties, some of them major.

First of all, there are certain parameter environments where the performance of the approximations is poor, sometimes a good deal worse than the asymptotic distribution (particularly in the case of the Edgeworth approximation). Unfortunately, the parameter environments for which this poor performance obtains are not at all unusual. As we might expect, given that the approximations are based on asymptotic series, this problem tends to become more widespread when sample sizes are small. Some indication of the wayward nature of these approximations in certain parameter environments is already available in [27] and [29]. Further documentation will be given in a companion article [33] to the present.

Secondly, although general formulae for the Edgeworth approximation are now available ([27], [39]) and widely applicable, the saddlepoint technique is still only practicable in specialized cases where the characteristic function is available or simple integral formulae for the pdf can be used (such as in the case of ratios [10]) or there exist a set of sufficient

statistics for the parameters to be estimated [13]. No doubt progress will be made in tackling some of these latter difficulties but, in the meantime, they remain a barrier to the general use of the procedure in analysing small sample distributions and influencing the way in which empirical work is actually undertaken.

Another difficulty that can arise in the use of the saddlepoint technique is that, for certain values of the argument of the pdf, singularities can occur within the strip of the imaginary axis containing the saddlepoint through which the path of integration is normally deformed. In such cases, this path of deformation is no longer permissable and special techniques must be used to smooth the approximation past the singularity; the resulting approximants are called uniform asymptotic expansions. Uniform approximants are typically much more complicated in form than the saddle-point approximation (an example is given in [30]). They are not always easy to extract and further work is required to splice them with the saddlepoint approximation, where it does exist, to cover the whole of the distribution.

Finally, it seems difficult to embody additional information on the distribution in question into these approximations. To take a simple example, in spite of the fact that the actual pdf is non negative and the df monotonic it is sometimes awkward to modify the Edgeworth approximations so that they share these properties. To take a more complicated example, we often know or can find the leading term in the series representation of the exact pdf (in many cases, without knowing the full expression for the pdf). This leading term frequently has a simple algebraic form and is instrumental in determining the behavior of the exact distribution in certain domains,

particularly the tails. Yet, even when this information is available, there seems to be no obvious way of building it into either the Edgeworth or the saddlepoint approximation. The resulting approximations, therefore, end up neglecting what is potentially very useful analytic information on the form of the distribution.

The purpose of the present paper is to introduce a new technique of approximating sampling distributions. The technique is very general and should be widely applicable in mathematical statistics and econometrics. It has the advantage, unlike the Edgeworth and saddlepoint approximations, of readily incorporating extraneous information on the distribution, even qualitative information. Moreover, since the technique is not based on an asymptotic series expansion in terms of the sample size or concentration parameter, accurate approximations can be obtained even in very small samples. The technique should, therefore, be most useful in cases where the Edgeworth and saddlepoint approximations run into difficulty. It turns out that the new approximation is close to the best uniform approximant in the class of certain rational functions. These approximants are discussed and the class of rational functions to be used is defined in section 2. A general theory of best uniform approximation in the context of density approximation is given in sections 3 and 4. These sections provide the theoretical basis for the new technique. Sections 5, 6 and 7 describe the procedure and give the general formulae needed in applications. In section 8 the method is applied to a simple simultaneous equations estimator, facilitating a comparison between the new and existing techniques of approximation.

2. A GENERAL CLASS OF DENSITY FUNCTIONS AND RATIONAL APPROXIMANTS

To fix ideas, we write the estimator or test statistic in which we are interested as θ_T . In what follows, we treat θ_T as a scalar so that, when dealing with estimators, we are in effect concentrating on the marginal distribution of individual components of a complete vector of estimates. The characteristic function (cf) of θ_T is written as $cf(s) = E\left(e^{is\theta_T}\right)$ and is assumed to be absolutely integrable. This implies that θ_T has a bounded, continuous pdf given on inversion by

$$(1) \quad pdf(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-isx}\, cf(s)ds.$$

Moreover, by the Riemann Lebesgue lemma, it follows from (1) that $pdf(x)\to 0$ as $x\to\pm\infty$. Thus, the effect of the integrability requirement on cf(s) is to confine our attention to the class of densities covered by the following assumption:

Assumption 1 θ_T has a continuous pdf which tends to zero at the limits of its domain of definition $(\pm\infty)$.

Note that the boundedness of the pdf now follows from its continuity and behavior at $\pm\infty$. Assumption 1 covers a wide variety of densities arising in econometric work. It can, in fact, be extended to allow for certain types of discontinuity and singularity but this complicates the development of the approximants that follow. In this paper, we will, therefore, keep to the class of densities defined by Assumption 1. This is sufficiently general to include all the usual simultaneous equations estimators and test statistics, as well as their extensions to models with lagged endogenous variables as regressors and autoregressive, moving average errors.

Having defined the class of density functions, the general problem of approximation takes the following form: for a particular density function pdf(x), find an approximating function which depends on a finite number of parameters whose values are selected in such a way that the approximating function is as close, in some sense, to the original density as possible over its entire domain of definition. Once stated in this way it is clear that there are two major components to the problem. The first is the form the approximating function should take. The second is the criterion of closeness of approximation to be used in selecting the best approximant. By a best approximant we mean the member (or members) of the given family of approximating functions whose closeness to the function pdf(x) cannot be improved by any other member of the same family. Thus, the second problem clearly raises the further question of whether or not there exists a best approximation to pdf(x) in the given family of approximants. This question of existence will be the subject of the next section. We now define the class of approximating functions and the measure of approximation to be used in the rest of the article.

Definition If $s(x)$ is a real continuous function satisfying $s(x)>0$ and $s(x)\to 0$ as $x\to\pm\infty$, then we define the class of rational approximating functions by

$$(2) \quad R_{n,n}(x;s,\gamma) = s(x)\frac{P_n(x)}{Q_n(x)} = s(x)\frac{a_o + a_1x + \ldots + a_nx^n}{b_o + b_1x + \ldots + b_nx^n} \qquad -\infty<x<\infty$$

where (i) the numerator and denominator are reduced to their lowest degree by the cancellation of identical factors; (ii) n is an even integer; and

(iii) $\gamma' = (a_o a_1, \ldots, a_n, b_o, b_1, \ldots, b_n) \in \Gamma$, the parameter space, which is defined as the following subspace of $2n+2$ dimensional Euclidean space

$$\Gamma = \{\gamma: \Sigma_{i=0}^{n} b_i^2 = 1 \ , \ Q_n(x) > 0 \quad \text{for all} \quad x \in (-\infty, \infty)\}.$$

The condition $\Sigma_{i=0}^{n} b_i^2 = 1$ on the parameter space Γ is a normalization which eliminates the redundancy in the coefficients of the rational function (2). Other normalizations such as $b_o = 1$ or $b_n = 1$ are possible and may be more useful in applications. We will, in fact, later use the normalization $b_o = 1$ in the application of section 8 but the present definition of Γ is retained for the theoretical development.

The condition $Q_n(x) > 0$ ensures that the rational fractions (2) have no poles on the real line and are, therefore, compatible with the class of density functions to be approximated. Since this is possible only when n is an even integer, we have introduced this requirement explicitly under (ii). On the other hand, when the denisty function we wish to approximate is non zero on part rather than all of the real axis, it is clear that this requirement may be relaxed. Moreover, if singularities in the density function do occur on the real axis we may remove the condition $Q_n(x) > 0$. If the position of the singularity is known, this can be incorporated directly into (2); otherwise it must be approximated and, for certain values of n , may not be captured by the approximation, although whether or not this occurs will depend on the technique that is used to construct the approximation.

We might consider working with the somewhat wider class of rational functions for which the numerator and denominator polynomials were not necessarily of the same degree. In certain applications it may seem appropriate to make such a generalization of the class of approximants, and the theory we develop can be modified to take this generalization into

account. However, there are various reasons why we do not choose to work with the more general class in developing our theory. The first is that the coefficient function s(x) will frequently be constructed so that it captures the behavior of the exact pdf(x) as x approaches the limits of its domain. A rational fraction of equal degree is then immediately compatible with this behavior. The second is that when the numerator and denominator are of equal degree, modifications to the coefficients that are designed to avoid unwanted zeros and poles in the final approximant are easier to make. That this is of particular importance will be seen in section 4 where the practical procedure we develop for obtaining a good approximant of the type (2) is based on modifying multiple-point Padé approximants, which in crude form will frequently possess zeros and poles that need to be removed in order to improve the approximation over the whole real line. Finally, numerical experience with rational function approximations in applied mathematics ([17], [24]) suggests that rational fractions with numerator and denominator of equal or near equal degree tend, on the whole, to give better approximations than those for which the degrees differ markedly. Taking an extreme case of comparison, polynomial approximations usually become unsatisfactory when it is necessary to approximate a function over a wide interval. Moreover, they lack the capacity to turn corners sharply and then go straight for long periods, particularly in a direction almost parallel to the horizontal axis. These properties are useful ones for a density function approximant to be capable of capturing. An important feature of rational fraction approximations is that even low degree fractions of the type (2) are flexible enough to assume this behavior. This is endorsed by the large number of numerical results with rational approximants reported by Hastings [18] and Hart [17]. It will also be confirmed in our own application of the technique reported in section 5.

In order to develop a theory for the goodness of approximation based on

members of the class (2) we introduce a norm to measure the error in the approximation. We will use the uniform norm (also known as the Tchebycheff or L_∞ norm) defined as

$$(3) \quad ||f(x)|| = \sup_{x\varepsilon(-\infty,\infty)} |f(x)|.$$

If we now let $f(x) = pdf(x) - R_{n,n}(x;s,\gamma)$ denote the approximation error, our problem is, for a given value of n and a given function $s(x)$, to find a value of γ which minimizes the maximum error[1]. At this value of γ, $R_{n,n}(x;s,\gamma)$ is then called a best uniform (or Tchebycheff) approximation to $pdf(x)$.

Other choices of norm are certainly possible and will generally lead to different best approximations, where they exist. However, for accurately approximating $pdf(x)$ over a wide interval the choice of the uniform norm seems very appropriate.

3. BEST UNIFORM APPROXIMATION BY RATIONAL FUNCTIONS

The theory of best uniform approximation of real continuous functions by rational fractions has a long history. One of the earliest discussions was undertaken by Tchebycheff [43]. Frobenius [16] and Padé [26] both systematically investigated the properties of a specialized class of rational approximants now known as Padé approximants (see section 4). In the complex domain, Runge [37] (see Rudin [36] chapter 13) established the possibility of uniform approximation of analytic functions by rational fractions with pre-assigned poles. A general theory of approximation in the complex domain by

1. Note that since $pdf(x)$ and $R_{n,n}(x)$ are continuous and have the same limits at $\pm\infty$, the maximum error will occur at a finite point on the real line. Note also that when there is no risk of misunderstanding we will abbreviate $R_{n,n}(x;s,\gamma)$ to $R_{n,n}(x)$.

rational functions was developed in the treatise by Walsh [44]. Extensive modern treatments of the subject covering all the classical results on the approximation of real valued functions are given in the volumes by Rice [34] and Meinardus [24]. Since the theory in this literature (with the exception of Walsh [44]) has been concerned with the approximation of functions which are defined over compact sets, the present section will be devoted to the development of a theory which is applicable over the whole real line and is, therefore, directly relevant to the problem of density function approximation. Our treatment of the problem will be based on the framework laid out in section 2 and will follow the lines in Rice [34], particularly his section 3.8.

To establish the existence of a best uniform approximant to a given pdf(x) in the class of rational fractions defined by (2) we need to show that there exists a set of parameters γ^* for which

$$(4) \quad ||R_{n,n}(x;s,\gamma^*) - \text{pdf}(x)|| = \inf_{\gamma\in\Gamma} ||R_{n,n}(x;s,\gamma) - \text{pdf}(x)||.$$

$$= \rho \quad , \text{ say.}$$

Now $0\leq\rho<\infty$ and we can find a sequence of rational fractions $\{R_{n,n}(x;s,\gamma^{(j)})\}$ for which

$$(5) \quad \rho_j = ||R_{n,n}(x;s,\gamma^{(j)}) - \text{pdf}(x)||$$

and

$$(6) \quad \lim_{j\to\infty} \rho_j = \rho.$$

It remains to prove that the parameter sequence $\{\gamma^{(j)}\}$ has a convergent subsequence which converges to a set of finite parameters. If we call the

latter γ^* then it will follow from (6) that γ^* satisfies (4). As discussed by Rice (pp. 26-27 of [34]), the crucial part of the proof of existence is to demonstrate that the parameters lie in a compact set[1]. First, we show that we may restrict our attention to bounded subsets of Γ .

Definition (Condition E of Rice[2])

The approximating function $R_{n,n}(x;s,\gamma)$ is said to satisfy Condition E for the norm $||\ ||$ if, given $M<\infty$, there is an $N<\infty$ such that

$$||R_{n,n}(x;s,\gamma)|| \leq M$$

implies that

$$\max_i |\gamma_i| \leq N$$

where $\gamma = (\gamma_i)$.

In view of (5) and (6), there is an integer j_o for which

$$(7) \quad ||R_{n,n}(x;s,\gamma^{(j)}) - pdf(x)|| \leq \rho+1$$

for all $j>j_o$. Moreover, from Assumption 1 it follows that there exists $K>0$ for which $||pdf(x)|| \leq K$ and, hence, using (7) we have the inequality

$$(8) \quad ||R_{n,n}(x;s,\gamma^{(j)})|| \leq K+\rho+1.$$

We now verify that Condition E holds for the approximating function $R_{n,n}(x;s,\gamma)$.

Lemma 1 The rational fraction $R_{n,n}(x;s,\gamma)$ defined by (2) satisfies Condition E for the uniform norm (3).

---

1. at least in this region of the overall parameter space Γ. We can, for example, exclude as irrelevant in terms of (4) those regions of Γ for which the parameters yield unbounded rational fractions.

2. See [34] page 27.

Proof

We consider the set

$$(9) \quad \{\gamma: \; ||R_{n,n}(x;s,\gamma)|| \leq M \;, M>0\}$$

Since $||R_{n,n(x;s,\gamma)}|| \leq M$ implies that, for a given number $L>0$,

$$\max_{x\in[-L,L]} |R_{n,n}(x;s,\gamma)| \leq M$$

it follows that (9) lies in the set

$$(10) \quad \{\gamma: \; \max_{x\in[-L,L]} |R_{n,n}(x;s,\gamma)| \leq M\}$$

Now for $x\in[-L,L]$ and taking $L>1$, we have

$$|R_{n,n}(x;s,\gamma)| \geq \frac{\min\limits_{x\in[-L,L]} |s(x)|}{\max\limits_{x\in[-L,L]} |Q_n(x)|} |P_n(x)|$$

$$(11) \qquad \geq \frac{s_L(L^2-1)^{\frac{1}{2}}}{(L^{2(n+1)}-1)^{\frac{1}{2}}} |P_n(x)|.$$

where

$$s_L = \min_{x\in[-L,L]} |s(x)| \geq 0.$$

Thus, when γ lies in the set (10), we have

$$\frac{M(L^{2(n+1)}-1)^{\frac{1}{2}}}{s_L(L^2-1)^{\frac{1}{2}}} \geq \max_{x\in[-L,L]} |P_n(x)| = \max_{x\in[-L,L]} |\sum_{i=0}^{n} a_i x^i|$$

and the polynomial $\Sigma_{i=0}^{n} a_i x^i$ is bounded uniformly on the interval $[-L,L]$. It follows that the coefficient parameters a_i are also bounded. Moreover,

$\Sigma_{i=0}^{n} b_i^2 = 1$ by definition, so that the parameter set (10) is bounded. By implication, the same is true for the set (9) and, thus, $R_{n,n}(x;s,\gamma)$ satisfies Condition E for the uniform norm. □

This Lemma shows that we can confine our attention to bounded subsets of Γ in searching for a best approximant. Hence, for all $j \geq j_o$ the sequence $\{\gamma^{(j)}\}$ introduced earlier lies in a bounded subset of Γ. We may, therefore, select a subsequence that converges to the vector γ^*, say. If we reindex the subsequence we can write for the individual components of γ, $\lim_{j\to\infty} a_i^{(j)} = a_i^*$ and $\lim_{j\to\infty} b_i^{(j)} = b_i^*$ for $i = 0,1, \ldots ,n$.

Now it is important to note that since Γ is not closed γ^* may or may not lie in Γ. If $\gamma^* \in \Gamma$ then $R_{n,n}(x;s,\gamma^*)$ is a rational function of the form defined in (2) and, in view of (4), is therefore a best uniform approximant of pdf(x). But, if $\gamma^* \notin \Gamma$ then $R_{n,n}(x;s,\gamma^*)$ is the limit of a sequence of functions and is not necessarily a rational function itself. In fact, it may not even be continuous (we give an example later in this section). However, the limit function $R_{n,n}(x;s,\gamma^*)$ will differ from a rational function, $R'_{n,n}(x;s,\gamma^*)$ say, only at a finite number of points. And, in fact, $R'_{n,n}(x;s,\gamma^*)$ is a best uniform approximant to pdf(x) in the class defined by (2).

The problem discussed in the last paragraph arises because although the denominator polynomial $Q_n(x) = Q_n(x;\gamma) > 0$ for $\gamma \in \Gamma$, this no longer necessarily holds when γ does not lie in Γ. Since γ^* is a subsequential limit of elements of Γ it follows that, in the limit, $Q_n^*(x) = Q_n(x;\gamma^*)$ can have at most n zeros. If we let $P_n^*(x) = P_n(x;\gamma^*)$ be the limit of the numerator polynomial as $\gamma^{(j)} \to \gamma^*$ it follows that there are at most

n points where $R'_{n,n}(x;s,\gamma^*) = s(x)P_n^*(x)/Q_n^*(x)$ is undefined. At all other points we must have $R_{n,n}(x;s,\gamma^{(j)}) = R'_{n,n}(x;s,\gamma^*)$. Moreover, since $R'_{n,n}(x;s,\gamma^*) = \text{pdf}(x) + \{R_{n,n}(x;s,\gamma^{(j)}) - \text{pdf}(x)\} + \{R'_{n,n}(x;s,\gamma^*) - R_{n,n}(x;s,\gamma^{(j)})\}$ it follows from (7) that for $j \geq j_o$

$$|R'_{n,n}(x;s,\gamma^*)| \leq K + \rho + 1 + |R'_{n,n}(x;s,\gamma^*) - R_{n,n}(x;s,\gamma^{(j)})|$$

and allowing $j \to \infty$ we deduce that

$$(12) \qquad |R'_{n,n}(x;s,\gamma^*)| \leq K + \rho + 1$$

Hence, for all x other than zeros of $Q_n^*(x)$, we have the inequality

$$(13) \qquad s(x)|P_n^*(x)| \leq (K + \rho + 1)\, Q_n^*(x).$$

By continuity, (13) holds also when $Q_n^*(x) = 0$. Thus any real zero of $Q_n^*(x)$ is also a zero of $P_n^*(x)$, since $s(x) > 0$ for all finite x . We, therefore, eliminate by cancellation each linear factor of $Q_n^*(x)$ corresponding to a real root of $Q_n^*(x) = 0$. We call the resulting rational fraction $R'_{n,n}(x's,\gamma^*)$ and note that for all values of x other than zeros of $Q_n^*(x)$

$$R'_{n,n}(x;s,\gamma^*) = R_{n,n}(x;s,\gamma^*)$$

while at the zeros $\{x_k : k = 1, \ldots, m \leq n\}$ of $Q_n^*(x)$

$$R'_{n,n}(x_k;s,\gamma^*) = \lim_{x \to x_k} R_{n,n}(x;s,\gamma^*).$$

Finally, we note that

$$||R_{n,n}(x;s,\gamma^*) - \text{pdf}(x)|| = \sup_{x \in (-\infty,\infty)} |R_{n,n}(x;s,\gamma^*) - \text{pdf}(x)|$$

$$= \max \left\{ \sup_{\substack{x \in (-\infty,\infty) \\ x \neq x_k}} |R_{n,n}(x;s,\gamma^*) - \text{pdf}(x)|, |R_{n,n}(x_k;s,\gamma^*) - \text{pdf}(x_k)| \right.$$

$$\left. i = 1, \ldots, m \right\}$$

$$\geq \sup_{x\in(-\infty,\infty)} |R'_{n,n}(x;s,\gamma^*) - \mathrm{pdf}(x)|$$

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

The rational function $R'_{n,n}(x;s,\ *)$ is, therefore, a best uniform approximation of pdf(x) and we have proved the following □

Theorem 1 (Existence of a best uniform approximant)

If pdf(x) satisfies Assumption 1 on $(-\infty,\infty)$, then there is a best uniform approximant to pdf(x) in the class of rational functions defined by (2).

To illustrate the problem that arises in the proof of this theorem because Γ is not closed, we consider the following density function of the Pareto distribution

$$\mathrm{pdf}(x) = \frac{ak^a}{x^{a+1}} \qquad a>0\ ,\ x\geq k>0$$

We consider the case in which $a = 2$ and a class of rational approximants of the form (2) is being used with $s(x) = 1/x$ $(x\geq k)$ and $n = 4$. Now consider the sequence of approximants defined by

$$R_{4,4}(x;s,\gamma^{(j)}) = \left(\frac{2k^2}{x}\right) \frac{\left(\frac{1}{1+k^4} - \frac{1}{2j^2}\right)^{1/2}x^2 - \left(\frac{k^4}{1+k^4} - \frac{1}{2j^2}\right)^{1/2}}{\left(\frac{1}{1+k^4} - \frac{1}{2j^2}\right)^{1/2}x^4 - \left(\frac{k^4}{1+k^4} - \frac{1}{2j^2}\right)^{1/2}x^2 + \frac{1}{j}}$$

Thus

$$\gamma^{(j)'} = \left[-\left(\frac{k^4}{1+k^4} - \frac{1}{2j^2}\right)^{1/2},0,\left(\frac{1}{1+k^4} - \frac{1}{2j^2}\right)^{1/2},0,0;\frac{1}{j},0,\ -\left(\frac{k^4}{1+k^4} - \frac{1}{2j^2}\right)^{1/2},0,\left(\frac{1}{1+k^4} - \frac{1}{2j^2}\right)^{1/2}\right]$$

and, as $j\to\infty$,

$$\gamma^{(j)'}\to\gamma^{*'} = \left[-\left(\frac{k^4}{1+k^4}\right)^{1/2},0,\left(\frac{1}{1+k^4}\right)^{1/2},0,0;0,0,\ -\left(\frac{k^4}{1+k^4}\right)^{1/2},0,\left(\frac{1}{1+k^4}\right)^{1/2}\right]$$

The limiting function is then

$$R_{4,4}(x;s,\gamma^*) = \begin{cases} \dfrac{2k^2}{x^3} & x > k \\ \\ 0 & x = k \end{cases}$$

which is not a rational function nor is it continuous on the interval $[k,\infty)$. However

$$R'_{4,4}(x;s,\gamma^*) = \frac{2k^2}{x^3} \qquad x \geq k$$

is rational, continuous and is clearly the best uniform approximant to pdf(x) on $[k,\infty)$.

4. A CONVERGENCE THEOREM

As the degree of the best approximating rational fraction increases, the error $E(n,s) = ||\text{pdf}(x) - R'_{n,n}(x:s,\gamma^*)||$ must be at least as small. In fact, as Theorem 3 below shows, $E(n,s) \to 0$ as $n \to \infty$, so that the best approximant $R'_{n,n}(x;s,\gamma^*)$ converges to pdf(x) as $n \to \infty$. It follows that, for any choice of density function satisfying Assumption 1, there is an arbitrarily close rational approximant. In this sense, the rational fractions of the class defined by (2) are dense in the set of density functions that satisfy Assumption 1.

Meinardus [24] proves a related theorem on the convergence of rational fractions to a continuous function over a bounded interval[1].

1. The Weierstrass Theorem ([24] p. 7) established the same result for polynomial approximants.

Theorem 2 (Meinardus)

(i) The real valued function $f(x)$ is continuous and non-negative over the interval $[-1,1]$ and is approximated by

$$(14) \quad R_{m,n}(x) = \frac{P_m(x)}{Q_n(x)}$$

where $P_m(x)$ and $Q_n(x)$ are polynomials with real coefficients and $Q_n(x)>0$ for $x\varepsilon[-1,1]$.

$$(ii) \quad E_{m,n}(f) = \inf_{R_{m,n}\varepsilon V_{m,n}} ||f - R_{m,n}||$$

where $V_{m,n}$ is the set of all rational functions as in (14).

Then

$$\lim_{n+m\to\infty} E_{m,n}(f) = 0$$

independent of the manner in which we pass to the limit.

Proof Meinardus [ ] pp. 158-160. □

Theorem 3 If pdf(x) satisfies Assumption 1 on $(-\infty,\infty)$ and

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

where $R'_{n,n}(x;s,\gamma^*)$ is the best uniform approximant to pdf(x) in the class of rational fractions defined by (2) then

$$\lim_{n\to\infty} E(n,s) = 0.$$

Proof

Let $\varepsilon>0$ be arbitrarily small. Then, by Assumption 1 and the definition of $R_{n,n}(x;s,\gamma)$, there exists an $L>0$ for which

$$\sup_{|x|>L} |pdf(x) - R'_{n,n}(x;s,\gamma^*)| < \varepsilon.$$

Now

$$||pdf(x) - R'_{n,n}(x;s,\gamma^*)|| = \max \left\{ \max_{x\varepsilon[-L,L]} |pdf(x) - R'_{n,n}(x;s,\gamma^*)|, \sup_{|x|>L} |pdf(x) - R'_{n,n}(x;s,\gamma^*)| \right\}$$

$$= \max \left\{ \max_{x\varepsilon[-L,L]} |pdf(x) - R'_{n,n}(x's,\gamma^*)|, \varepsilon \right\}$$

It only remains to show that there is an n_o for which $n>n_o$ implies that

$$(15) \quad \max_{x\varepsilon[-L,L]} |pdf(x) - R'_{n,n}(x;s,\gamma^*)| < \varepsilon$$

For then, since ε is arbitrarily small, we can approximate pdf(x) by $R'_{n,n}(x;s,\gamma^*)$ over the whole real line as closely as we please for sufficiently large n. Hence, $E(n,s)\to 0$ as $n\to\infty$.

In fact, (15) follows from Theorem 2. We need only transform $x = Ly$ with $-1\le y\le 1$ and setting

$$R'_{n,n}(x;s,\gamma^*) = s(Ly)\,\frac{P_n'(Ly)}{Q_n'(Ly)} = \bar{s}(y)\,\frac{\bar{P}_n'(y)}{\bar{Q}_n'(y)},$$

$$pdf(x) = pdf(Ly) = \overline{pdf}(y)$$

we have

$$\max_{x\varepsilon[-L,L]} |pdf(x) - R'_{n,n}(x)| = \max_{y\varepsilon[-1,1]} \bar{s}(y) \left| \frac{\overline{pdf}(y)}{\bar{s}(y)} - \frac{\bar{P}_n'(y)}{\bar{Q}_n'(y)} \right|$$

Since $\overline{pdf}(y)/\bar{s}(y)$ is continuous and non-negative over $[-1,1]$ it follows by Theorem 2 that (15) holds for n sufficiently large $\square$

5. LOCAL EXPANSIONS FOR DENSITIES

The theory of the last two sections shows that for a given pdf in the class defined by Assumption 1 there exists a best rational fraction approximant of the type (2) and that, as we increase the degree of the approximant, this converges to pdf(x) over the entire real axis. In any practical situation, of course, we will need to prescribe the degree of the approximant to be used and attempt to find the best approximant in the given class. This normally requires numerical methods and the algorithms discussed in the literature[1] rely on knowledge of the true function values at a grid of points as well as, in certain cases, the function derivatives. This seems too much to expect in an econometric context where, even in those cases where the exact density function is known in analytic form, numerical computations are often impossible because of convergence problems with the multiple series representation of the density or the inadequate tabulations of the special polynomials that appear in the analytic expressions.

We are, therefore, left with the problem of how, in a given situation, to get close to the best approximant in the class (2) without having to rely on arbitrary evaluations of the exact distribution. The solution we present to this problem in the present and succeeding section of the paper is based on the idea of using the local behavior of the true density in the body of the distribution and in the tails to construct a global approximation of the

---

1. See for example [24] pp. 170-171 and [34] ch. 6.

form (2). In principle, the procedure we develop for moving from local to global density approximations can be based on the knowledge of local behavior at an arbitrary set of points. But, in practice, it will be sufficient to use information concerning the local behavior of the density in the tails and around the centre of the distribution. The application we consider in section 8 will show that this information is sufficient to secure excellent global approximations to rather complicated density functions even with rational fractions of lower degree.

Local behavior of density functions can take the form of expansions about the value of the function at a certain point or perhaps estimates of the function values obtained from Monte Carlo simulations. We will deal with the case where some analytic information from local expansions is available while, at the same time, it should be clear how the procedure we develop can also be used to accommodate Monte Carlo evidence.

Our present analytic knowledge of the exact distribution of a variety of econometric estimators and test statistics show that there exists an asymptotic expansion of the density function in ascending powers of x^{-1} as the argument x approaches the limits of its domain ($\pm\infty$ or $+\infty$) . In general, we can write the expansion about infinity in the form

$$(16) \quad \text{pdf}(x) \sim t(|x|)\{\alpha_0 + \alpha_1/x + \alpha_2/x^2 + \alpha_3/x^3 + \alpha_4/x^4 \ldots \}$$

as $x \to \pm\infty$. The coefficient function $t(|x|) \to 0$ as $|x| \to \infty$ and, in the case of most of the common simultaneous equations estimators, is of the form $t(|x|) = |x|^{-k}$ where $k \geq 2$. Thus, in the case of the two stage least squares estimator $k = \ell+2$ where ℓ is the degree of overidentication in the equation being estimated. An expansion of the type (16) was developed by

Sargan and Mikhail [41] for the instrumental variable estimator and was used by Sargan [40] in the analysis of Monte Carlo estimates of moments that do not exist.

At points $\{d_i : i = 1, \ldots, I\}$ where pdf(x) is continuously differentiable to an appropriate order we have the Taylor expansions

$$(17) \quad \text{pdf}(x) = \beta_{i0} + \beta_{i1}(x-d_i) + \beta_{i2}(x-d_i)^2 + \beta_{i3}(x-d_i)^3 + \beta_{i4}(x-d_i)^4 + \ldots \qquad (i = 1, \ldots, I)$$

In a number of cases, we also have the analytic form of the leading term in the series representation of the density. If we denote this leading term by $w(x)$, then it will be useful to consider extensions of the expansion (17) which take the form

$$(18) \quad \text{pdf}(x) = w(x)\{\beta_{i0} + \beta_{i1}(x-d_i) + \beta_{i2}(x-d_i)^2 + \beta_{i3}(x-d_i)^3 + \beta_{i4}(x-d_i)^4 + \ldots\}$$

Two obvious choices of the points d_i are: (i) the origin, particularly for certain test statistics like the "t" ratio; and (ii) the true value of the relevant parameter, when pdf(x) refers to the marginal distribution of a certain estimator.

Although expansions such as (16) and (17) usually produce good approximations only in the immediate neighborhood of the point of expansion, they can be used to construct approximations which perform well outside the immediate locality of the approximation, while retaining the good behavior of the original expansions within the locality. With reference to (17), the fourth degree polynomial in x may yield a good approximation to pdf(x) in

a neighborhood of the point d_i ; but, in most cases, its performance will rapidly deteriorate outside of this neighborhood and it will be quite inadequate as an approximation on the tails. On the other hand, the coefficients β_{il} in the expansion (17) usually contain information which can produce greatly improved approximations outside the range in which the expansion (17) itself is immediately useful. That this is so is demonstrated by the extensive practical experience with Padé approximants in the applied mathematics literature. These approximants are rational fractions for which the corresponding Taylor series matches the Taylor series expansion of a given function to as many powers as is possible. In the present context, we can refer to the following example used by Baker [ ]:

$$(19) \qquad f(x) = \left(\frac{1+2x}{1+x}\right)^{\frac{1}{2}} = 1 + \frac{1}{2}x - \frac{5}{8}x^2 + \frac{13}{16}x^3 - \frac{141}{128}x^4 + \ldots$$

The Taylor series for $f(x)$ in (19) has radius of convergence equal to $\frac{1}{2}$. Yet as x becomes large $f(x)$ is a well bahaved function which tends to $\sqrt{2}$ as $x \to \infty$. Using only the first three coefficients $1, \frac{1}{2}, -\frac{5}{8}$ in (18), we construct the Padé approximant

$$(20) \qquad \frac{1 + (7/4)x}{1 + (5/4)x} = 1 + \frac{1}{2}x - \frac{5}{8}x^2 + \frac{25}{32}x^3 + \ldots$$

This has the same Taylor series expansion about the origin as $f(x)$ to $0(x^2)$ and it tends to $7/5 = 1.4$ as $x \to \infty$. Thus, using only three coefficients in a local expansion about the origin, the Padé approximation (19) provides an approximation at infinity to $f(x)$ which differs at the second decimal place. Even within the radius of convergence of the Taylor expansion (20) outperforms the Taylor expansion: for instance, at $x = 1/4$ $f(x) = 1.0954451$, the first three terms of the Taylor expansion give 1.0859375 while (20) equals

1.0952381 providing at least another decimal place of accuracy.

This example suggests that Padé approximants can have the useful property of accelerating the convergence of a given power series within its circle of convergence, while at the same time considerably extending the domain over which truncated series expansions can give useful results. These features make Padé approximants attractive for constructing first step rational fraction approximations from the information embodied in purely local density expansions such as (16), (17) or (18). Section 6 will be devoted to the algebraic details of this construction and will give the appropriate formulae.

Since the coefficients in the local expansions are needed in the construction of rational fraction approximants, we now give an analytic procedure for extracting local density expansions such as (16) and (17). It will be useful to first make explicit the general form of the characteristic function.

Assumption 2 (i) The characteristic function $cf(s)$ has the general form

$$(21)\quad cf(s) = cf_1(s) + cf_2(s) + cf_3(s)$$

where

$$cf_1(s) = e^{i\eta s} \sum_{m=0}^{M-1} p_m (is)^m ,$$

$$cf_2(s) = e^{i\eta s}|s|^{\mu} \sum_{k=0}^{K} \sum_{\ell=0}^{L(k)} q_{k\ell}|s|^k (\ell n|s|)^{\ell} \qquad \mu \geq M;\ L(k) = 0 \text{ or } 1 \text{ for all } k$$

$$cf_3(s) \in L(-\infty,\infty) \quad , \text{ i.e. is absolutely integrable over the interval } (-\infty,\infty) \text{ , for } j=0,1, \ldots ,N \text{ where } N \text{ is the smallest integer } \geq \mu+K+1$$

$$cf_3^{(j)}(s) \to 0 \quad \text{as} \quad s \to \pm\infty .$$

(ii) The behavior of $cf(s)$ as $s \to 0$ is given by the asymptotic series

expansion

$$cf(s) \sim e^{i\eta s} \left\{ \sum_{m=0}^{M-1} p_m (is)^m + |s|^{\mu} \sum_{k=0}^{\infty} \sum_{\ell=0}^{L(k)} q_{k\ell} |s|^k (\ell n|s|)^{\ell} \right\}.$$

This assumption is sufficiently general to include a very wide class of distributions and should apply to most econometric estimators and test statistics both in classical and non-classical (including dynamic model) situations. The first component of the characteristic function $cf_1(s)$ is analytic and ensures that integral moments will exist to order M-1 if this is an even integer and M-2 if M-1 is odd [23]. In cases where the distribution does not possess all its moments, the second component, $cf_2(s)$, of (21) is important in the local behavior of cf(s) in the neighborhood of the origin and is, as we shall see below, instrumental in determining the form of the tails of pdf(x).

We start with the following two basic results which relate the tail behavior of density functions to the regularity properties of the characteristic function. They follow without difficulty from the standard discussions on this subject in the literature;[1] but they also demonstrate that we need to go somewhat further to extract a tail expansion of the form (16).

Lemma 2 If the distribution with density pdf(x) and characteristic function cf(s) has finite $(M-1)^{'th}$ absolute moment, then cf(s) is M-1 times continuously differentiable and the derivatives $cf^{(n)}(s) \to 0$ as $s \to \pm\infty$ for each $n = 0,1, \ldots, M-1$

Proof The first statement follows by dominated convergence from the existence of the $(M-1)^{'th}$ absolute moment. The behavior of the derivatives at $\pm\infty$

1. See, for example, Feller [15] or Lukacs [23].

follows from the representation

$$cf^{(n)}(s) = \int_{-\infty}^{\infty} e^{isx}(ix)^n pdf(x)dx$$

and, since $(ix)^n pdf(x)$ is absolutely integrable on $(-\infty,\infty)$, the Riemann-Lebesgue Lemma ensures that $cf^{(n)}(s)\to 0$ as $s\to\pm\infty$ for each $n = 0,1, \ldots ,M-1$. □

Lemma 3 (Erdéyli) If cf(s) is M-1 times continuously differentiable, if $cf^{(n)}(s)\to 0$ as $s\to\pm\infty$ and if $cf^{(n)}(s)$ is absolutely integrable for each $n = 0,1, \ldots ,M-1$, then

$$pdf(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-isx}cf(s)ds = \frac{(ix)^{-M+1}}{2\pi}\int_{-\infty}^{\infty} e^{-isx}cf^{(M-1)}(s)ds = o(x^{-M+1})$$

as $x\to\infty$.

Proof Erdéyli [14] p. 47

Lemma 3 shows that $pdf(x)\to 0$ as $x\to\pm\infty$ at least as fast as x^{-M+1}; but this is, in general, not a very sharp result. For, if the $(M-1)^{'th}$ absolute moment of the distribution exists and pdf(x) satisfies Assumption 1 then we would expect that $pdf(x) = O(x^{-M-\delta})$ for some $\delta>0$. For example, in the case of the Cauchy distribution $cf(s) = e^{-|s|}$ and Lemma 3 demonstrates that $pdf(x) = o(1)$ whereas, in fact, $pdf(x) = O(x^{-2})$. Thus, Lemmas 2 and 3 are not very helpful in providing local expansions about infinity of the form (16).

However, a sharper result which does lead directly to the asymptotic expansion (16) can be obtained from the more explicit representation of the characteristic function (21) and the theory of Fourier Transforms of generalized functions and their asymptotic expansions ([20], [22]).

Theorem 4 If the distribution with density pdf(x) and characteristic function cf(s) satisfies Assumptions 1 and 2 then pdf(x) has the following asymptotic expansion as $|x|\to\infty$

$$\text{pdf}(x) = \frac{1}{\pi|x|^{\mu+1}} \sum_{r=0}^{K} \left[\sum_{k+\ell=r} c_{k\ell} \left(\text{sgn}(x)\right)^k + \sum_{\substack{k+\ell+m=r \\ m\geq 1}} c_{k\ell m} \left(\text{sgn}(x)\right)^k \right.$$

$$\left. + \sum_{k+\ell=r} d_{k\ell}\, \ell n|x| \left(\text{sgn}(x)\right)^k \right] x^{-r} + o\left(|x|^{-N}\right)$$

where the coefficients $c_{k\ell}$, $c_{k\ell m}$ and $d_{k\ell}$ in this expansion are defined by equations (26) and (27) below and N is the least integer $\geq \mu+k+1$.

Proof This is based on the theory of asymptotic expansions of Fourier Transforms as developed in Lighthill [22] and Jones [20].

We will use the notation $\text{ft}_i(x)$ to denote the inverse Fourier Transform of $\text{cf}_i(s)$. Now, since the functions $\text{cf}_i(s)$ for i=1,2 do not lie in $L(-\infty,\infty)$, the $\text{ft}_i(x)$ cannot be defined in the usual way but do exist as generalized functions. In particular, the $\text{cf}_i(s)$ can be defined as generalized functions since there exists a G>0 for which $(1+s^2)^{-G}\text{cf}_i(s)\varepsilon L(-\infty,\infty)$[1]; the $\text{ft}_i(x)$ are then defined as the generalized functions obtained as the inverse Fourier Transforms of the generalized functions $\text{cf}_i(s)$[2] .

Starting with $\text{cf}_1(s)$ we write

$$\text{cf}_1(s) = \lim_{t\to 0+} e^{i\eta s} \sum_{m=0}^{M-1} p_m e^{-|s|t} (is)^m$$

and then, by definition,

---

1. See, for example, Lighthill [22], page 21
2. [22], page 18.

$$ft_1(x) = \lim_{t\to 0+} \left\{ \frac{1}{2\pi} \sum_{m=0}^{M-1} p_m \int_{-\infty}^{\infty} e^{-isx} e^{i\eta s} e^{-|s|t} (is)^m ds \right\}$$

$$= \frac{1}{2\pi} \sum_{m=0}^{M-1} p_m (-1)^m \lim_{t\to 0+} \left\{ \frac{d^m}{dx^m} \int_{-\infty}^{\infty} e^{-i(x-\eta)s} e^{-|s|t} ds \right\}$$

$$= \frac{1}{2\pi} \sum_{m=0}^{M-1} p_m (-1)^m \lim_{t\to 0+} \left\{ \frac{d^m}{dx^m} \left[\int_{o}^{\infty} e^{\left[-i(x-\eta)-t\right]s} ds \right.\right.$$

$$\left.\left. + \int_{o}^{\infty} e^{\left[i(x-\eta)-t\right]s} ds \right] \right\}$$

$$(22) \qquad = \frac{1}{2\pi} \sum_{m=0}^{M-1} p_m (-1)^m \delta^{(m)}(x-\eta)$$

where $\delta(y)$ is the Dirac delta function and $\delta^{(m)}(y)$ its m'th derivative. We deduce the asymptotic behavior of $ft_1(x)$ as $x\to\infty$ immediately from (22) as

$$(23) \qquad ft_1(x) = o(x^{-k})$$

for any value of $k>0$.

The second component is

$$cf_2(s) = e^{i\eta s} \sum_{k=0}^{K} \{q_{k0}|s|^{\mu+k} + q_{k1}|s|^{\mu+k} \ell n|s|\}$$

$$= e^{i\eta s} \sum_{k=0}^{K} \left[\left(q_{k0} + q_{k1} \frac{\partial}{\partial \alpha} \right) |s|^{\alpha} \right]_{\alpha=\mu+k}$$

$$= \lim_{t\to 0+} e^{i\eta s} \sum_{k=0}^{K} \left[\left(q_{k0} + q_{k1} \frac{\partial}{\partial \alpha} \right) |s|^{\alpha} e^{-|s|t} \right]_{\alpha=\mu+k}$$

On inversion, we obtain

$$ft_2(x) = \lim_{t\to 0+} \left\{ \frac{1}{2\pi} \sum_{k=0}^{K} \left[\left(q_{k0} + q_{k1} \frac{\partial}{\partial\alpha} \right) \int_{-\infty}^{\infty} e^{-isx} e^{i\eta s} e^{-|s|t} |s|^{\alpha} ds \right]_{\alpha=\mu+k} \right\}$$

$$= \lim_{t\to 0+} \left\{ \frac{1}{2\pi} \sum_{k=0}^{K} \left[\left(q_{k0} + q_{k1} \frac{\partial}{\partial\alpha} \right) \left\{ \int_{0}^{\infty} e^{-(iy+t)s} s^{\alpha} ds \right.\right.\right.$$

$$\left.\left.\left. + \int_{0}^{\infty} e^{-(-iy+t)s} s^{\alpha} ds \right) \right]_{\substack{\alpha=\mu+k \\ y=x-\eta}} \right\}$$

$$= \frac{1}{2\pi} \sum_{k=0}^{K} \left[\left(q_{k0} + q_{k1} \frac{\partial}{\partial\alpha} \right) \left(\lim_{t\to 0+} \Gamma(\alpha+1) \left\{ (t+iy)^{-\alpha-1} + (t-iy)^{-\alpha-1} \right\} \right) \right]_{\substack{\alpha=\mu+k \\ y=x-\eta}}$$

$$= \frac{1}{2\pi} \sum_{k=0}^{K} \left[\left(q_{k0} + q_{k1} \frac{\partial}{\partial\alpha} \right) \Gamma(\alpha+1) |y|^{-\alpha-1} \left\{ e^{-\frac{i\pi}{2} \operatorname{sgn}(y)(\alpha+1)} + e^{\frac{i\pi}{2} \operatorname{sgn}(y)(\alpha+1)} \right\} \right]_{\substack{\alpha=\mu+k \\ y=x-\eta}}$$

$$= \frac{1}{2\pi} \sum_{k=0}^{K} \left[\left(q_{k0} + q_{k1} \frac{\partial}{\partial\alpha} \right) 2\Gamma(\alpha+1) \cos\{\tfrac{1}{2}\pi(\alpha+1)\} |y|^{-\alpha-1} \right]_{\substack{\alpha=\mu+k \\ y=x-\eta}}$$

$$= \frac{1}{\pi} \sum_{k=0}^{K} \left[\Gamma(\alpha+1) \cos\{\tfrac{1}{2}\pi(\alpha+1)\} \left\{ q_{k0} + \left(\psi'(\alpha+1) - \tfrac{1}{2}\pi \tan\{\tfrac{1}{2}\pi(\alpha+1)\} \right.\right.\right.$$

$$\left.\left.\left. - \ell n|y| \right) q_{k1} \right\} |y|^{-(\alpha+1)} \right]_{\substack{\alpha=\mu+k \\ y=x-\eta}}$$

where $\psi'(z) = \Gamma'(z)/\Gamma(z)$, the logarithmic derivative of the Gamma function [ ].

Thus, the asymptotic behavior of $ft_2(x)$ as $x\to\infty$ is given by the series

$$(24) \quad ft_2(x) = \frac{1}{\pi} \sum_{k=0}^{K} \sum_{\ell=0}^{\infty} \left[\Gamma(\alpha+1) \cos\{\tfrac{1}{2}\pi(\alpha+1)\} \left\{ q_{k0} + \left(\psi'(\alpha+1) - \tfrac{1}{2}\pi \tan\{\tfrac{1}{2}\pi(\alpha+1)\} \right.\right.\right.$$

$$\left.\left.\left. - \ell n|x| + \sum_{m=1}^{\infty} \frac{1}{m} \left(\frac{\eta}{x}\right)^m \right) q_{k1} \right\} \frac{(\alpha+1)_{\ell}}{\ell!} \left(\frac{\eta}{x}\right)^{\ell} |x|^{-(\alpha+1)} \right]_{\alpha=\mu+k}$$

where $(\alpha+1)_\ell = \Gamma(\alpha+1+\ell)/\Gamma(\alpha+1)$.

The third component of the characteristic function is $cf_3(s)$. Now $ft_3(x)$ is the inverse Fourier Transform of $cf_3(s)$ so that $(-ix)^N ft_3(x)$ has the inverse Fourier Transform $cf_3^{(N)}(s)$. By Assumption 2, $cf_3^{(N)}(s)$ is absolutely integrable over $(-\infty,\infty)$ and it follows from the Riemann Lebesgue lemma that $(-ix)^N ft_3(x) = o(1)$ as $x\to\infty$. This last result together with (23) and (24) imply that as $x\to\infty$

$$pdf(x) = ft_1(x) + ft_2(x) + ft_3(x)$$

$$(25) \qquad = \frac{1}{\pi|x|^{\mu+1}} \sum_{k=0}^{K} |x|^{-k} \left[q'_{k0} + q'_{k1} \left\{ \sum_{m=1}^{\infty} \frac{1}{m} \left(\frac{\eta}{x}\right)^m - \ell n|x| \right\} \right] \sum_{\ell=0}^{\infty} \frac{(\mu+k+1)_\ell}{\ell!} \left(\frac{\eta}{x}\right)^\ell$$

$$+ o(|x|^{-N})$$

$$(26) \qquad = \frac{1}{\pi|x|^{\mu+1}} \sum_{r=0}^{K} \left[\sum_{k+\ell=r} \frac{q'_{k0}(\mu+k+1)_\ell \eta^\ell}{\ell!} \left(\mathrm{sgn}(x)\right)^k \right.$$

$$+ \sum_{\substack{k+\ell+m=r \\ m\geq 1}} \frac{q'_{k1}(\mu+k+1)_\ell \eta^{\ell+m}}{\ell! m} \left(\mathrm{sgn}(x)\right)^k$$

$$\left. - \sum_{k+\ell=r} \frac{q'_{k1}(\mu+k+1)_\ell \eta^\ell}{\ell!} \ell n|x| \left(\mathrm{sgn}(x)\right)^k \right] x^{-r} + o(|x|^{-N})$$

where

$$q'_{k0} = \Gamma(\mu+k)\cos\{\tfrac{1}{2}\pi(\mu+k+1)\}\left\{ q_{k0} + \left(\psi'(\mu+k+1) - \tfrac{1}{2}\pi \tan\{\tfrac{1}{2}\pi(\mu+k+1)\} \right) q_{k1} \right\} ,$$

$$q'_{k1} = \Gamma(\mu+k)\cos\{\tfrac{1}{2}\pi(\mu+k+1)\} q_{k1}$$

and N is the least integer $\geq \mu+k+1$. We rewrite (26) in the form

$$(27)\quad \frac{1}{\pi|x|^{\mu+1}} \sum_{r=0}^{K} \left[\sum_{k+\ell=r} c_{k\ell} \Big(\operatorname{sgn}(x)\Big)^k + \sum_{\substack{k+\ell+m=r \\ m\geq 1}} c_{k\ell m} \Big(\operatorname{sgn}(x)\Big)^k \right.$$

$$\left. - \sum_{k+\ell=r} d_{k\ell} \ell n|x| \Big(\operatorname{sgn}(x)\Big)^k \right] x^{-r} + o(|x|^{-N})$$

and this establishes the result. □

To illustrate the use of Theorem 4, we take the simple example of the Cauchy distribution with $cf(s) = e^{-|s|}$. In this case

$$cf_1(s) = 1$$

$$cf_2(s) = |s| \sum_{k=0}^{K} \frac{(-1)^{k+1}}{(k+1)!} |s|^k$$

$$cf_3(s) = \frac{|s|^{k+2}}{(k+2)!} e^{-\Theta|s|} \quad , \qquad 0<\Theta<1$$

and we deduce from (26) by setting $\mu = 1$, $\eta = 0$ and $q_{kl} = 0$ that

$$pdf(x) = \frac{1}{\pi x^2} \sum_{r=0}^{K} \frac{(-1)^{r+1}\Gamma(r-2)\cos\{\frac{1}{2}\pi(r+2)\}}{(r+1)!} \Big(\operatorname{sgn}(x)\Big)^r x^{-r} + o(|x|^{-K-2})$$

$$= \frac{1}{\pi x^2} \sum_{r=0}^{K} (-1)^{r+1} \cos\{\tfrac{1}{2}\pi(r+2)\} |x|^{-r}$$

Now $\cos\{\frac{1}{2}\pi(r+2)\}$ equals $(-1)^{\frac{1}{2}r+1}$ when r is an even integer and zero when r is odd. Hence, setting $r=2n$ and $K=2$, we have

$$pdf(x) = \frac{1}{\pi x^2} \sum_{n=0}^{N} (-1)^n (x^2)^{-n} + o\Big((x^2)^{-(N+1)}\Big)$$

This expansion can be verified directly from the density function $pdf(x) = [\pi(1+x^2)]^{-1}$ itself.

Theorem 5 If $s^N cf(s)$ is absolutely integrable over $(-\infty,\infty)$ then the local expansion of $pdf(x)$ about the point $x = d_i$ is given by

$$(28) \quad pdf(x) = \sum_{j=0}^{N-1} \beta_{ij}(x-d_i)^j + O(x-d_i)^N$$

where

$$\beta_{ij} = \frac{1}{2\pi j!} \int_{-\infty}^{\infty} (-is)^j e^{-isd_i} cf(s)ds \qquad j = 0,1, \ldots ,N-1.$$

Proof Since $s^N cf(s) \varepsilon L(-\infty,\infty)$ we expand the exponential e^{-isx} in the inversion formula

$$pdf(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-isx} cf(s)ds$$

about the point $x = d_i$ and integrate term by term giving

$$\frac{1}{2\pi j!} \sum_{j=o}^{N-1} \int_{-\infty}^{\infty} (-is)^j e^{-isd_i} cf(s)ds(x-d_i)^j + \frac{1}{2\pi N!} \int_{-\infty}^{\infty} (-is)^N e^{-is\tilde{d}_i} cf(s)ds(x-d_i)^N$$

where $\tilde{d}_i$ lies on the line segment connecting x and d_i . □

Local expansions of the type discussed in this section for the tails and the body of the distribution can also be extracted under similar conditions for the distribution function. These expansions will be useful in the development of a corresponding theory of global approximation for the distribution function rather than the density and will be discussed in a later paper. Expansions of this type are already given for many of the common distributions in the statistical literature ([44]).

6. MULTIPLE-POINT PADÉ APPROXIMANTS

As discussed in the previous section, Padé approximants can be used to improve the convergence properties of local Taylor expansions and have the useful additional property that they frequently extend the domain over which these local expansions provide good approximations. This section will show how Padé approximants can be derived from the local density expansions (16) and (17). These approximants will provide a preliminary set of rational fractions. They can then be used directly as approximations to pdf(x) or modified so that they belong to the class of rational fractions (2) and have satisfactory global behavior. The question of modifying the preliminary rational fractions will be taken up in section 7.

We start by writing the density function in the form

$$(29) \quad pdf(x) = s(x)pdf_s(x) \qquad -\infty<x<\infty$$

where $s(x)$ is a real continuous function satisfying $s(x)>0$ over the entire real axis and $s(x)\to 0$ as $x\to\pm\infty$. This representation of pdf(x) reconciles with the class of rational fractions defined in (2) and allows us to accommodate information about the coefficient functions $t(|x|)$ and $w(x)$ which appear in the local density expansions (16) and (18). In many cases, $s(x)$ will represent the leading term in the multiple series representation of the density pdf(x) and, in such cases, $s(x)$ will usually be identical to $t(|x|)$ and $w(x)$. When this leading term in the density is unknown, a suitable alternative will be to set $s(x) = t(|x|)$ directly or some modified form of $t(|x|)$ which has the same asymptotic behavior but which is well behaved elsewhere on the real axis. If necessary, the

expansion (18) can then be adjusted to take account of this modification so that (17) is correct to the same number of terms.

If we write the local expansions of $pdf_s(x)$ in the form

$$(30) \quad pdf_s(x) \sim \alpha_0 + \alpha_1 x^{-1} + \alpha_2 x^{-2} + \alpha_3 x^{-3} + \alpha_4 x^{-4} + \ldots \qquad x \to \pm\infty$$

$$(31) \quad pdf_s(x) \sim \beta_{i0} + \beta_{i1}(x-d_i) + \beta_{i2}(x-d_i)^2 + \beta_{i3}(x-d_i)^3 + \beta_{i4}(x-d_i)^4 + \ldots \quad (i=1, \ldots, I), \; x \to d_i$$

our problem is to construct a rational fraction of the form

$$(32) \quad [n/n] = \frac{P_n(x)}{Q_n(x)} = \frac{a_0 + a_1 x + \ldots + a_n x^n}{b_0 + b_1 x + \ldots + b_n x^n} \qquad n = \text{an even integer}$$

which has the same local behavior as (30) and (31) and to as high an order as possible. Such a rational fraction is called a multiple-point Padé approximant and is discussed by Baker [4, ch. 8][1]

The equations which define (32) can best be introduced by considering the approximant based on the Taylor series about a single point. We take the case of (31) with $d_i = 0$ and normalize (32) by setting $b_0 = 1$. This normalization ensures that $Q_n(0) = 1 > 0$ so that the $[n/n]$ approximant will not have a pole at the origin (this now being the point of expansion of the Taylor series (31)). The coefficients of $[n/n]$ are now determined by the equation

$$(33) \quad pdf_s(x) Q_n(x) - P_n(x) = 0\left(x^{2n+1}\right)$$

---

1. This work together with [5] and the collection of articles in [38] provide a systematic coverage of the extensive literature on the theory of Padé approximants and their applications, particularly in mathematical physics.

Explicitly, we have the relations

$$
\begin{aligned}
&\beta_{i0} && && && && && = a_0 \\
&\beta_{i1} + \beta_{i0}b_1 && && && && && = a_1 \\
&\beta_{i2} + \beta_{i1}b_1 + \beta_{i0}b_2 && && && && && = a_2 \\
&\beta_{i3} + \beta_{i2}b_1 + \beta_{i1}b_2 + \beta_{i0}b_3 && && && && && = a_3 \\
&\quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad \ldots \quad . \quad . \quad . \quad . \\
&\beta_{in} + \beta_{in-1}b_1 + \beta_{in-2}b_2 + \beta_{in-3}b_3 + \ldots + \beta_{i0}b_n && && && && && = a_n \\
&\beta_{in+1} + \beta_{in}b_1 + \beta_{in-1}b_2 + \beta_{in-2}b_3 + \ldots + \beta_{i1}b_n && && && && && = 0 \\
&\quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad \ldots \quad . \quad . \quad . \quad . \\
&\beta_{i2n} + \beta_{i2n-1}b_1 + \beta_{i2n-2}b_2 + \beta_{i2n-3}b_3 + \ldots + \beta_{in}b_n && && && && && = 0
\end{aligned}
\tag{34}
$$

which comprise $2n+1$ equations in the $2n+1$ required coefficients of $[n/n]$ viz $\{a_0, a_1, \ldots, a_n; b_1, \ldots, b_n\}$. Baker [4][1] has proved that, although a solution to (34) does not necessarily exist for all positive integers n there is an infinite subsequence $\{n_j\}$ for which the Padé approximant $[n_j/n_j]$ exists for any formal power series with $\beta_{i0} \neq 0$. Further, when the approximant $[n/n]$ exists, it is unique.[2]

We see, by inspection of (34), that in order to compute the coefficients of $[n/n]$ we need the coefficients in the local expansion (31) to order $2n$. Even for low values of n, this is likely to become prohibitive quite quickly when dealing with the distribution of an econometric statistic because of the increasing difficulty in extracting higher order coefficients and the complications of the resulting formulae. Moreover, in view of the smoothness of most density functions, in practice there will be little advantage to be gained from increasing the order of contact at a particular point past $n = 3$

1. Theorem 2.4 of [4].
2. Theorem 1.1 of [4]

or 4. In many cases, $n = 2$ will be sufficient to provide a highly satisfactory local density approximant.[1]

Multiple point expansions provide an excellent means of enabling us to reduce the order of contact at individual points to within manageable limits while extending the domain over which the final approximant will perform well. Thus, a two point Padé approximant $[n/n]$ might be based on the first $n+1$ equations of (34) which require local expansion coefficients up to β_{in} and a corresponding set of n equations with expansion coefficients up to order $n-1$ for a point other than the origin. These equations will then yield an approximant with contact of order n at the origin and $n-1$ at the second point.

If one of the points of local expansion is infinity, then the equations take on a slightly different form. In this case, rather than (33) we require

$$(35) \quad pdf_s(x)Q_n(x) - P_n(x) = 0\left(x^{-2n-1}\right)$$

as $x \to \infty$. We then have the following explicit relations from (30), (32) and (35)

$$
\begin{aligned}
&\alpha_0 b_n && && && && = a_n \\
&\alpha_1 b_n && + \alpha_0 b_{n-1} && && && = a_{n-1} \\
&\alpha_2 b_n && + \alpha_1 b_{n-1} && + \alpha_0 b_{n-2} && && = a_{n-2} \\
(36)\quad &. && . && . && . \ldots . && . \\
&\alpha_n b_n && + \alpha_{n-1} b_{n-1} && + \alpha_{n-2} b_{n-2} && + \ldots + \alpha_0 && = a_0 \\
&\alpha_{n+1} b_n && + \alpha_n b_{n-1} && + \alpha_{n-1} b_{n-2} && + \ldots + \alpha_1 && = 0 \\
&. && . && . && . \ldots . && . \\
&\alpha_{2n} b_n && + \alpha_{2n-1} b_{n-1} && + \alpha_{2n-2} b_{n-2} && + \ldots . && . = 0
\end{aligned}
$$

---

1. These issues will be taken up in greater depth in a later paper.

As with (34), this is a system of $2n+1$ equations in the same number of unknown coefficients.

In the general case of expansions about arbitrary points d_i as in (31), the equations which determine the coefficients take the form

$$(37)\qquad \left\{\sum_{j=0}^{\infty} \beta_{ij}(x-d_i)^j\right\} Q_n\left[d_i + (x-d_i)\right] - P_n\left[d_i + (x-d_i)\right] = 0\left[(x-d_i)^{2n+1}\right]$$

$$i = 1, \ldots, I$$

If we write $y = x-d_i$ and expand $Q_n(d_i+y)$ and $P_n(d_i+y)$ as

$$Q_n(d_i+y) = \sum_{k=0}^{n} b_k(d_i+y)^k = \sum_{k=0}^{n} b_k^{(i)} y^k = Q_n^{(i)}(y)$$

$$P_n(d_i+y) = \sum_{k=0}^{n} a_k(d_i+y)^k = \sum_{k=0}^{n} a_k^{(i)} y^k = P_n^{(i)}(y)$$

we have

$$(38)\qquad b^{(i)} = K^{(i)} b \quad , \quad a^{(i)} = K^{(i)} a$$

where

(39)

$$K^{(i)} = \begin{bmatrix} 1 & d_i & d_i^2 & d_i^3 & \cdots & d_i^n \\ 0 & 1 & 2d_i & 3d_i^2 & \cdots & \binom{n}{1} d_i^{n-1} \\ 0 & 0 & 1 & 3d_i & \cdots & \binom{n}{2} d_i^{n-2} \\ \cdot & \cdot & \cdot & \cdot & \cdots & \cdot \\ 0 & 0 & 0 & 0 & \cdots & 1 \end{bmatrix}, b^{(i)} = \begin{bmatrix} b_0^{(i)} \\ b_1^{(i)} \\ \cdot \\ \cdot \\ b_n^{(i)} \end{bmatrix} , \; b = \begin{bmatrix} b_0 \\ b_1 \\ \cdot \\ \cdot \\ b_n \end{bmatrix}$$

$$a^{(i)} = \begin{bmatrix} a_o^{(i)} \\ a_1^{(i)} \\ \cdot \\ \cdot \\ a_n^{(i)} \end{bmatrix} \quad , \quad a = \begin{bmatrix} a_o \\ a_1 \\ \cdot \\ \cdot \\ a_n \end{bmatrix} \quad .$$

and (37) becomes

$$(40) \qquad \left(\sum_{j=0}^{\infty} \beta_{ij} y^j\right) Q_n^{(i)}(y) - P_n^{(i)}(y) = 0\left(y^{2n+1}\right)$$

which is of the same form as (33) but in the transformed coefficients.

To work in terms of the original coefficients we can use the transformation matrix $K^{(i)}$ in the case of the vector $a^{(i)}$ as in (38) but in view of the normalization on the vector b we partition $K^{(i)}$ and b as follows

$$K^{(i)} = \left[\begin{array}{c|c} 1 & \underset{\sim}{k}^{(i)\prime} \\ \hline 0 & K_{22}^{(i)} \end{array}\right] \quad , \quad k^{(i)\prime} = \left[d_i, d_i^2, d_i^3, \ldots, d_i^n\right]$$

$$b' = \left[b_0, b_*'\right] = \left[1, b_*'\right]$$

We then have from (38)

$$b_0^{(i)} = k^{(i)\prime} b_* \quad , \quad b_*^{(i)} = K_{22}^{(i)} b_*$$

We now define

$$G^{(i)} = \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 \\ -\beta_{i0} & 0 & 0 & \cdots & 0 \\ -\beta_{i1} & -\beta_{i0} & 0 & \cdots & 0 \\ -\beta_{i2} & -\beta_{i1} & -\beta_{i0} & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ -\beta_{in-1} & -\beta_{in-2} & -\beta_{in-3} & \cdots & -\beta_{i0} \end{bmatrix} \quad , \quad g^{(i)} = \begin{bmatrix} \beta_{i0} \\ \beta_{i1} \\ \vdots \\ \beta_{in} \end{bmatrix}$$

The first $n+1$ equations of (34) can be written as

$$a + G^{(i)} b_* = g^{(i)}$$

or, in the general case of a local expansion about the point d_i (not necessarily the origin)

$$(41) \quad a^{(i)} + G^{(i)} b_*^{(i)} = g^{(i)} + g^{(i)} b_0^{(i)}$$

Transforming back to the original coefficients in the rational fraction we get

$$K^{(i)} a + G^{(i)} K_{22}^{(i)} b_* = g^{(i)} + g^{(i)} k^{(i)\prime} b_*$$

or

$$(42) \quad K^{(i)} a + \left[G^{(i)} K_{22}^{(i)} - g^{(i)} k^{(i)\prime} \right] b_* = g^{(i)}$$

The system of equations (42) holds for each point of local expansion i.e. for $i = 1, \ldots, I$ in our original notation (see (31)).

Note that as we have constructed (42) the system involves $n+1$ equations. In practice, it may be convenient to use fewer equations at each point, thus reducing the order of contact of the Padé approximant at each point and requiring less analytic information about the expansion coefficients. The procedure allows us to make up for this reduction in the number of equations at each point by increasing the number of points we use in developing the approximant. This process has the additional advantage of improving the global nature of the final approximation.

In the general case we let $K^{(i)}$ be $(m_i+1) \times (n+1)$ $G^{(i)}$ be $(m_i+1) \times m_i$, $K_{22}^{(i)}$ be $m_i \times n$, $g^{(i)}$ be $(m_i+1) \times 1$ and $k^{(i)}$ be $n \times 1$.

The complete system of equations which determine the coefficients in the $[n/n]$ Padé approximant are then based on (36) and (42) and take the following general form:

(43)

$$
\begin{array}{c}
\overset{\leftarrow\ n+1\ \rightarrow}{} \quad \overset{\longleftarrow\quad n\quad\longrightarrow}{} \\
\left[\begin{array}{c|c}
K^{(1)} & G^{(1)}K_{22}^{(1)} - g^{(1)}k^{(1)\prime} \\
\hline
K^{(I)} & G^{(I)}K_{22}^{(I)} - g^{(I)}k^{(I)\prime} \\
\hline
0_{rx(n+1)} & \begin{array}{cccc}
-\alpha_{r+1} & -\alpha_{r+2} & \cdots & -\alpha_{n+r} \\
-\alpha_{r} & -\alpha_{r+1} & \cdots & -\alpha_{n+r-1} \\
\vdots & \vdots & \cdots & \vdots \\
-\alpha_{2} & -\alpha_{3} & \cdots & -\alpha_{n+1}
\end{array} \\
\hline
I_{n+1} & \begin{array}{cccc}
-\alpha_{1} & -\alpha_{2} & \cdots & -\alpha_{n} \\
-\alpha_{0} & -\alpha_{1} & \cdots & -\alpha_{n-1} \\
0 & -\alpha_{0} & \cdots & -\alpha_{n-2} \\
\vdots & \vdots & \cdots & \vdots \\
0 & 0 & \cdots & -\alpha_{0}
\end{array}
\end{array}\right]
\left[\begin{array}{c} a_0 \\ a_1 \\ \vdots \\ a_n \\ \hline b_1 \\ b_2 \\ \\ \\ b_n \end{array}\right]
=
\left[\begin{array}{c} g^{(1)} \\ \hline g^{(I)} \\ \hline 0 \\ \vdots \\ 0 \\ \hline -\alpha_0 \\ 0 \\ \vdots \\ 0 \end{array}\right]
\end{array}
\begin{array}{l}
\uparrow \\ m_1 \\ \downarrow \\
\uparrow \\ m_I \\ \downarrow \\
\uparrow \\ \\ \\ \\ \downarrow \\
\uparrow \\ m_\infty \\ \downarrow
\end{array}
$$

where $m_\infty = n+r+1$. In (43) we need to select I , the m_i and m_∞ in such a way that $\Sigma_{i=1}^{I} m_i + m_\infty \geq 2n+1$. In the final block of equations in (43) we will often select $r = 0$ so that $m_\infty \leq n+1$. As with the case of the local expansions (31) about the points d_i this will reduce the number of final expansion coefficients that are required to solve for the $[n/n]$ Padé coefficients. In some cases (the application in section 8 turns out to be such a case) we may have more than enough coefficients, so that

$\Sigma_{mi} + m_\infty > 2n+1$ and we may neglect some equations of (43) to obtain a solution. When this happens it would seem preferable to neglect those equations which refer to higher order points of contact of the Padé approximant with the local expansions than those which refer to lower order points of contact.

7. MODIFYING THE PADÉ APPROXIMANT

By solving (43) for the Padé coefficients we obtain a preliminary rational fraction $[n/n]$ as in (32). This can be used to construct an approximation to pdf(x) of the form $s(x)[n/n] = s(x)P_n(x)/Q_n(x)$. In some cases, this will turn out to be a perfectly satisfactory approximant. In others, it will need to be modified to produce a good approximation to pdf(x) over a wide interval. This is because there is nothing in the procedure outlined in section 6 which prevents the occurrence of zeros in the polynomial $Q_n(x)$ on the real axis. These zeros induce poles in the approximant and will need to be eliminated if the approximation is to perform well, unless the zeros appear in remote and irrelevant regions of the distribution. My experience to date suggests that the latter is not usually the case. The normal occurrence is for the procedure in section 6 to produce a preliminary approximant with either no poles at all or a pair of poles, at least one of which lies within the main body of the distribution.

In addition to unwanted poles, the $[n/n]$ Padé approximant may become zero at a finite number of points on the real axis. Since we will, in general, have $pdf(x)>0$ for all finite x we will normally wish to eliminate the zeros of the approximant unless they occur well outside the

region of interest in the distribution.

When poles occur, they are typically found in the bridging region between the points of local density expansion used in (43) to construct the Padé approximant. This suggests that an obvious way of helping to remove unwanted poles is to introduce an additional point of local expansion in (43), perhaps at the price of reducing the order of contact at another point. However, there is no guarantee that this method will eliminate poles and it has the disadvantage of requiring additional information about the distribution to be operational. The procedure we suggest below does not suffer from these disadvantages. It will eliminate the poles and it is sufficiently flexible to allow for additional information about the distribution to be incorporated at the time of modification, if that information is available.

Before we outline the procedure, it may be worth mentioning that the occurrence of unwanted poles in Padé approximants is a long standing problem. The presence of poles in the approximant is one of the reasons why it is difficult to prove general theorems about the convergence of Padé approximants to a given function as the degree of the approximant increases. Many of the general results that are available[1] concern the convergence of subsequences as $n \to \infty$. For an example of non-convergence, we can cite Chui [8] who proves that in the general family of entire functions there exists a function for which the sequence of $[n/n]$ Padé approximants is divergent everywhere in the whole complex plane except at the origin.

The procedure we suggest for modifying Padé approximants so that they are well behaved over the whole real axis is based on the following simple idea. If zeros of the numerator and denominator polynomials $P_n(x)$ and

---

1. See, for example, Part II of Baker [4].

$Q_n(x)$ occur on the real axis, they will occur in pairs since n is even. We then replace the real roots of the associated quadratic equations by complex conjugate pairs in such a way that we preserve, as far as possible, the known behavior of the function at the points of local expansion. Various degrees of sophistication are possible in the practical application of this method. In fact, as we will demonstrate in the application of section 8 even crude adjustments which preserve only contact of order one at the points of local expansion seem to work remarkably well. After we have adjusted the coefficients in the Padé approximant so that $P_n(x)>0$ and $Q_n(x)>0$ throughout the real axis we simply numerically integrate and rescale so that the area under the curve is unity.

Some of the principles involved in the method outlined in the last paragraph can be illustrated in the case of an approximant with $n = 4$. Let the $[n/n]$ Padé approximant extracted by the procedure of section 6 with points of local expansion at $x = 0$ and $x^{-1} = 0$ be given by

$$(44) \qquad [4/4](x) = \frac{\sum_{i=0}^{4} a_i x^i}{\sum_{i=0}^{4} b_i x^i} = \frac{a_4(x-\gamma_1)(x-\gamma_2)(x-\gamma)(x-\bar{\gamma})}{b_4(x-\delta_1)(x-\delta_2)(x-\delta)(x-\bar{\delta})}$$

where γ_1, γ_2 denote real zeros of the numerator and δ_1, δ_2 denote real zeros of the denominator. $(\gamma,\bar{\gamma})$ and $(\delta,\bar{\delta})$ are complex conjugate pairs. We start by rewriting (44) in the form

$$(45) \qquad [4/4](x) = \frac{a_4\{x^2 - (\gamma_1+\gamma_2)x + \gamma_1\gamma_2\}(x-\gamma)(x-\bar{\gamma})}{b_4\{x^2 - (\delta_1+\delta_2)x + \delta_1\delta_2\}(x-\delta)(x-\bar{\delta})}$$

We now propose to modify the coefficients of the quadratics in braces so that $[4/4](x)>0$ for all real x while retaining the same behavior as (45) in the

neighborhood of $x = 0$ and $x^{-1} = 0$. We, therefore, define the family of functions

$$(46) \qquad [4/4](x;\Theta) = \frac{a_4\{cx^2+dx+e\}(x-\gamma)(x-\bar{\gamma})}{b_4\{fx^2+gx+h\}(x-\delta)(x-\bar{\delta})}$$

where $\Theta' = (c,d,e,f,g,h)$ is a vector of real parameters to be chosen. To ensure equivalent local behavior in (45) and (46) we restrict our choice of Θ so that

(i) $c/f = 1$; and

(ii) $e/h = \gamma_1\gamma_2/\delta_1\delta_2$.

Now (i) will ensure that $[4/4](x;\Theta)\to a_4/b_4$ as $x\to\pm\infty$ and (ii) that $[4/4](x;\Theta)\to a_4\gamma_1\gamma_2|\gamma|^2/b_4\delta_1\delta_2|\delta|^2$ as $x\to 0$.

In most cases we find that the zeros (γ_1,γ_2) and (δ_1,δ_2) occur with the same sign patterns. This is because the zeros and singularities lie in the intervals between the points (here 0 and $\pm\infty$) of local expansion. If we take the case where both $\gamma_1\gamma_2>0$ and $\delta_1\delta_2>0$, our task is then to raise f and h from their original values in (45) so that the discriminant $g^2-4fh<0$. This will require proportional changes in c and e so that (i) and (ii) remain valid. Often these automatic changes in c and e will be sufficient to ensure that there are no zeros in (46). If they are not, some small adjustment in the value of d will normally suffice. There is an added advantage to adjusting the value of d in that simple hand calculations will show what adjustments in this parameter will improve the order of contact of (46) at the points of local expansion while preserving the desired global behavior of $[4/4](x;\Theta)>0$ for all x . Various other scenarios of parameter changes are possible but those we have illustrated should indicate some of the

relevant considerations and the ease with which they may be performed.

The family of rational fractions (46) based on Padé approximants have introduced extra flexibility in the approximating procedure. The idea is essentially to partially reparameterize a first stage Padé approximant so that we can achieve good global behavior by sacrificing some degree of contact at the points of local expansion. But with the new family of approximating rational fractions (46) we have the opportunity to adjust the parameters to take account of any additional information about the distribution that has not already been used in the equations (43) that define the original coefficients; perhaps less precise information based on, for example, Monte Carlo work with the same distribution.

An obvious alternative procedure for modifying the Padé approximant (45), but which I have not yet tried in application, is to use splines to bridge the intervals in which singularities and zeros occur. This method may be particularly useful in cases where zeros and singularities occur together in close proximity.

8. AN APPLICATION TO A SIMULTANEOUS EQUATIONS ESTIMATOR

We consider the single structural equation

$$(47) \quad y_1 = \beta y_2 + Z_1\gamma_1 + u$$

where y_1 and y_2 are vectors of T observations on two endogenous variables, Z_1 is a $T \times K_1$ matrix of observations on K_1 exogenous variables and u is a vector of random disturbances. The reduced form equations for y_1 and y_2 are

$$(48)\qquad [y_1 \vdots y_2] = [Z_1 \vdots Z_2] \begin{bmatrix} \pi_{11} & \pi_{12} \\ \pi_{21} & \pi_{22} \end{bmatrix} + [v_1 \vdots v_2]$$

where Z_2 is a $T \times K_2$ matrix of observations of K_2 exogenous variables excluded from (47). We assume that the usual standardizing transformations have been carried out so that (i) $T^{-1}Z'Z = I_K$ where $[Z = Z_1 \vdots Z_2]$, $k = k_1 + k_2$ and (ii) the rows of $[v_1 \vdots v_2]$ are independent and identically distributed normal vectors with zero mean and covariance matrix equal to the identity matrix. We also assume that (47) is identified so that $K_2 \geq 1$.

The two stage least squares estimator 2SLS of β in (47) is given by the ratio $\hat{\beta} = y_2'Ry_1/y_2'Ry_2$ where $R = Z_2Z_2'$. The exact density function of $\hat{\beta}$ is known to be the doubly infinite series [35]

$$(49)\qquad \text{pdf}(x) = \frac{1}{\beta\left(\frac{1}{2},\frac{K_2}{2}\right)(1+x^2)^{(K_2+1)/2}} \frac{e^{-\frac{\mu^2}{2}(1+\beta^2)}}{} \sum_{j=0}^{\infty} \frac{\left(\frac{K_2+1}{2}\right)_j}{\left(\frac{K_2}{2}\right)_j j!} \left\{\frac{\mu^2}{2}\frac{(1+\beta x)^2}{1+x^2}\right\}^j {}_1F_1\left[\frac{K_2-1}{2}, j + \frac{K_2}{2}; \frac{\mu^2\beta^2}{2}\right]$$

which depends on the three parameters β, K_2 and $\mu^2 = T\pi_{22}'\pi_{22}$.

The extensive tabulations in Anderson and Sawa [2] show that (49) may be adequately approximated by the asymptotic normal only when μ^2 is very large (the size of μ^2 required for the asymptotic distribution to provide an adequate approximation is itself contingent on the size of β and K_2. Even for moderate values of β and K_2, the computations in [2] show that extremely large values of μ^2 (well over 1,000) are required to secure a satisfactory approximation).

As discussed in the introduction, other approximations to (49) which

perform satisfactorily for a range of parameter values are the Edgeworth ([1], [2]) and saddlepoint [19] approximations. But, when μ^2 is small both these approximations become inadequate.

In Figures 1, 2 and 3 we illustrate the inadequacy of these three different methods of approximation in the case where $\beta = 0.6$, $\mu^2 = 4.0$ and $K = 4, 10$. This is rather

FIGURES 1, 2 AND 3 ABOUT HERE

an extreme case where μ^2 is very low (around the lower limit of μ^2 values found by Anderson et. al. [3] in their numerical computations of key parameters for actual econometric models). It has been chosen to test the adequacy of the new method of approximation discussed in the earlier sections of this paper specifically in a case where the existing methods break down.

Figures 4-8 detail the approximants obtained at each stage of the procedure outlined in the previous sections of the paper.[1] In the first stage of the procedure we need to select the coefficient function $s(x)$ as in (29) above. A crude choice would be $s(x) = [1+|x|^{K_2+1}]^{-1}$ since this has the same asymptotic behavior of $|x|^{-(K_2+1)}$, which is in this case the coefficient function in the tail expansion (16), and since this function is also well behaved elsewhere on the real line. An alternative and better choice is the leading term in the density expansion (49) viz

$$(50) \qquad s(x) = \left[B\left(\frac{1}{2}, \frac{K_2}{2}\right)(1+x^2)^{(K_2+1)/2}\right]^{-1}$$

This is, in fact, the pdf of $\hat{\beta}$ under the null hypothesis that $\beta = 0$ and $\pi_{22} = 0$.[2] As mentioned previously in the introduction, leading terms such as (50) in multiple series representations of density functions can usually be derived without much difficulty and will often be available even

1. We deal specifically with the case $\mu^2 = 4$, $k = 4$, $\beta = 0.6$. A later paper will detail more fully some numerical experience with modified Pade approximants.

2. c.f. Basmann [7].

in cases where an analytic form for the exact density has not been obtained. In the present case, a few elementary manipulations show that $\hat{\beta}$ takes the form of a standard normal variate divided by the square root of a chi square with K_2 degrees of freedom, with the numerator and denominator independent. The statistic $\hat{\beta}$ is, therefore, proportional to a t-variate with K_2 degrees of freedom, leading to a pdf of the form given by (50).

Writing $pdf(x) = s(x)pdf_s(x)$ as in (29) we then extract the local expansions (30) and (31) for $pdf_s(x)$. The expansions we use are for the tails $(x^{-1} = 0)$ and the origin $(x = 0)$. The coefficients that appear in (30) and (31) are given by

$$\alpha_0 = \sum_{j=0}^{\infty} \frac{\left(\frac{K+1}{2}\right)_j}{\left(\frac{K}{2}\right)_j j!} \left(\frac{\mu^2}{2}\right)^j W(K,j)$$

$$\alpha_1 = \sum_{j=1}^{\infty} \frac{\left(\frac{K+1}{2}\right)_j}{\left(\frac{K}{2}\right)_j} \frac{2\beta}{(j-1)!} \left(\frac{\mu^2}{2}\right)^j W(K,j)$$

$$\alpha_2 = \sum_{j=1}^{\infty} \frac{\left(\frac{K+1}{2}\right)_j}{\left(\frac{K}{2}\right)_j (j-1)!} \left(\frac{\mu^2}{2}\right)^j \left[(2j-1)\beta^2 - 1\right] W(K,j)$$

$$\alpha_3 = \sum_{j=1}^{\infty} \frac{\left(\frac{K+1}{2}\right)_j}{\left(\frac{K}{2}\right)_j} \frac{2\beta}{(j-1)!} \left(\frac{\mu^2}{2}\right)^j \left[\frac{(2j-1)(2j-2)}{3!} \beta^2 - j\right] W(K,j)$$

$$\alpha_4 = \sum_{j=1}^{\infty} \frac{\left(\frac{K+1}{2}\right)_j}{\left(\frac{K}{2}\right)_j (j-1)!} \left(\frac{\mu^2}{2}\right)^j \left[\frac{2(2j-1)(2j-2)(2j-4)}{4!} \beta^4 - j(2j-1)\beta^2 + \frac{(j+1)}{2}\right] W(K,j)$$

and

$$\beta_0 = \sum_{j=0}^{\infty} \frac{\left(\frac{K+1}{2}\right)_j}{\left(\frac{K}{2}\right)_j j!} \left(\frac{\mu^2\beta^2}{2}\right)^j W(K,j)$$

$$\beta_1 = \sum_{j=1}^{\infty} \frac{\left(\frac{K+1}{2}\right)_j}{\left(\frac{K}{2}\right)_j (j-1)!} \frac{\mu^{2j}\beta^{2j-1}}{2^{j-1}} W(K,j)$$

$$\beta_2 = \sum_{j=1}^{\infty} \frac{\left(\frac{K+1}{2}\right)_j}{\left(\frac{K}{2}\right)_j (j-1)!} \left(\frac{\mu^{2j}\beta^{2j-2}}{2^j}\right) (1-\beta^2)W(K,j)$$

$$+ \sum_{j=2}^{\infty} \frac{\left(\frac{K+1}{2}\right) 2\beta^{-2}}{\left(\frac{K}{2}\right)_j (j-2)!} \left(\frac{\mu^2\beta^2}{2}\right)^j W(K,j)$$

$$\beta_3 = \sum_{j=1}^{\infty} \frac{\left(\frac{K+1}{2}\right)_j (-j)}{\left(\frac{K}{2}\right)_j (j-1)!} \left(\frac{\mu^{2j}\beta^{2j-1}}{2^{j-1}}\right) W(K,j)$$

$$\beta_4 = \sum_{j=1}^{\infty} \frac{\left(\frac{K+1}{2}\right)_j}{\left(\frac{K}{2}\right)_j (j-1)!} \mu^{2j} \left[\frac{\beta^{2j}(j+1)}{2^{j+1}} - \frac{\beta^{2j-2}j}{2^j}\right] W(K,j)$$

$$+ \sum_{j=2}^{\infty} \frac{\left(\frac{K+1}{2}\right)_j}{\left(\frac{K}{2}\right)_j (j-2)!} \left(\frac{\mu^2}{2}\right)^j \left[\beta^{2j-2}j(j+1) + \tfrac{1}{2}\beta^{2j-4}\right] W(K,j)$$

where we have dropped the subscript on K_2 for convenience and where

$$W(K,j) = e^{-\frac{1}{2}\mu^2(1+\beta^2)}\,{}_1F_1\left(\frac{K-1}{2}\ ,\ j+\frac{K}{2}\ ;\frac{\mu^2\beta^2}{2}\right)$$

$$= e^{-\frac{1}{2}\mu^2}\,{}_1F_1\left(j+\tfrac{1}{2}\ ,\ j+\frac{K}{2}\ ;\ -\frac{\mu^2\beta^2}{2}\right)$$

by Kummer's transformation [42].

Figure 4 details the local density approximations to pdf(x) based on (30) and (31) with the coefficient function $s(x)$ as in (50). The approximations are good in the locality of the points of expansion, the origin and the tails, but start to deteriorate rapidly as we move out of the immediate vicinity. The right hand tail expansion seems particularly good.

Figure 5 shows the $[4/4]$ Padé approximant to the density (49). This has the form (in the notation of (44))

$$(51)\qquad R_{4,4}(x;s) = s(x)[4/4](x) = s(x)\,\frac{a_4(x-\gamma_1)(x-\gamma_2)(x-\gamma)(x-\bar{\gamma})}{b_4(x-\delta_1)(x-\delta_2)(x-\delta)(x-\bar{\delta})}\ .$$

where

$$a_4 = 4.533619 \qquad b_4 = 1.221628$$

$$\gamma_1 = -\ 1.158240 \qquad \delta_1 = -\ 3.567599$$

$$\gamma_2 = -\ 0.537379 \qquad \delta_2 = -\ 0.485485$$

$$\gamma,\bar{\gamma} = -\ 2.133352 \pm 0.732053i \qquad \delta,\bar{\delta} = 0.310396 \pm 0.613123i$$

As a first step approximant $R_{4,4}(x)$ is rather good, with problems occurring only in the left tail at the singularity $x = \delta_2$ (i.e. $x-\beta = \delta_2-\beta = -1.085845$ for $\beta = 0.6$) and at the two zeros $x = \gamma_2,\gamma_1$ (note that the second pole occurs outside the region of immediate interest in the distribution).

The next stage in the procedure is to modify the Padé approximant (51) along the lines suggested in section 7. We note that in the denominator the

quadratic $x^2 - (\delta_1+\delta_2)x + \delta_1\delta_2 = x^2 + 4.05344x + 1.7333$ has discriminant 9.497176 . To remove the real zeros we propose to replace this quadratic by

(52) $1.5x^2 + 4x + 3$

where we have raised the constant and the coefficient of x^2 and simply rounded the coefficient of x . According to the ideas outlined in section 7, we now need to proportionately adjust the coefficient of x^2 and the constant term in one of the quadratics in the numerator. We select the quadratic $x^2 - (\gamma_1+\gamma_2)x + \gamma_1\gamma_2 = x + 1.695619x + 0.622414$ which we need to modify in any case to remove the unwanted zeros of $R_{4,4}(x)$. Making the proportional adjustments recommended to this quadratic we get $1.5x^2 + 1.695619x + 1.077276$. This gives us the following modified Padé approximant after one change of coefficients

$$(53) \quad R^{(a)}_{4,4}(x) = s(x)\,\frac{a_4\{1.5x^2 + 1.695619x + 1.077276\}(x-\gamma)(x-\bar{\gamma})}{b_4\{1.5x^2 + 4x + 3\}(x-\delta)(x-\bar{\delta})}$$

This function is graphed in Figure 6 against the exact density. We see that the singularity and zero problems have been eliminated and the performance of the approximation is remarkably good. We note some reduction in the order of contact at the points of local expansion, particularly the origin (or taking into account the change of origin on the graph $x-\beta = -0.6$) .

As suggested in section 7, it is worthwhile to modify at least one of the remaining coefficients to improve the order of contact at the points of local expansion. Note that the success of this procedure can be measured against the original Padé approximant in the relevant localities so we do not need a graph of the exact density to do so. Comparing the Padé and modified Padé approximants in Figures 5 and 6 it is clear that the order of contact

of the modified Padé at the origin will be improved if we raise the derivative at this point ($x = 0$ i.e. $x-\beta = -0.6$) . This will be achieved by raising the coefficient of x in the quadratic in braces in the numerator of (53). We make a change in this coefficient from 1.69 to 2.0 giving the new modified Padé approximant

$$(54) \quad R^{(b)}_{4,4}(x) = s(x)\,\frac{a_4\{1.5x^2 + 2.0x + 1.077\}(x-\gamma)(x-\bar{\gamma})}{b_4\{1.5x^2 + 4x + 3\}(x-\delta)(x-\bar{\delta})} \quad .$$

This function is graphed in Figure 7. Even with the rather crude adjustments we have made (54) is really an exceptionally close approximation to the true density and is well behaved over the whole real axis. A final adjustment can be made by renormalizing so that the area under (54) is unity. The adjusted curve is displayed in Figure 8.

9. CONCLUSION

This paper has introduced a new technique of approximating probability density functions. The approximating functions belong to a family of rational fractions and are sufficiently flexible to be capable of producing good approximants to a very wide class of density functions. The theory developed in sections 3 and 4 indicates that this family of rational fractions contains approximants which are best in a well defined sense and which will perform well in reproducing the form of the exact density functions over the entire real axis. The practical procedure for finding good approximants in this family that is discussed in the paper is based on the use of multiple-point Padé approximants to construct global approximations from purely local information about the density. These multiple-point Padé

approximants are then modified to ensure that they have good global behavior and to incorporate any additional information that may be available concerning the density. The application in section 8 to an already well established test area for density approximations illustrates that the procedure can produce exceptionally good approximations even in cases where existing methods break down. Further refinement of the ideas laid out in section 7 on modifying the initial Padé approximant should lead to fine approximations which are very close to the best uniform approximants discussed in sections 2 and 3.

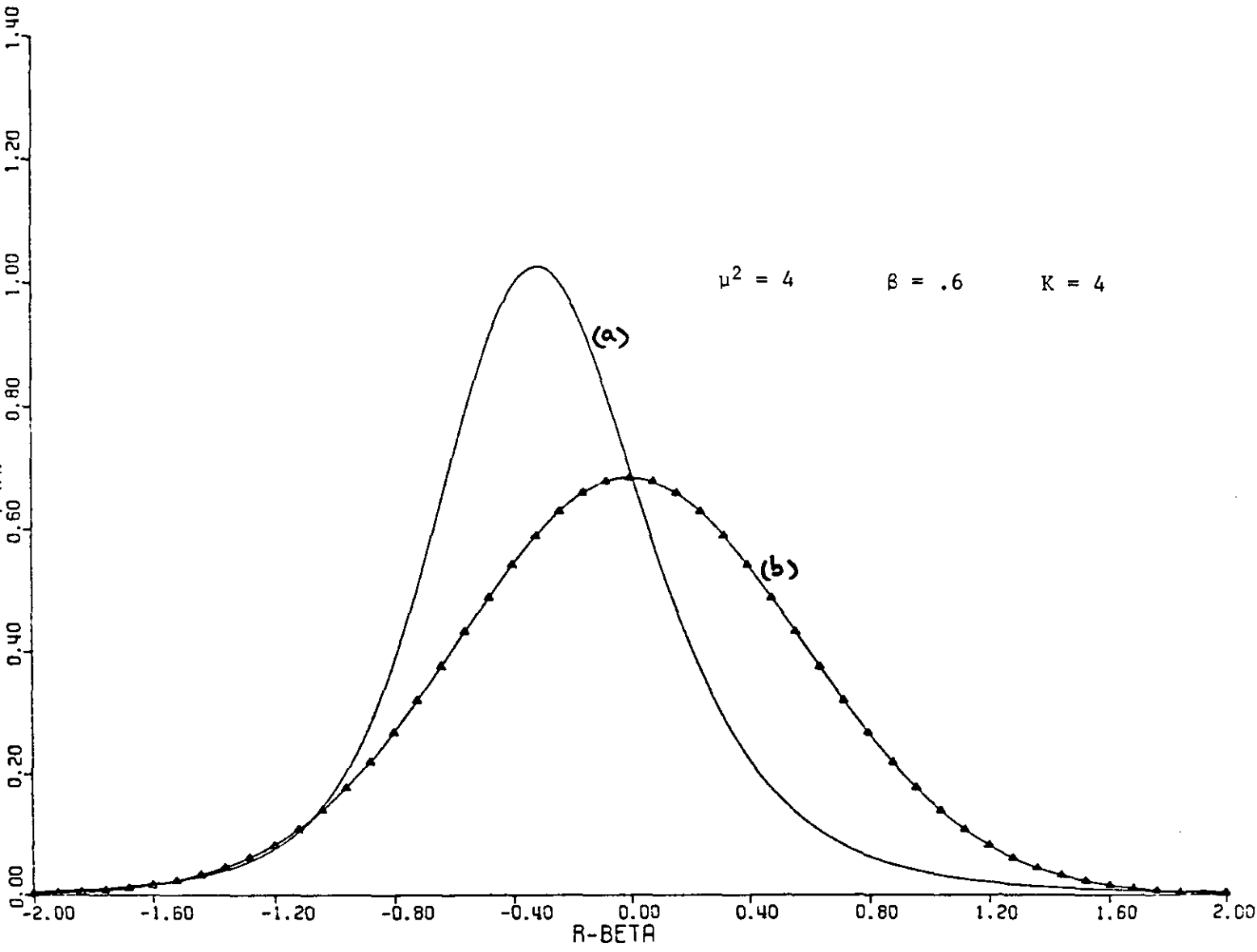

FIGURE 1A Asymptotic Normal Approximation

(a) exact density

(b) asymptotic normal

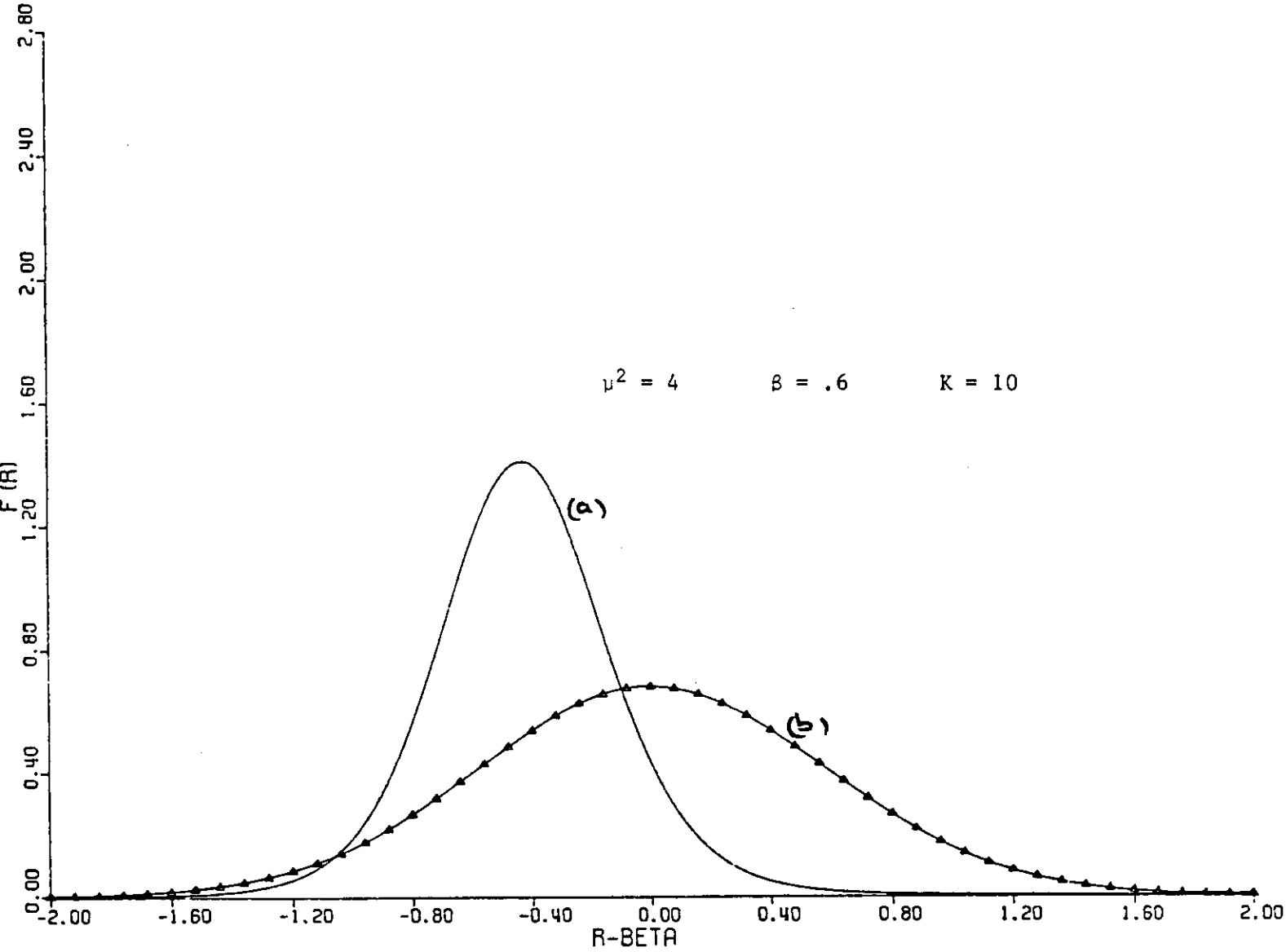

FIGURE 1B Asymptotic Normal Approximation

(a) exact density
(b) asymptotic normal

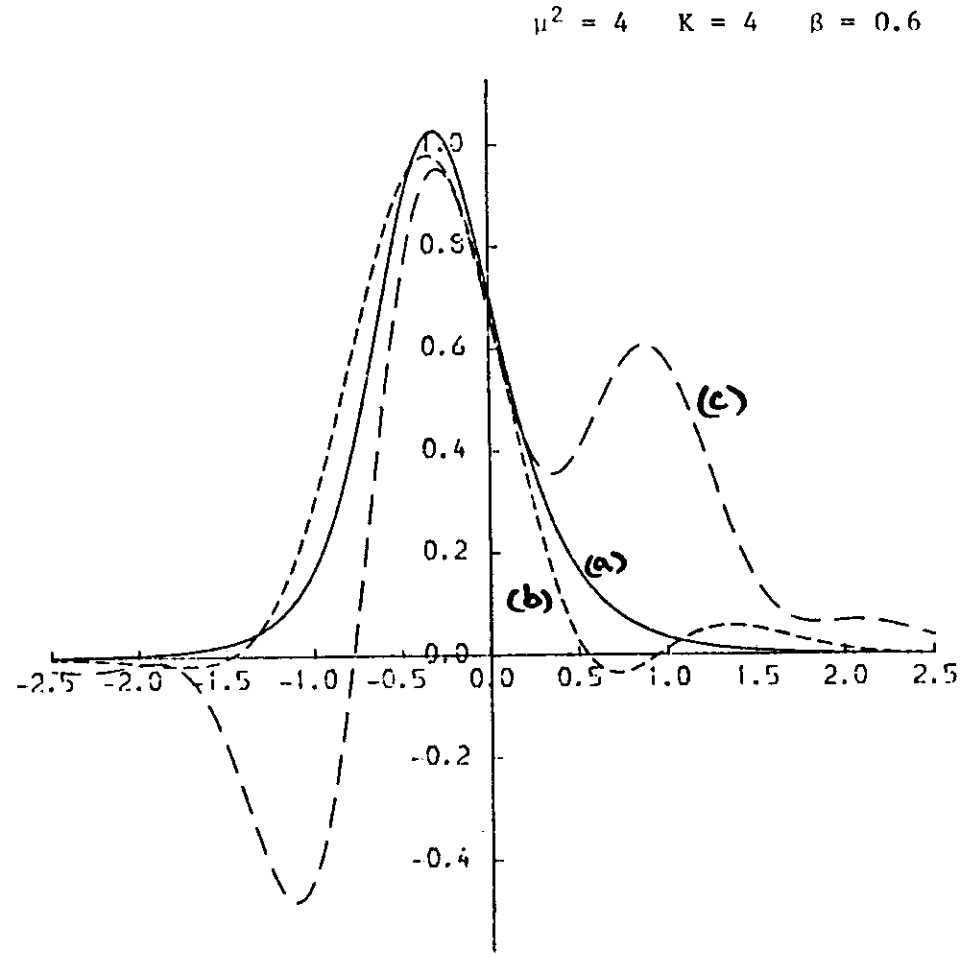

FIGURE 2A Edgeworth Approximation

(a) exact density
(b) $O(T^{-1/2})$
(c) $O(T^{-1})$

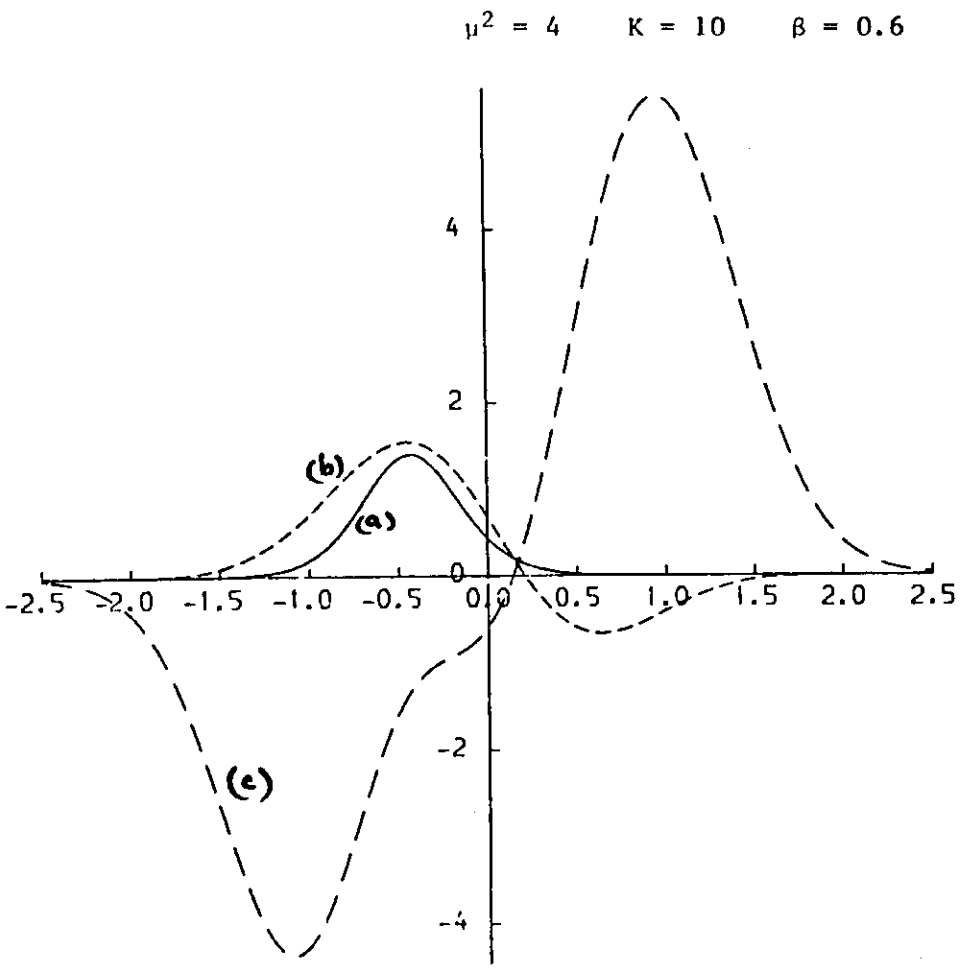

FIGURE 2B Edgeworth Approximation

(a) exact
(b) $O(T^{-1/2})$
(c) $O(T^{-1})$

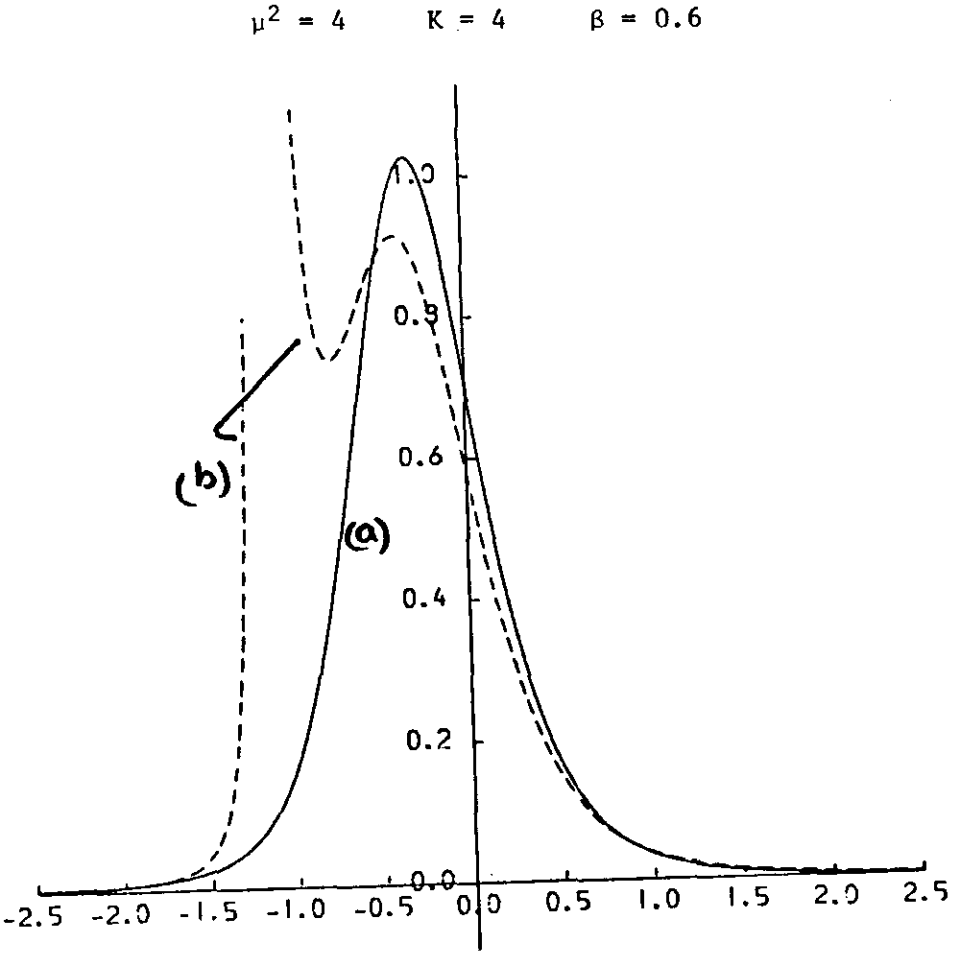

FIGURE 3A Saddlepoint Approximation

(a) exact density
(b) saddlepoint

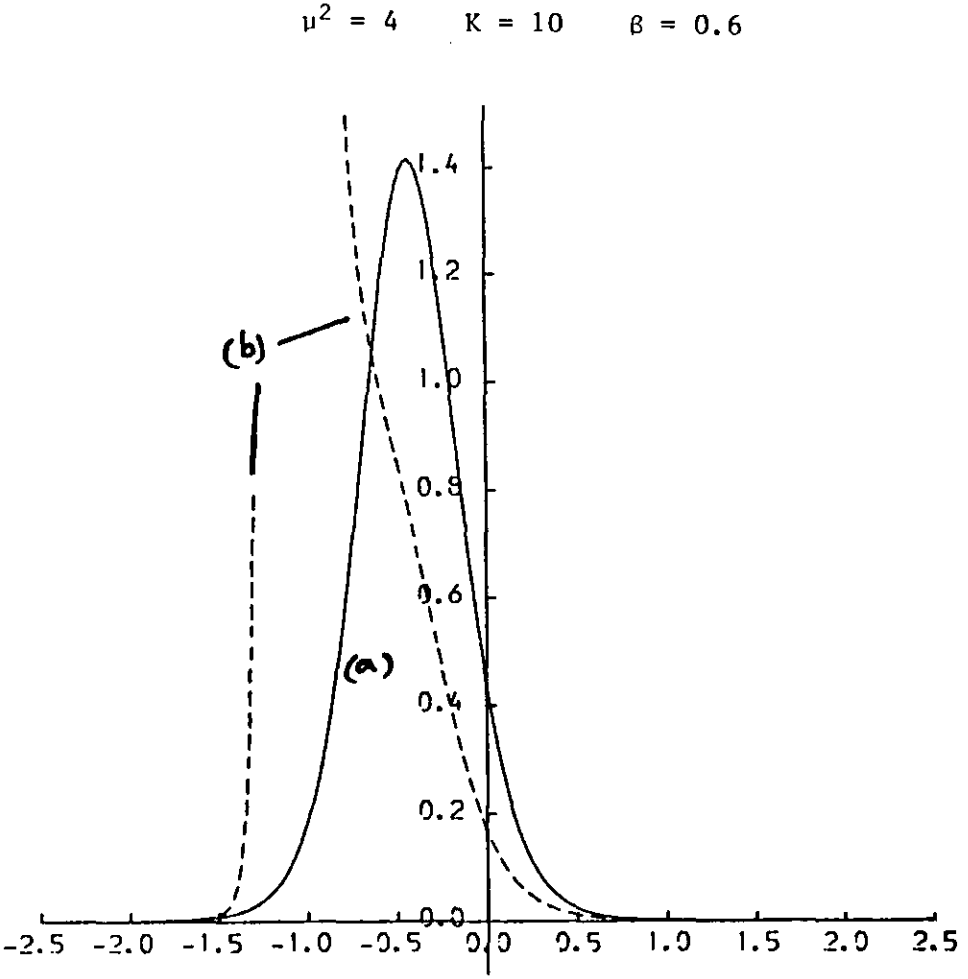

FIGURE 3B Saddlepoint Approximation

(a) exact density
(b) saddlepoint

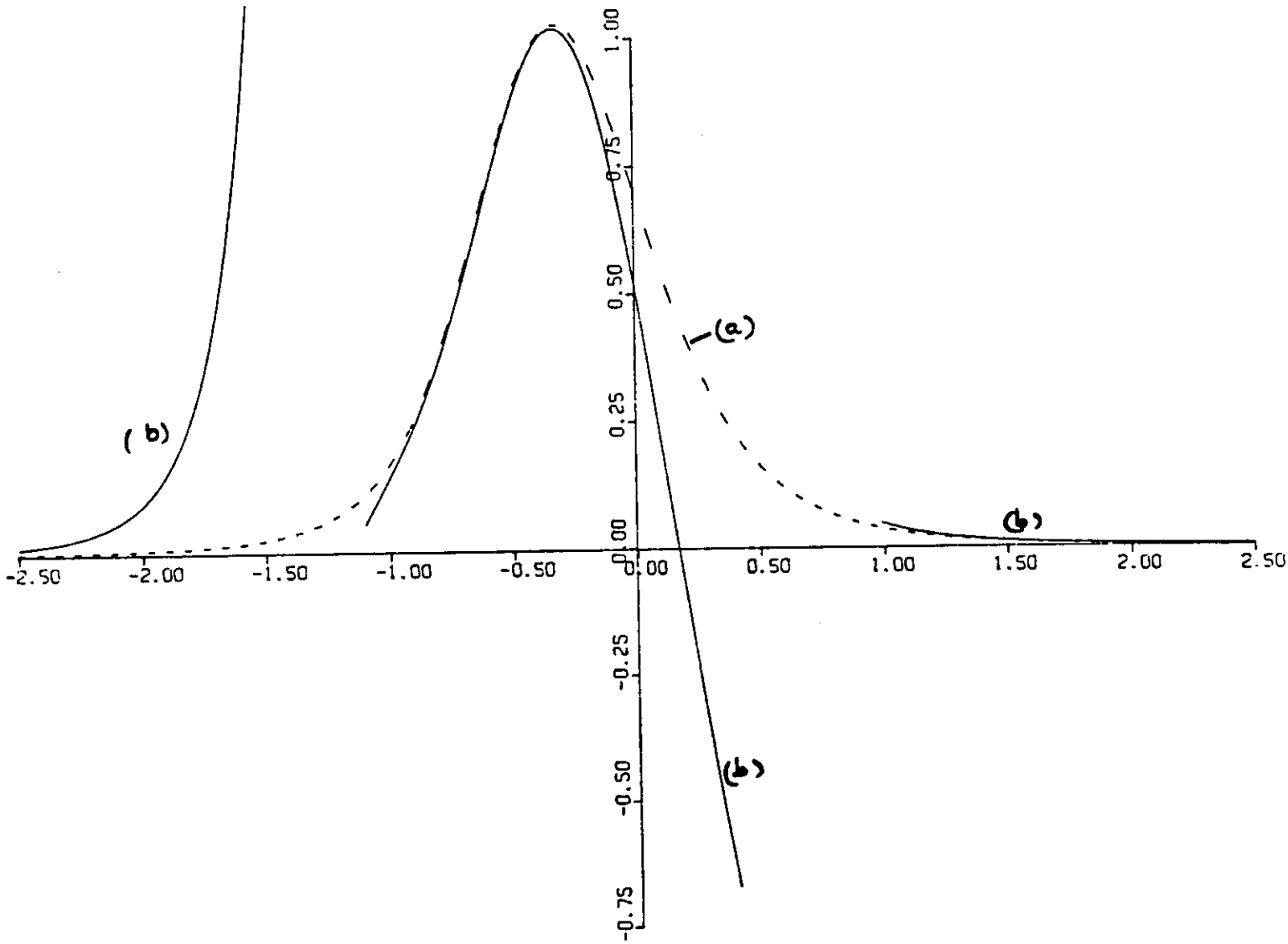

FIGURE 4 Local density approximations at the origin and in the tails to pdf(x)

(a) exact density

(b) local approximations

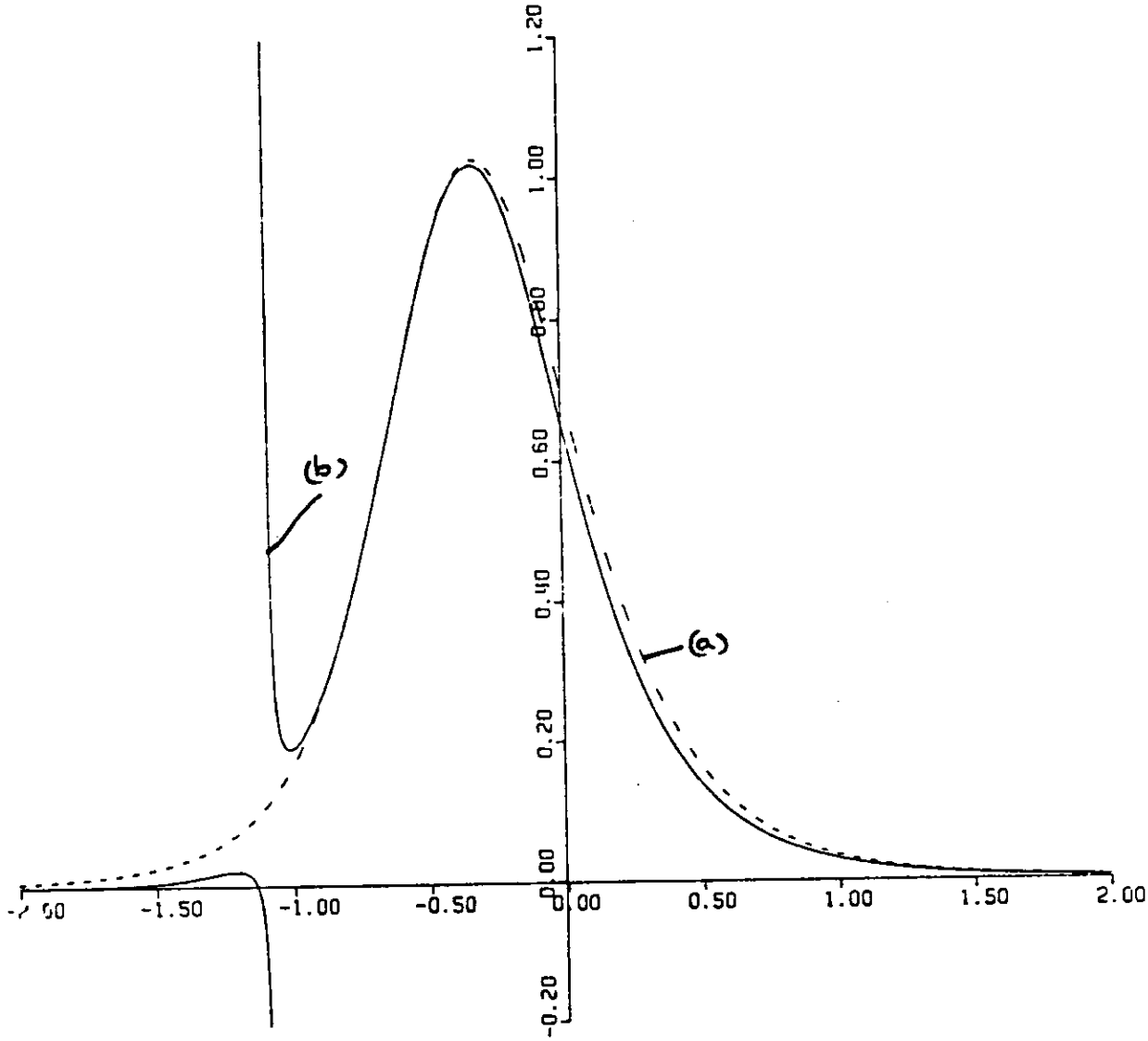

FIGURE 5 First step [4/4] Pade approximant to pdf(x)

(a) exact density

(b) Padé approximant

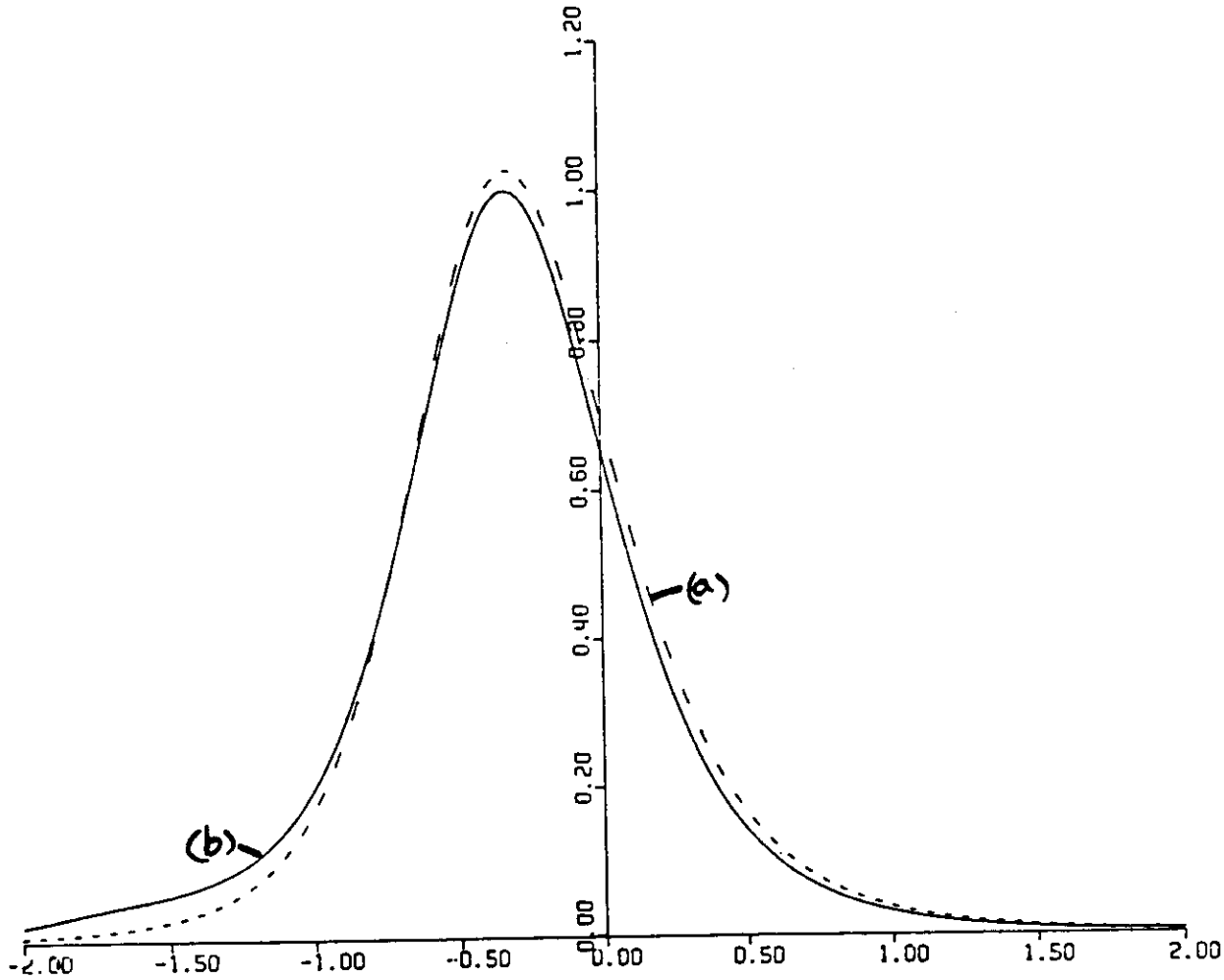

FIGURE 6 Modified Pade approximant to pdf(x): first change of coefficients

(a) exact density
(b) modified Padé approximant

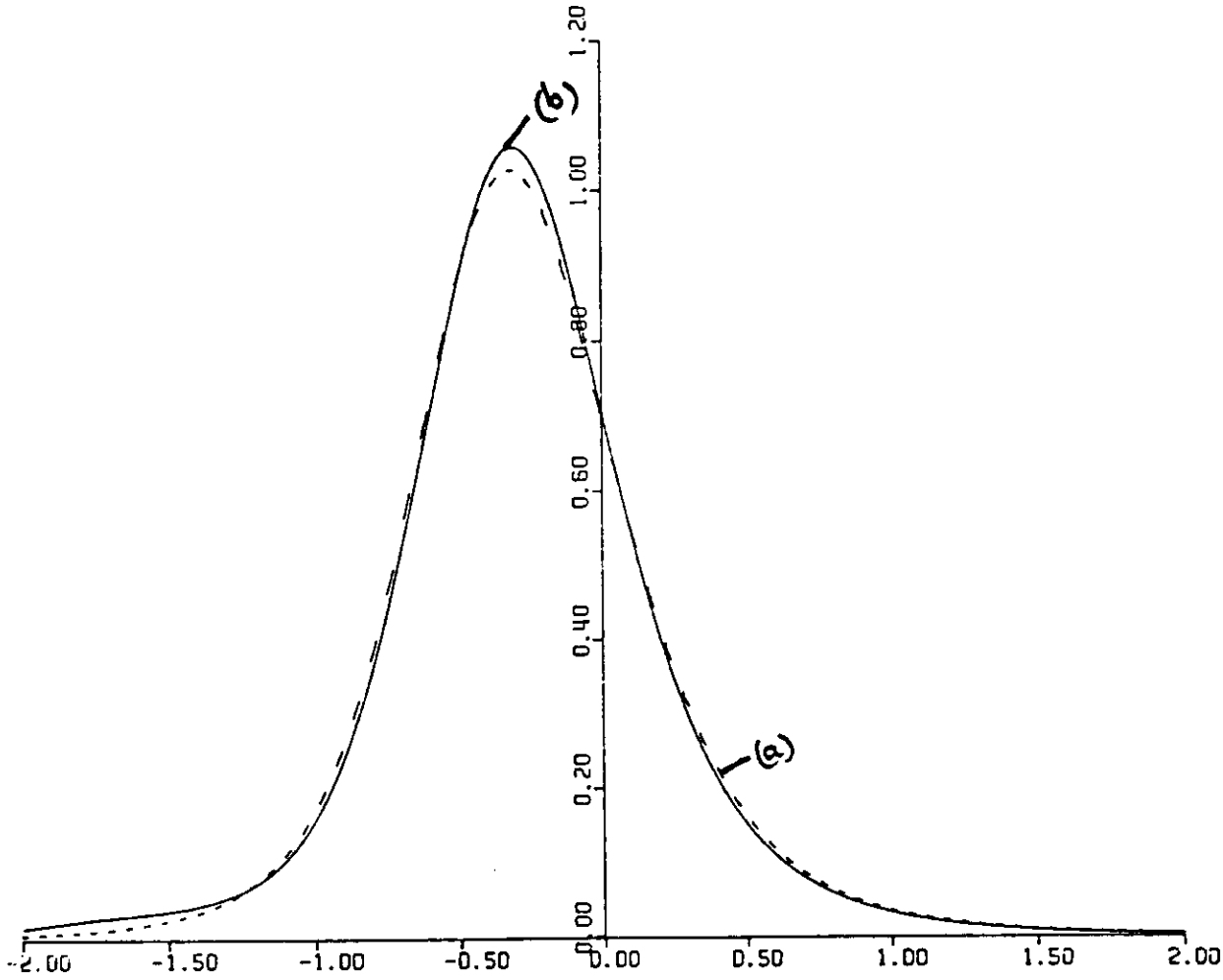

FIGURE 7 Modified Pade approximant to pdf(x): second change of coefficients

(a) exact density

(b) modified Padé approximant

REFERENCES

[1] Anderson, T. W. and T. Sawa, "Distribution of Estimates of Coefficients of a Single Equation in a Simultaneous System and Their Asumptotic Expansions," Econometrica, 41, 1973, pp. 683-714.

[2] --------, "Evaluation of the Distribution Function of the Two-Stage Least Squares Estimate," Econometrica, 47

[3] Anderson, T. W., K. Morimune, and T. Sawa, "The Numerical Values of Some Key Parameters in Econometric Models," Stanford University Technical Report No. 270 (IMISS).

[4] Baker, G. A., The Essentials of Padé Approximants, Academic Press, 1975.

[5] Baker, G. A. and J. L. Gammel, The Padé Approximant in Theoretical Physics, Academic Press, 1970.

[6] Barndorff-Nielson, O. and D. R. Cox, "Edgeworth and Saddlepoint Approximations with Statistical Applications," Journal of Royal Statistical Society, 1979.

[7] Basmann, R. L., "Exact Finite Sample Distributions and Test Statistics: A Survey and Appraisal," chapter 4 in M. D. Intriligator and D. A. Kendrick, Frontiers of Quantitative Economics, Vol. II, Amsterdam: North Holland, 1974.

[8] Chui, C. K., "Recent Results on Padé Approximants and Related Problems," in G. G. Lorenz et. al., Approximation Theory II, Academic Press, 1976.

[9] Daniels, H. E., "Saddlepoint Approximations in Statistics," Annals of Mathematical Statistics, 25, 1954, pp. 631-650.

[10] --------, "The Approximate Distribution of Serial Correlation Coefficients," Biometrika, 43, 1956, pp. 169-185.

[11] --------, "Exact Saddlepoint Approximations," Biometrika, 1980.

[12] Durbin, J., "The Approximate Distribution of Partial Serial Correlation Coefficients Calculated from Residuals from Regression on Fourier Series," London School of Economics, 1979.

[13] --------, "Approximations for Densities of Sufficient Statistics," London School of Economics, 1979.

[14] Erdéyli, A., Asymptotic Expansion, Dover, 1956.

[15] Feller, W., An Introduction to Probability Theory and Its Applications, Vol. II, Wiley, 1971.

[16] Frobenius, G., "Ueber relationem zwischen den näherungsbruchen von potenzreihen," Journal für die Reine und Angewandte Matematik, 90, 1881, pp. 1-17.

[17] Hart, J. F., Computer Approximations, Wiley, 1968.

[18] Hastings, C., Approximations for Digital Computers, Princeton, 1955.

[19] Holly, A. and P. C. B. Phillips, "A Saddlepoint Approximation to the Distribution of the k-Class Estimator in a Coefficient in a Simultaneous System," Econometrica, 47, 1979, pp. 1527-1548.

[20] Jones, D. S., Generalized Functions, McGraw Hill, 1966.

[21] Kendall, M. G. and A. Stuart, The Advanced Theory of Statistics, Vol. I, London: Griffin, 1969.

[22] Lighthill, M. J., An Introduction to Fourier Analysis and Generalized Functions, Cambridge University Press, 1958.

[23] Lukacs, E., Characteristic Functions, London: Griffin, 1970.

[24] Meinardus, G., Approximation of Functions: Theory and Numerical Methods, Vol. 13, Springer, 1967.

[25] Oberhettinger, F., Fourier Transforms of Distributions and Their Inverses, Academic Press, 1973.

[26] Padé, H., "Sur la representation approchee d'unefonction par des fractions rationelles," Annales Scientifiques de l'Ecole NormaleSuperievre, 9, 1892, pp. 1-93 (supplement).

[27] Phillips, P. C. B., "Approximations to Some Finite Sample Distributions Associated with a First Order Stochastic Difference Equation," Econometrica, 45, 1977a, pp. 463-486.

[28] -------, "A General Theorem in the Theory of Asymptotic Expansions as Approximations to the Finite Sample Distributions of Econometric Estimators," Econometrica, 45, 1977b, pp. 1517-1534.

[29] -------, "Edgeworth and Saddlepoint Approximations in the First-Order Non-Circular Autoregression," Biometrika, 65, 1978a, pp. 91-98.

[30] -------, "A Note on the Saddlepoint Approximation in a First Order Non-Circular Autoregression," Cowles Foundation Discussion Paper No. 487, Yale University, 1978b.

[31] -------, "Finite Sample Theory and the Distributions of Alternative Estimators of the Marginal Propensity to Consume," Review of Economic Studies, 47, 1980a, pp. 183-224.

[32] -------, "The Exact Finite Sample Density of Instrumental Variable Estimators in an Equation with n+1 Endogenous Variables," Econometrica, 48, 1980b, pp. 861-878.

[33] -------, "Approximating the Density Functions of Simultaneous Equations Estimators," (in preparation).

[34] Rice, J. R., The Approximation of Functions, Volumes I and II, Addison Wesley, 1964 and 1967.

[35] Richardson, D. H., "The Exact Distribution of a Structural Coefficient Estimator," Journal of the American Statistical Association, 63, 1968, pp. 1214-1226.

[36] Rudin, W., Real and Complex Analysis, McGraw Hill, 1974.

[37] Runge, C., "Zur theorie der ein deutigen analytischen funktioner," Acta Mathematica, 6, 1885, pp. 229-244.

[38] Saff, E. B. and R. S. Varga, Padé and Rational Approximation, Academic Press, 1977.

[39] Sargan, J. D., "Econometric Estimators and the Edgeworth Approximation," Econometrica, 44, 1976, pp. 421-428; "Erratum," Econometrica, 15, p. 272.

[40] -------, "On Monte Carlo Estimates of Moments Which Are Infinite," Advances in Econometrics, (forthcoming 1980).

[41] Sargan, J. D. and W. M. Mikhail, "A General Approximation to the Distribution of Instrumental Variable Estimators," Econometrica, 39, 1971, pp. 131-169.

[42] Slater, L. J., Confluent Hypergeometric Functions, Cambridge University Press, 1960.

[43] Tchebycheff, P. L., "Sur les questions de minima qui se rattachent a la représentation par la methode des moindres carrés," Oeuvres I, pp. 273-378.

[44] Walsh, J. L., "interpolation and Approximation by Rational Functions in the Complex Domain," American Mathematical Society, 1965 (originally published 1935).

[45] Zelen, M. and N. C. Severo, "Probability Functions," chapter 26 in M. Abramowitz and I. A. Stegun, Handbook of Mathematical Functions, Dover, 1965.