Review of Economic Studies (1980) XLVII, 183–224
© The Society for Economic Analysis Limited

0034-6527/80/00090183 $02.00

Finite Sample Theory and the Distributions of Alternative Estimators of the Marginal Propensity to Consume

P. C. B. PHILLIPS
Yale University and University of Birmingham

"One eminent statistician once said that a large part of the work in statistics involves being clever with approximations. I think this is what we need in sumultaneous equation estimation: clever approximations for some of these five-line long formulas and tables for non-standard distributions to help the people who do the real work in economics" Kmenta (1974, p. 275).

1. INTRODUCTION

In one of the pioneering papers of finite sample theory in econometrics Bergstrom[1] (1962) made a comparative study of the small sample behaviour of two alternative estimators of the marginal propensity to consume in the basic stochastic Keynesian model. The estimators he considered were least squares (OLS) and maximum likelihood, the latter being equivalent to two stage least squares (2SLS) in this just identified case.[2] At the time Bergstrom wrote the arguments for and against these estimators in small sample situations were largely unresolved, and the lack of evidence[3] one way or another provided ample motivation for the study. Bergstrom's comparison was based on the exact finite sample densities of the two estimators, which were derived mathematically on the hypothesis of a normally distributed structural disturbance on the consumption function and a non-random investment series. Numerical evaluation of these densities for various parameter values (including the sample size) then made it possible to tabulate the probabilities with which each estimator lay outside certain specified regions of the true value. These calculations led to the important conclusion that

> "... the use of the maximum likelihood estimator gives the greater probability of obtaining a very accurate estimate but the greater probability, also, of making a large error" (p. 487)

and that

> "... for samples of 10 or more observations, generated by the basic, stochastic Keynesian model (with realistic values of the parameters) the maximum likelihood estimator of the marginal propensity to consume is the 'better' general purpose estimator of this parameter" (p. 489).

Naturally, investigators are interested in the general question of which is the "better" general purpose estimator in a given situation with realistic sample sizes, where we interpret "better" in terms of the distribution of an estimator being more concentrated about the true value of the parameter. This question can be adequately dealt with only when an investigator has enough information about the sampling distributions of the estimators he is considering to assess their relative merits, at least for parameter values in the region within which he has to work. This assessment will necessarily involve an element of personal judgement by the investigator, taking into account the importance to him of errors of different magnitudes in estimates of the parameters in which he is interested. Ideally this could be measured by a well specified loss function, which would need to be constructed in the light of the intended applications of the estimated model and the consequences of sampling errors, perhaps even including the sign of such errors, in these intended applications.

In a stimulating survey paper Basmann (1974) has recently pointed to another reason why comparative studies of the distributions of various estimators are of value. Basmann argues from the standpoint of research economy. Detailed information on the small sample distributions of estimators can be used to help determine how accurate our data needs to be before we can confidently express a preference for one estimator or another in a given situation. This argument underlines our need for useful knowledge about the finite sample effects of misspecification; in this context, measurement error misspecification. In fact, the mathematical study of the effects of specification error on sampling distributions in econometrics has only recently commenced. We will discuss some of the results that have so far been obtained and the work that is under way in this particular area in Section 2 below.

For these and other important reasons it is of value to us to be informed about the small sample behaviour of commonly used econometric estimators and test statistics and to be guided by the results of comparative studies between various estimators in models which correspond as closely as possible to those encountered in empirical research. It therefore seems worthwhile to extend the range of realistic models to which our comparative studies of econometric estimators refer and to develop techniques and the associated computer software to enable an empirical researcher to extract information about the small sample behaviour of various estimators and statistics he may be considering for use and to do so explicitly in the context of the sample size, the particular model specification and the exogenous series with which he may be working. The present paper is concerned, in part, with such an aim. In Section 5 of the paper, we illustrate the approach by taking the Keynesian model used by Bergstrom extended by the introduction of an adjustment lag in the consumption function. A large proportion of empirical studies of the consumption function incorporate such a lag, and it is of interest in itself to consider the extent to which Bergstrom's conclusions carry over to this case. The resulting model involves the two complications of simultaneity and the presence of a lagged endogenous variable, but we maintain the simplifying assumption of a normally distributed and serially independent structural disturbance. Moreover, the approach we illustrate in this Section of the paper together with some of the computer software we have developed can be applied in more general econometric models.

The specific model we consider in Section 5 belongs to the class of models we refer to as being non-classical. By a classical model in the context of the conventional simultaneous equations framework we mean a model with normally distributed disturbances and non-random exogenous variables (this is the terminology introduced by Sargan (1976*a*)). Almost all the finite sample theory that has been developed in econometrics since the fundamental work of Basmann and Bergstrom has been founded substantially on these two rather limiting assumptions. On the other hand, these classical assumptions do enable us to derive a number of useful exact finite sample results. In Section 2 of the paper, we will briefly review some of the developments in this particular area that seem most relevant to the application we are considering later in the paper.[4] When we relax the classical assumptions and allow, in particular, for the presence of lagged endogenous variables in the regressor set, an exact theory is no longer within practical reach.[5]

However, it seems likely that good approximations to the small sample distributions of econometric estimators and test statistics can be obtained in such cases.[6] One approximation that offers interesting possibilities in this respect is based on the Edgeworth series expansion of the exact distribution and the resulting approximation (which we call the Edgeworth approximation) has already been applied with some success in dynamic models (Phillips (1977*a*)).

Many economists and econometricians may not yet have had the opportunity to familiarize themselves with the literature that has been emerging in this area. A further aim of the present paper, therefore, is to provide some background on the nature of the Edgeworth approximation and some associated asymptotic expansions and discuss the results that are now available for us to use (Sections 3 and 4 of the paper). I should mention at this point that the paper is not intended as a completely general survey. However, it is hoped that the less formal and more wide ranging discussion of the early sections of this paper will help to provide a useful introduction to this general area of current research. The sections that follow Section 4 then provide an application of the theory to the distribution of OLS and 2SLS estimators of the coefficients in a simple consumption function that involves lagged consumption in the set of regressors.

2. SOME DEVELOPMENTS IN EXACT FINITE SAMPLE THEORY IN ECONOMETRICS

Seven years after the publication of Bergstrom's paper, Sawa (1969) continued the comparison of OLS and 2SLS estimators, this time in the more general set up of a single equation with two endogenous variables (and possibly some exogenous variables) embedded in a system of structural equations.[7] Concentrating on the coefficient of the right hand side endogenous variable, Sawa derived the exact finite sample densities of the two estimators[8] under the classical assumptions of normally distributed disturbances and non-random exogenous variables. The numerical calculations of these densities that were published by Sawa supported the general conclusion reached by Bergstrom in the simpler model, and provide a good deal of additional insight into the way the shape of these densities respond to changes in the values of key parameters.[9]

The shape of the densities turns out (not unexpectedly in the case of OLS) to be most sensitive to the magnitude of the covariance between the right hand side endogenous variable and the structural disturbance: the larger is this parameter, the more serious is the bias of the OLS estimator compared with that of the 2SLS estimator. From the mathematical form of the densities, Sawa notes also that, in the case of the OLS estimator, moments of order less than $T-1$ exist (where T is the sample size) and those of higher order do not; whereas, in the case of the 2SLS estimator, moments of order less than K_2 exist (where K_2 is the number of excluded exogenous variables) and those of higher order do not.[10] Thus, moments of the 2SLS estimator in this context certainly exist up to the degree of overidentification (i.e. K_2-1).[11] The result means, also, that in models where T is a good deal larger than K_2, the OLS estimator will possess more moments and its distribution will, therefore, be characterized by tails which are thinner than those of the distribution of the 2SLS estimator. In the Keynesian model used by Bergstrom $K_2=1$, so that no integral moments of the maximum likelihood estimator exist, which explains the phenomenon of large errors having a greater probability of occurrence with the use of the maximum likelihood estimator than with OLS.

The papers by Richardson (1968) and Sawa (1969) dealt with a single structural equation containing two endogenous variables. The mathematical forms of the exact densities of the OLS and 2SLS estimators in this case are similar and can be represented as double infinite series. They can be simplified somewhat to single series by use of the confluent hypergeometric function (Lebedev (1965)) and the leading term in the series reveals the order to which moments exist. Otherwise, these series are not easy to interpret and are sometimes slow to converge (and hence are not as useful as might be expected for numerical computations). In many situations, we are more interested in the distribution functions than the densities and Anderson and Sawa (1973) obtain double series expressions for the former with terms involving incomplete beta integrals. Anderson and Sawa consider the distribution of k-class estimators (for non-stochastic k) in the same paper and, for these estimators, the general expression for the densities and distribution functions involve fourth order infinite series. In the same set up of a single equation containing two endogenous variables, Mariano and Sawa (1972) obtained the exact density of the limited information maximum likelihood estimator (LIML) of the coefficient of the right hand side endogenous variable and verified that this density does not possess moments of any integer order.[12]

Very little work has been published so far for estimators in structural equations containing more than two endogenous variables. Basmann *et al.* (1972) extract the exact joint density of 2SLS estimators in a just identified equation containing three endogenous variables. Basmann (1974, pp. 251–252) quotes a result due to Richardson for the same set up but with an arbitrary number of degrees of overidentification. In Basmann's notation this last result characterizes the subclass

$$\bigcup_{N=1}^{\infty} H_{2,N}$$

where $H_{n,N}$ denotes the joint distribution on $\mathbb{R}^n$ of the 2SLS estimators of the coefficients of the n right hand side endogenous variables in an equation with N degrees of overidentification. In some more recent work, Sargan characterizes the class $\bigcup_{n=1}^{\infty} H_{n,0}$ corresponding to a just identified equation containing $n+1$ endogenous variables (for any n). In Appendix A(ii), I give the results I have obtained elsewhere (Phillips (1978*b*)) which characterize

$$H = \bigcup_{n=1}^{\infty} \bigcup_{N=0}^{\infty} H_{n,N}$$

corresponding to an equation containing $n+1$ endogenous variables (for any n) and an arbitrary number of degrees of overidentification N. The expression for the joint density is an infinite series whose terms involve a matrix argument hypergeometric function (Herz (1955), James (1975), Johnson and Kotz (1972)). The leading term in this joint density is similar to a multivariate Cauchy distribution when $N=0$ and a multivariate t-distribution (Johnson and Kotz (1972, p. 134)) when $N>0$. From this expression we can verify directly Basmann's conjecture (Basmann (1961) and (1963*a*)) that integer moments of the 2SLS estimator exist up to the degree of overidentification (N). Mariano (1972) earlier verified this conjecture in the case of even order moments and Hatanaka (1973) did so for both odd and even moments. Ullah and Nagar (1974) have derived an expression for the exact mean of the 2SLS estimator in the special case of $n=2$ (i.e. for an equation containing 3 endogenous variables).[13]

The usual approach taken in deriving the mathematical form of the exact density functions of OLS and 2SLS estimators is to write these estimators first of all as functions of non-central Wishart matrices. When we concentrate on the coefficients of the right hand side endogenous variables these functions turn out to have a fairly simple form. A typical representation (compare (11) below and Mariano (1972)) takes the form $A_{22}^{-1}a_{21}$ where the matrix

$$A = \begin{bmatrix} a_{21} & a_{21}' \\ a_{21} & A_{22} \end{bmatrix}$$

has a non-central Wishart distribution of order $n+1$ (a_{11} is a scalar and A_{22} is an $n \times n$ matrix); the covariance matrix of this Wishart distribution is the covariance matrix of the endogenous variables included in the equation; the degrees of freedom depend on the estimator being considered; and the means sigma matrix (the matrix of non-centrality parameters) is a matrix quadratic form in the reduced form parameters. Modern methods of multivariate analysis (based in large part on the work of Herz (1955), James (1964) and Constantine (1963)) enable us to employ a convenient mathematical representation of the joint density of the matrix A above (in terms of matrix argument hypergeometric functions). To obtain the joint density function of the estimator we first need to transform variates so that we are working directly with the function $A_{22}^{-1}a_{21}$. Then we are left with an integration over the space of a_{11} and the matrix space of A_{22} in order to extract the required density. This integration turns out to be difficult in the case where $n > 1$. I give the final result for the density in Appendix A(ii). When $n = 1$ it turns out that an alternative approach based on contour integration can be used to extract the density and I outline the approach in Appendix A(i). The derivation of the distribution of the LIML estimator involves similar techniques. So far no work has appeared on the distribution of this estimator other than in the $n = 1$ case (Mariano and Sawa (1972), Anderson (1974)). However, it should not be difficult to extract a formula for the density in the case of general n and N using some recent extensions of the zonal polynomials of James (1964) and work on this is currently under way.

All of the above work deals with correctly specified models. Naturally, we are also interested in the effects of different types of misspecification on the form of the finite sample distributions of estimators and test statistics. Earlier, in the Introduction, we mentioned the importance of measurement error misspecification. We also wish to discover the effect of standard forms of specification errors, such as omitted variables (both exogenous and endogenous) and the inclusion of extraneous variables, on the small sample behaviour of different estimators. The first systematic study of this latter type of question has been done by Hale *et al.* (1978) and Rhodes and Westbrook (1977). The exact results of these authors deal with the case of two included endogenous variables. Their conclusions, based partly on an analysis of the bias and mean squared errors of the estimators, suggest that the comparative advantage held by 2SLS over OLS in correctly specified equations can be considerably weakened when the equation is misspecified; in a number of cases, the distribution of the OLS estimator was found to be the more concentrated about the true value of the parameter. Clearly, these results need to be borne in mind in making judgements about the appropriate use of estimators in simultaneous equations.

As suggested earlier, a major difficulty with multiple series representations of exact density functions and exact moment formulae is that they cannot be implemented for numerical calculations as easily as might be expected. Sometimes the series are slow to converge and in the more general cases mentioned above the formulae rely on matrix argument hypergeometric functions whose series representations are in terms of zonal polynomials (see James (1975) and Subrahmaniam (1976) for recent surveys of work in this area). These polynomials are currently tabulated up to order 12 and no general formulae for them have yet been derived. In many cases, the available tabulations of the polynomials will be insufficient to secure reliable numerical values for the density.[14]

Thus we may often need to rely on approximations of various types even when the exact formulae are available. In the next section we turn to examine the Edgeworth approximation which is currently attracting interest among econometricians and which seems capable of providing good approximations under certain conditions even in quite complicated models.[15]

3. AN INTRODUCTION TO ASYMPTOTIC EXPANSIONS AND SOME BACKGROUND OF THE EDGEWORTH APPROXIMATION[16]

The present section is intended as an introduction to the theory of asymptotic expansions of the Edgeworth type. As with other areas of research, it is helpful to see this theory in an historical context alongside the development of closely related (and, sometimes, better known) work. As the first step in our discussion we, therefore, select an appropriate setting for the theory. The most natural, in the present case, is the collection of limit theorems which together comprise what we commonly refer to as asymptotic theory. To help the reader maintain bearings in the ensuing discussion I have prepared a diagram (see Appendix B) illustrating certain aspects of asymptotic theory, with branches leading to refinements of central limit theory and useful asymptotic expansions which will be discussed more fully in the text. These asymptotic expansions form the basis of certain types of approximations we can use for small sample distributions in econometric work.

The Edgeworth approximation is based on a finite number of terms (usually the first two or three terms) of an asymptotic series expansion of the distribution function (or density) of the estimator of test statistic under consideration. We now regard this particular asymptotic expansion (or Edgeworth expansion, as it is called) as a refinement of the associated limit theorem which gives us the asymptotic distribution of the estimator or test statistic. But, this is not the way in which the series was first discovered. Historically, the Edgeworth series is closely related to another series which is known as the Gram–Charlier series (or, more precisely, the Gram–Charlier A-series). The idea behind the latter series is to represent the density $f_T(x)$ of some appropriately standardized statistic[17] as a linear combination of the standardized normal density and its successive derivatives. That is

$$f_T(x) = a_0 i(x) + \frac{a_1}{1!} i'(x) + \frac{a_2}{2!} i''(x) + \dots \qquad \dots(1)$$

where the a_r $(r = 0, 1, 2, \dots)$ are constant coefficients and $i(x) = (2\pi)^{\frac{1}{2}} \exp(-\frac{1}{2}x^2)$. The idea goes back to Tchebycheff ((1860), (1890)) as early as the mid-nineteenth century and was pursued later by Gram (1879) and Charlier (1905).[18] The derivatives of $i(x)$ have the simple representation

$$i^{(r)}(x) = \frac{d^r i(x)}{dx^r} = (-1)^r H_r(x) i(x) \qquad (r = 1, 2, \dots)$$

where $H_r(x)$ is a polynomial in x of degree r and the set $\{H_r(x) | r = 1, 2, \dots\}$ from an orthogonal set of polynomials (called Hermite polynomials) with respect to the normal distribution. In particular,

$$\int_{-\infty}^{\infty} H_r(x) H_s(x) i(x) dx = \begin{cases} r! & \text{when } s = r \\ 0 & \text{otherwise} \end{cases}$$

so that if we multiply both sides of (1) by $H_r(x)$ and, proceeding formally, integrate term by term on the right hand side we obtain

$$a_r = (-1)^r \int_{-\infty}^{\infty} H_r(x) f_T(x) dx. \qquad \dots(2)$$

Thus, the coefficients a_r $(r = 1, 2, \dots)$ become related to the moments of the underlying distribution. In fact, since $H_r(x)$ is a polynomial, it is clear that a_r will comprise a linear combination of the moments of the underlying distribution.

Of course, the operations leading to (2) are, as they stand, purely formal and ignore questions such as the convergence of the series (1). The latter question has, indeed, been investigated in the mathematical literature and sufficient conditions found. But, these impose severe restrictions on the tails of $f_T(x)$ and in practice exclude most distributions of interest.[19]

The fact that we cannot as a rule assert convergence of the series does not mean, however, that a finite number of terms will not provide a good approximation. And what the Edgeworth series succeeds in doing is to break down the coefficients a_r into components and then reassemble the series so that it is a proper asymptotic series in the mathematical sense. Such a series has the property that when we truncate the series after a finite number of terms the remainder has the same order of magnitude (usually taken in terms of powers of $T^{-\frac{1}{2}}$) as the first neglected term.[20] In the Edgeworth series, terms of the Gram–Charlier series are grouped according to their order of magnitude in terms of an underlying parameter such as $T^{-\frac{1}{2}}$. Let us take the underlying parameter to be $T^{-\frac{1}{2}}$; then the Edgeworth series has the form

$$f_T(x) = i(x)[1+\textstyle\sum_{j=1}^{\nu} P_j(x)T^{-j/2}]+R_{\nu T}(x) \qquad \ldots(3)$$

where $P_j(x)$ is a polynomial in x and $R_{\nu T}(x)$ is the remainder in the series after $\nu+1$ terms. The series given by the right side of (3) will then be an asymptotic series if $R_{\nu T}(x) = O(T^{-(\nu+1)/2})$. Sometimes, we can go further and say that the order of magnitude of $R_{\nu T}(x)$ holds uniformly in x; clearly, this is a stronger result.

To take an example of (1) and (3) let us go back to the simplest case of the standardized statistic

$$Z_T = \sqrt{T}\{T^{-1}(X_1+\ldots+X_T)-m\}/\sigma \qquad \ldots(4)$$

where the X_t ($t=1, \ldots, T$) are independent and identically distributed random variables with mean m, variance σ^2 and finite moments (and, hence, cumulants) up to order $\nu+2$. We denote by k_j the cumulant of X_t of order j so that the jth cumulant of $(X_t-m)/\sigma$ is

$$k'_j = k_j/\sigma^j \qquad (j=1, \ldots, \nu+2)$$

and the jth cumulant of Z_T is

$$\begin{aligned} k''_j &= k'_j/T^{(j/2)-1} \qquad (j=1, \ldots, \nu+2) \\ &= O(T^{-(j/2)+1}) \qquad \ldots(5) \end{aligned}$$

(see, for instance, Cramér (1946, p. 225)). If $f_T(x)$ is the density of Z_T the Gram–Charlier series (1) becomes in this case[21]

$$f_T(x) = i(x) - \frac{1}{3!}\frac{k_3}{T^{\frac{1}{2}}}i^{(3)}(x) + \frac{1}{4!}\frac{k'_4}{T}i^{(4)}(x) - \frac{1}{5!}\frac{k'_5}{T^{\frac{3}{2}}}i^{(5)}(x) + \frac{1}{6!}\left(\frac{k'_6}{T^2}+\frac{10(k'_3)^2}{T}\right)i^{(6)}(x)+\ldots \qquad \ldots(6)$$

By simple rearrangement of (6) we obtain the first few terms in the Edgeworth series as follows

$$\begin{aligned} f_T(x) &= i(x) - \frac{1}{3!}\frac{k'_3}{T^{\frac{1}{2}}}i^{(3)}(x) + \frac{1}{T}\left\{\frac{1}{4!}k'_4 i^{(4)}(x) + \frac{10}{6!}(k'_3)^2 i^{(6)}(x)\right\} \\ &\quad + \text{terms of higher order in } T^{-\frac{1}{2}} \\ &= i(x)\left[1+\frac{1}{T^{\frac{1}{2}}}\left\{\frac{k'_3}{3!}H_3(x)\right\} + \frac{1}{T}\left\{\frac{k'_4}{4!}H_4(x)+\frac{10}{6!}(k'_3)^2 H_6(x)\right\}\right] \\ &\quad + \text{terms of higher order in } T^{-\frac{1}{2}}. \qquad \ldots(7) \end{aligned}$$

Clearly, (7) has the same general form as (3). We note that $P_1(x)$ is a polynomial of degree 3 in x and $P_2(x)$ is a polynomial of degree 6; in general, $P_j(x)$ in (3) is a polynomial of degree $3j$. We note that (7) can be written in the alternative form

$$f_T(x) = \exp\left\{-\frac{1}{T^{\frac{1}{2}}}\frac{k'_3}{3!}D^3 - \frac{1}{T}\frac{k'_4}{4!}D^4+\ldots\right\}i(x), \quad D = d/dx \qquad \ldots(7')$$

and this is the form originally suggested by Edgeworth (1905).

By formal inspection of (6) and (7) we see that we may expect accuracy up to order T^{-1} with knowledge of cumulants of order six in (6) but only cumulants of order four in (7). Although Edgeworth derived the form of (7) in 1905, a rigorous demonstration of the order of magnitude of the remainder terms was not available until Cramér's papers appeared in 1925 and 1928. Cramér showed that (7) is indeed a proper asymptotic series and his results were made more readily available with the publication of his monograph on probability in 1937. Cramér also verified that the corresponding expansion for the distribution function of Z_T, obtained by formally integrating the series (7), is a proper asymptotic series. Simpler proofs of the results came later and one of the most lucid modern discussions is given by Feller (1970). Feller's proof of the validity of the expansion as an asymptotic series rests on two simple conditions. First, if $\phi_x(t)$ is the characteristic function of the component variates X_t, then

$$\int_{-\infty}^{\infty} |\phi_x(t)|^r dt < \infty$$

for some $r \geqq 1$. This condition ensures the existence of the density $f_T(x)$. The second condition requires that X_t possess cumulants up to order $\nu+2$ for the validity of the expansion up to $\nu+1$ terms. Thus, if we continue the development (7) up to $\nu+1$ terms, writing the expansion as in (3) the error $R_{\nu T}(x)$ on the first $\nu+1$ terms of the expansions is from Feller's result

$$R_{\nu T}(x) = \mathrm{o}(T^{-\nu/2}) \quad \text{uniformly in } x.$$

That is, $R_{\nu T}(x)$ tends to zero as $T \to \infty$ faster than $T^{-\nu/2}$. Note that we can now write (3) as

$$f_T(x) = i(x)[1 + \textstyle\sum_{j=1}^{\nu-1} P_j(x) T^{-j/2}] + R_{\nu-1T}(x) \qquad \ldots(8)$$

and $R_{\nu-1T}(x) = \mathrm{O}(T^{-\nu/2})$ so that the error has the same order of magnitude as the first neglected term.

The Edgeworth approximation to $f_T(x)$ is obtained by taking the first few terms in the series on the right side of (8). From the asymptotic nature of the series we know that the error on the Edgeworth approximation will tend to zero faster than the error on the normal approximation (the very first term of the series) as $T \to \infty$. While this suggests that the Edgeworth approximation may give useful results in many cases, it is important to stress that this theory covers only the *order of magnitude* of the error as $T \to \infty$; it says nothing about the *absolute magnitude* of the error in particular cases. For this reason, although the Edgeworth approximation is appealing in itself, the need for a theory which provides explicit bounds on the error has long been recognized. Unfortunately, only a few results have been obtained. The central result, which bounds the error on the normal approximation, was obtained by Berry and Esseen in the early 1940's (Berry (1941), Esseen (1945)).[22] This theorem tells us that if $F_T(x)$ is the distribution function of Z_T and $I(x)$ the distribution function of a standard normal variate, then

$$\sup_x |F_T(x) - I(x)| \leqq \frac{C\beta_3}{T^{1/2}\sigma^3} \qquad \ldots(9)$$

where β_3 is the third absolute moment of X_t and C is an absolute, universal constant. The inequality (9) has been shown to hold for values of C as low as 2·031.[23] The Berry–Esseen Theorem has been extended to cover the case of non-identically distributed variates (see Petrov (1975), chapter V) and has recently been generalized by Sazonov (1968), Bhattacharya (1975) and Sweeting (1977) to include multivariate distributions. As yet there are no extensions to cases of importance in econometrics, and this seems to be a worthwhile area of future investigation.[24]

On the other hand, the last seven or eight years have witnessed the emergence of a good deal of literature concerned with the use of Edgeworth type approximations in econometrics. The first published paper in the area is by Sargan and Mikhail in 1971 but the idea of using Edgeworth approximations in econometric work certainly goes back somewhat farther. At least two papers in the area were presented at the 1970 World Congress (Sargan (1970) and Mariano (1973)) and I have noticed an abstract by Sargan in *Econometrica* as early as 1964, which reports the application of Gram–Charlier series to the distribution of the 2SLS estimator. I say as early as 1964 because the development of this type of approximation theory in econometrics should be viewed in the light of corresponding developments in mathematical statistics. At this point in time, there was no theory in the statistical literature dealing with statistics whose moments may not exist, as is the case with the 2SLS estimator. In a survey paper, Wallace had suggested in 1958 that expansions could be constructed for quite general functions of sample moments,[25] but no rigorous theory was available. In 1967, Chambers published a paper which went a long way towards filling this gap. He developed Edgeworth expansions for multivariate statistics more general than standardized means, gave conditions for their validity and algorithms for their computation; he also derived expansions for quite general vector functions of other multivariate statistics (such as sample moments), and gave computational algorithms in this case as well.

A seminal paper in the area of econometric applications was published by Sargan in 1975. In this paper, Sargan proved a very general theorem on the validity of Edgeworth expansions for sample distributions of quite general estimators and statistics with limiting normal distributions (including all the usual simultaneous equations estimators and t-ratio test statistics). The approach taken in this paper was to write the error in an estimator, say $\hat{\beta}-\beta$, as a function of a more basic set of statistics comprising the errors in the sample moments of the data. Thus, writing $\hat{\beta}-\beta=e_T(p, w)$ in notation close to that of Sargan (1975), the vector of more basic statistics is partitioned into a subvector p of normally distributed variates, and a vector w, statistically independent of p. In addition to the normality requirement on p, Sargan's theorem imposes a smoothness and invertibility condition on the function $e_T(\cdot)$ and demands that $\sqrt{T}w$ have bounded moments of all orders. But, the important feature of this result is that whereas p and w have finite moments of all orders, no similar condition is placed on the error function $e_T(p, w)$ and, hence, $\hat{\beta}$. Thus, the theorem applies to the important cases of econometric estimators whose finite sample moments may exist only up to a certain order.

A simple example of the application of the Sargan result is given by the 2SLS estimator of the coefficient vector β in the single equation

$$y_1 = Y_2\beta + Z_1\gamma + u \qquad \ldots(10)$$

of a simultaneous equations model. $y_1(T\times 1)$ and $Y_2(T\times n)$ are an observation vector and observation matrix, respectively, of the included endogenous variables, Z_1 is a $T\times K_1$ matrix of observations of included exogenous variables and u is a normally distributed disturbance with zero mean and covariance matrix $\sigma^2 I$. If Z is the $T\times K$ matrix of observations of all K exogenous variables, then the 2SLS estimator of β is given by[26] (c.f. Mariano (1977, pp. 490–491))

$$\hat{\beta} = (Y_2'RY_2)^{-1}(Y_2'Ry_1) \qquad \ldots(11)$$

where $R = Z(Z'Z)^{-1}Z' - Z_1(Z_1'Z_1)^{-1}Z_1'$. We take the reduced form equations for y_1 and Y_2 as

$$y_1 = Z\pi_1 + v_1; \qquad Y_2 = Z\Pi_2 + V_2$$

and define

$$p_1 = \frac{Z'y_1}{T} - \left(\frac{Z'Z}{T}\right)\pi_1 = \frac{Z'v_1}{T}; \qquad P_2 = \frac{Z'Y_2}{T} - \left(\frac{Z'Z}{T}\right)\Pi_2 = \frac{Z'V_2}{T}$$

and a selection matrix S for which $Z_1 = ZS$ (i.e. S is a matrix which selects those columns of Z occurring in Z_1). Then

$$\hat{\beta} = [(P_2+M\Pi_2)'M^{-1}(P_2+M\Pi_2)-(P_2+M\Pi_2)'S(S'MS)^{-1}S'(P_2+M\Pi_2)]^{-1}$$
$$\cdot [(P_2+M\Pi_2)'M^{-1}(p_1+M\pi_1)-(P_2-M\Pi_2)'S(S'MS)^{-1}S'(p_1+M\pi_1)] \quad \ldots(12)$$

where $M = T^{-1}Z'Z$. We now let p be the vector formed from the components of p_1 and P_2 and under the assumption of normally distributed disturbances and non-random exogenous variables it follows that p has a multivariate normal distribution. The error in some linear combination of the estimator $h'\hat{\beta}$ (where h is a known constant vector) is then, from (12), a simple rational function[27] in the components of p. This rational function satisfies the smoothness and invertibility conditions required for the application of Sargan's theorem and, therefore, the distribution of $\sqrt{T}h'(\hat{\beta}-\beta)$ admits a valid Edgeworth expansion.

While this theorem has great generality it does not extend to situations where there are lagged endogenous variables as regressors. For, in this case, we cannot always assert the normality of a subset of the sample moments[28] (such as p_1 and P_2 in the above example) even if the disturbances are normally distributed; and a more general theory is needed to establish the validity of the Edgeworth expansion in such cases. In another important paper, Sargan (1976*a*) has specialized the results and algorithms given earlier by Chambers to derive general formulae which apply to this case. In particular, Sargan details the formulae for the Edgeworth approximation to the distribution function based on terms up to $O(T^{-1})$: that is, the first three terms of the expansion. An independent proof of the validity of the Edgeworth expansion in this more general setting, together with a discussion of the conditions under which the expansion applies, is contained in Phillips (1977*b*). In these more general situations, it is convenient to write the error in an estimator such as $\hat{\beta}$ as a function of a single vector of underlying variates, q. We then have $\hat{\beta}-\beta = e_T(q)$ and our conditions for the validity of the Edgeworth approximation are of two types.[29] First, some general smoothness and invertibility conditions on the error function $e_T(\ )$ which broadly parallel those in the earlier framework of Sargan (1975). Second, a set of conditions on the stochastic properties of the vector of underlying variates q. These essentially require that $\sqrt{T}q$ has cumulants of the same order of magnitude in $1/\sqrt{T}$ as would a standardized mean (compare (5), above) and that the distribution of $\sqrt{T}q$ itself admits a valid Edgeworth expansion. It should be pointed out at this point that, although this theorem has great generality, the complete verification of the applicability of the result to regression models with lagged endogenous variables requires that the conditions of the theorem be rigorously checked out in this context. As yet, this has not been done.[30]

In addition to the above results, a number of general theorems on the validity of Edgeworth series expansions have recently been published in the probability literature. Unfortunately, none of these are of sufficient generality to apply in time series models. But since they have an important bearing on our subject matter in this section and are related in other ways, which I will mention, to methods that have been used in econometrics I will briefly discuss them here.

Dealing with the case of standardized sums of independent, identically distributed random variables, Ibragimov (1967), gave necessary as well as sufficient conditions for the validity of the Edgeworth expansion. Chibisov (1972) has proved the validity of an asymptotic expansion of the distribution of a multivariate statistic that can itself be represented in the form of an asymptotic series whose terms are polynomial functions of standardized means of independent identically distributed random vectors i.e. a statistic of the form

$$Z_T = h_0(S_T) + \textstyle\sum_{j=1}^{k} T^{-j/2}h_j(S_T) \quad \ldots(13)$$

with $S_T = T^{-\frac{1}{2}}\sum_{t=1}^{T} X_t$ and the X_t independent, identically distributed random p-vectors with zero mean vector and finite moments up to the rth order ($r>2$) and where the $h_j(\cdot)$

are vectors of polynomials of dimension $s \leqq p$ with the components of $h_0(\cdot)$ being linear.[31] One interesting feature of Chibisov's proof of the asymptotic expansion of the distribution of Z_T is the treatment of the change of dimension from the basic statistic S_T to the statistic of interest Z_T. The problems resulting from this change of dimension had not been dealt with in the earlier literature (such as the article by Chambers (1967)), although the proofs in Sargan (1975) and Phillips (1977*b*) provide an alternative way of overcoming this problem.[32]

The representation of a statistic in the form (13) often does not hold precisely but with a remainder whose stochastic order is $T^{-(k+1)/2}$ as $T \to \infty$. Chibisov (1973) gives some conditions under which this will be so and in the 1972 paper establishes asymptotic expansions for the distribution of Z_T in this more general case.[33] Expansions such as (13) and their counterparts with remainders are important in econometrics because moments of a finite number of terms of the expansion lead us to the Nagar approximations to the moments of the estimators. The validity of these moment approximations has been considered by Sargan (1974).

As yet there have been few published applications of Edgeworth approximations in econometric work that enable us to evaluate the accuracy of the approximations in different regions of the distribution.[34] This is particularly true of time series models. In an earlier study (Phillips (1977*a*)), I concentrated on the first order autoregression and found that, when the model was very stable, the approximations based on the first two and three terms of the series were very close to the exact distribution (computed by numerical integration) even for quite small sample sizes (such as $T = 10$). The simple model used in this article did not involve a constant term or exogenous variables. The fact that Orcutt and Winokur (1969) in their sampling experiment discovered quite different sampling behaviour between estimates of the constant and the coefficient in an autoregression suggest that a parallel study on the same model with a fitted mean or the presence of an exogenous variable would be of interest. The present paper goes somewhat further in Sections 5–6 by considering the added complication of simultaneity.

4. LARGE DEVIATION EXPANSIONS

As discussed in the last section, series expansions of the Edgeworth type can be viewed as extensions of the limit theorems which give us the asymptotic distribution of our estimators and test statistics. They, therefore, belong to the same branch in the theory of probability as the classical central limit theorem (see the diagram in Appendix B). Moreover, they share a common limitation with classical central limit theory: namely, that they are often not very informative about the tails or limiting tails of a statistic of interest. To clarify this remark it is helpful to refer back to the case of a standardized sum Z_T of T independent and identically distributed random variables $\{X_t: t = 1, \ldots, T\}$ with a common distribution such that $E(X_t) = 0$ and $E(X_t^2) = \sigma^2$. Then, classical theory tells us that

$$F_T(x) = P(Z_T \leqq x) \to I(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt \qquad \ldots(14)$$

as $T \to \infty$, which is of interest when $x = O(1)$ as $T \to \infty$. But, when the argument x is allowed to vary with T, the statement of the above theorem can be trivial. For instance, if $x \to -\infty$ as $T \to \infty$ then both sides of (14) tend to zero: the theory is uninformative about the relative rate of convergence and this is why the relative error of the asymptotic normal approximation can be large in the tails even for quite large T. In such cases, what we are often really interested in is the behaviour of the limiting tails of $F_T(-x)$ and $1 - F_T(x)$ so that it is more useful here to consider the ratios of tail probabilities

$$\frac{F_T(-x)}{I(-x)} \quad \text{and} \quad \frac{1 - F_T(x)}{1 - I(x)}$$

under the assumption that $x \to \infty$ as $T \to \infty$. If the limiting tails are normal then these ratios will converge to unity as $T \to \infty$. Clearly, the rate at which $x \to \infty$ with T determines how deep in the tails we are concentrating. When $x = o(\sqrt{T})$ as $T \to \infty$ a number of important results have been obtained and these constitute the theory of large deviations.[35,36]

For instance the theorem for tail probabilities of standardized sums corresponding to the classical result (14) tells us that if $x \geqq 0$ and $x = o(\sqrt{T})$ as $T \to \infty$ then (Ibragimov and Linnik (1971) and Petrov (1968))

$$\frac{P(Z_T > x)}{1 - I(x)} = \exp\left\{\frac{x^3}{\sqrt{T}}\Psi\left(\frac{x}{\sqrt{T}}\right)\right\}\left(1 + O\left(\frac{x+1}{\sqrt{T}}\right)\right) \qquad \ldots(15)$$

and

$$\frac{P(Z_T \leqq -x)}{I(-x)} = \exp\left\{-\frac{x^3}{\sqrt{T}}\Psi\left(\frac{-x}{\sqrt{T}}\right)\right\}\left(1 + O\left(\frac{x+1}{\sqrt{T}}\right)\right) \qquad \ldots(16)$$

where

$$\Psi(z) = \psi_0 + \psi_1 z + \psi_2 z^2 + \ldots \qquad \ldots(17)$$

is a power series whose coefficients ψ_i $(i = 0, 1, \ldots)$ depend on the cumulants of X_t and which converges in a neighbourhood of $z = 0$.[37]

Although this theorem is clearly stronger than (14) it also depends on the stronger condition that

$$E\{\exp(a|X_t|)\} < \infty \qquad \ldots(18)$$

for some $a > 0$ so that the moment generating function of the component variates exists and the corresponding characteristic function is analytic in a strip of the imaginary axis (Lukacs (1970)). The main import of (15) and (16) is that the limiting tails of Z_T are normal only if x does not tend to infinity too fast (to be precise $x = o(T^{\frac{1}{6}})$). For if x tends to infinity as fast or faster than a constant multiple of $T^{\frac{1}{6}}$ then the limiting tails of Z_T are not normal but will depend on the coefficients in the power series $\Psi(z)$. Thus, if $x \geqq 0$ and $x = O(T^{\frac{1}{4}})$ as $T \to \infty$ it is easy to see that

$$P(Z_T > x) = (1 - I(x))\exp\left\{\psi_0\frac{x^3}{\sqrt{T}} + \psi_1\frac{x^4}{\sqrt{T}}\right\}\left(1 + O\left(\frac{x+1}{\sqrt{T}}\right)\right) \qquad \ldots(19)$$

and, more generally, if $x \geqq 0$ and $x = O(T^{k/2(k+2)})$ for some positive integer k then

$$P(Z_T > x) = (1 - I(x))\left\{\exp\frac{x^3}{\sqrt{T}}\Psi^{[k]}\left(\frac{x}{\sqrt{T}}\right)\right\}\left(1 + O\left(\frac{x+1}{\sqrt{T}}\right)\right) \qquad \ldots(20)$$

where $\Psi^{[k]}(z)$ represents the first k terms of the series $\Psi(z)$ in (17) above. Similar results hold for the negative tail.

In practice, we are frequently concerned with approximating the tails of the distribution of a test statistic whose exact distribution is unknown. In such cases, where x may be quite large relative to $\sqrt{T}$ it is known that the Edgeworth approximation can lead to unsatisfactory results, including negative probabilities.[38] An alternative which should be available in many cases is to use the first few terms in a large deviation expansion such as (19) or (20). Note that these expansions have the advantage that they are positive for all x (although not necessarily *less* than unity) and might be expected to do well at least for a certain region in the tails. Some time ago, Chernoff (1956) pointed out the relevance of this type of limit theory in statistical applications but, to my knowledge, there have as yet been few applications.

One limitation to the immediate application in econometrics of large deviation limit theory and its associated expansions is the fact that virtually all the results established so far in the probability literature apply only in the case of standardized sums of independent random variables (or vectors). In two recent articles, I have explored the validity and form of large deviation expansions such as (19) and (20) in a more general setting. The first article (Phillips (1977*c*)) establishes a large deviation limit theorem for multivariate statistics which are more general than standardized means but which depend on the sample size T in much the same way as $T\to\infty$.[39] The second article (Phillips (1976)) derives general formulae for large deviation expansions such as (19) and (20) when we are interested in approximating the tails of the sampling distribution of statistics which can be represented as quite general functions of the first and second sample moments of the data. Thus, the setting is the same as that in Sargan (1976*a*) and Phillips (1977*b*) dealing with the Edgeworth expansion, so that the formulae derived should apply in a number of different models including those with lagged endogenous variables and, in some cases, non-normal errors. The expansion obtained is then applied to tail probabilities of the least squares estimator of the coefficient in a first order autoregression. The numerical results suggest that this new approximation can give good results in the region of the tail between 10 per cent and 1 per cent in this example but the performance of the approximation is more sensitive to the stability of the model than the Edgeworth approximation.[40] Both these approximations, therefore, need to be used with care when approximating tail probabilities in dynamic models.

5. ALTERNATIVE ESTIMATORS OF THE MARGINAL PROPENSITY TO CONSUME

The model we use in our application of the theory in Section 3 is the system

$$C_t=\alpha Y_t+\beta C_{t-1}+u_t \qquad (t=\ldots,-1,0,1,2,\ldots) \qquad \ldots(21)$$

$$Y_t=C_t+I_t \qquad (t=\ldots,-1,0,1,2,\ldots) \qquad \ldots(22)$$

where the variables C_t, Y_t and I_t represent consumption, income and investment, respectively. We assume that the disturbances u_t $(t=\ldots,-1,0,1,2,\ldots)$ are serially independent and identically distributed as $N(0,\sigma_u^2)$; and I_t is taken to be a non-random exogenous variable whose sample second moment converges to a finite positive constant as the sample size tends to infinity. As Bergstrom argued in 1962, the model given by (21) and (22) is, in spite of its simplicity, still the kernel of many macroeconometric models. The lag in the consumption function lends an additional feature of realism and can be justified on the basis of a number of theories. But, not all of these theories are consistent with a serially independent disturbance, so our assumptions about the stochastic properties of u_t narrow down the field of application of the model.[41] Nevertheless, it is hoped that the model provides a useful starting point in the development of a small sample theory for dynamic simultaneous equation models.

Writing $\delta=\beta/(1-\alpha)$ and under the stability condition $|\delta|<1$ we derive from (21) and (22) the final form equations

$$C_t=\left(\frac{\alpha}{1-\alpha}\right)\sum_{s=0}^{\infty}\delta^s I_{t-s}+\left(\frac{1}{1-\alpha}\right)\sum_{s=0}^{\infty}\delta^s u_{t-s}=m_t+w_t, \qquad \ldots(23)$$

say, and

$$Y_t=I_t+m_t+w_t.$$

We define the $T\times 1$ observation vectors $c'=(C_1, C_2, \ldots, C_T)$, $c'_{-1}=(C_0, C_1, \ldots, C_{T-1})$, $y'=(Y_1, Y_2, \ldots, Y_T)$ and $d'=(I_1, I_2, \ldots, I_T)$ so that the OLS estimators α^* and β^* of α and β in (21) are given by

$$\alpha^*=\frac{(c'_{-1}c_{-1})(y'c)-(y'c_{-1})(c'_{-1}c)}{(y'y)c'_{-1}c_{-1})-(y'c_{-1})^2} \quad\text{and}\quad \beta^*=\frac{(y'y)(c'c_{-1})-(c'_{-1}y)(y'c)}{(y'y)(c'_{-1}c_{-1})-(y'c_{-1})^2}.$$

But, using (22) we can write these formulae as

$$\alpha^* = \frac{(c'_{-1}c_{-1})(c'c+d'c)-(c'c_{-1}+d'c_{-1})(c'_{-1}c)}{(c'c+2d'c+d'd)(c'_{-1}c_{-1})-(c'c_{-1}+d'c_{-1})^2} \qquad \text{...(24)}$$

and

$$\beta^* = \frac{(c'c_{-1})(c'c+2d'c+d'd)-(c'c_{-1}+d'c_{-1})(c'c+d'c)}{(c'c+2d'c+d'd)(c'_{-1}c_{-1})-(c'c_{-1}+d'c_{-1})^2}. \qquad \text{...(25)}$$

Our object now is to write α^* and β^* in terms of standardized sample moments of the underlying data so that the functional representation of the estimators corresponding to that described in Section 2 in connection with the papers by Sargan (1976*a*) and Phillips (1977*b*). We let $x' = (C_0, C_1, \ldots, C_T)$, define the $(T+1)\times(T+1)$ matrices

$$A_1 = \begin{bmatrix} 1 & 0 & \ldots & 0 & 0 \\ 0 & 1 & \ldots & 0 & 0 \\ \cdot & \cdot & \ldots & \cdot & \cdot \\ 0 & 0 & \ldots & 1 & 0 \\ 0 & 0 & \ldots & 0 & 0 \end{bmatrix}, \qquad A_2 = \begin{bmatrix} 0 & \frac{1}{2} & \ldots & 0 & 0 \\ \frac{1}{2} & 0 & \ldots & 0 & 0 \\ \cdot & \cdot & \ldots & \cdot & \cdot \\ 0 & 0 & \ldots & 0 & \frac{1}{2} \\ 0 & 0 & \ldots & \frac{1}{2} & 0 \end{bmatrix}$$

and

$$A_3 = \begin{bmatrix} 0 & 0 & \ldots & 0 & 0 \\ 0 & 1 & \ldots & 0 & 0 \\ \cdot & \cdot & \ldots & \cdot & \cdot \\ 0 & 0 & \ldots & 1 & 0 \\ 0 & 0 & \ldots & 0 & 1 \end{bmatrix}$$

and introduce the variables

$$q_i = \{x'A_ix - E(x'A_ix)\}/T \qquad (i = 1, 2, 3)$$
$$q_i = \{b'_ix - E(b'_ix)\}/T \qquad (i = 4, 5)$$

where $b'_4 = (d', 0)$ and $b'_5 = (0, d')$. We note that $c'_{-1}c_1 = x'A_1x$, $c'c_{-1} = x'A_2x$, $c'c = x'A_3x$, $d'c_{-1} = b'_4x$ and $d'c = b'_5x$.

By setting $\mu = T^{-1}d'd$, $\mu_i = T^{-1}E(x'A_ix)$ $(i = 1, 2, 3)$ and $\mu_p = T^{-1}E(b'_ix)$ $(p = 4, 5)$ we find the representations

$$\alpha^* = \{(q_1+\mu_1)(q_3+q_5+\mu_3+\mu_5)-(q_2+q_4+\mu_2+\mu_4)(q_2+\mu_2)\}$$
$$\div\{(q_3+2q_5+\mu+\mu_3+2\mu_5)(q_1+\mu_1)-(q_2+q_4+\mu_2+\mu_4)^2\}, \qquad \text{...(26)}$$

$$\beta^* = \{(q_2+\mu_2)(q_3+2q_5+\mu+\mu_3+2\mu_5)-(q_2+q_4+\mu_2+\mu_4)(q_3+q_5+\mu_3+\mu_5)\}$$
$$\div\{(q_3+2q_5+\mu+\mu_3+2\mu_5)(q_1+\mu_1)-(q_2+q_4+\mu_2+\mu_4)^2\}. \qquad \text{...(27)}$$

In a similar way we find the following representations of the 2SLS estimators $\hat{\alpha}$ and $\hat{\beta}$ of α and β

$$\hat{\alpha} = \frac{(q_1+\mu_1)(q_5+\mu_5)-(q_2+\mu_2)(q_4+\mu_4)}{(q_1+\mu_1)(q_5+\mu_5+\mu)-(q_4+\mu_4)(q_2+q_4+\mu_2+\mu_4)} \qquad \text{...(28)}$$

and

$$\hat{\beta} = \frac{(q_2+\mu_2)(q_5+\mu_5+\mu)-(q_5+\mu_5)(q_2+q_4+\mu_2+\mu_4)}{(q_1+\mu_1)(q_5+\mu_5+\mu)-(q_4+\mu_4)(q_2+q_4+\mu_2+\mu_4)}. \qquad \text{...(29)}$$

The representations (26) to (29) express each estimator as a rational function of the elements q_i, $(i = 1, ..., 5)$, and are suitable for the application of the algorithm in Sargan (1976*a*) to derive the Edgeworth approximation. The algorithm requires derivatives of these functions up to the third order and cumulants of the elements q_i up to the fourth order. To compute the derivatives, a computer programme was written[42] to obtain analytical derivatives of polynomials and rational functions up to the third order. The programme reads in only the form of the polynomials and the value of the coefficients; it evaluates derivatives at any required point although, for the purposes of the present application, derivatives are evaluated at the origin.[43]

6. THE CHARACTERISTIC FUNCTION AND CUMULANTS OF THE SAMPLE MOMENTS OF THE DATA

Under our assumption of a normally distributed disturbance it follows that the vector x is normal with mean vector m and covariance matrix Ω where

$$m = \begin{bmatrix} m_0 \\ m_1 \\ \vdots \\ m_T \end{bmatrix}, \qquad \Omega = \tau^2 \begin{bmatrix} 1 & \delta & \dots & \delta^T \\ \delta & 1 & \dots & \delta \\ \cdot & \cdot & & \cdot \\ \delta^T & \delta^{T-1} & \dots & 1 \end{bmatrix}$$

$\tau^2 = \sigma_u^2\{(1-\alpha)^2 - \beta^2\}^{-1}$ and $\delta = \beta/(1-\alpha)$. The characteristic function of the vector $Q' = (Q_1, ..., Q_5)$ where $Q_i = x'A_ix$ $(i = 1, 2, 3)$ and $Q_p = b_p'x$ $(p = 4, 5)$ is given by[44]

$$\bar{\theta}(t) = |I - 2iG(t)\Omega|^{-\frac{1}{2}} \exp\{\tfrac{1}{2}(m + i\Omega f(t))'(\Omega - 2i\Omega G(t)\Omega)^{-1}(m + i\Omega f(t)) - \tfrac{1}{2}m'\Omega^{-1}m\} \quad \dots(30)$$

where

$$G(t) = t_1A_1 + t_2A_2 + t_3A_3 \quad \text{and} \quad f(t) = t_4b_4 + t_5b_5.$$

Writing the second characteristic (or cumulant generating function) of Q as $\bar{\lambda}(t) = \log(\bar{\theta}(t))$ we note from the relationship $q_i = T^{-1}Q_i - \mu_i$ that the second characteristic of the vector $\sqrt{T}q$, where $q' = (q_1, ..., q_5)$ is

$$\begin{aligned} \lambda(t) &= \bar{\lambda}(t/\sqrt{T}) - iT^{\frac{1}{2}}\textstyle\sum_{j=1}^{5}\mu_jt_j \\ &= -\tfrac{1}{2}\log\det\left(I - \frac{2i}{\sqrt{T}}G(t)\Omega\right) + \tfrac{1}{2}\left(m + \frac{i}{\sqrt{T}}\Omega f(t)\right)'\left(\Omega - \frac{2i}{\sqrt{T}}\Omega G(t)\Omega\right)^{-1}\left(m + \frac{i}{\sqrt{T}}\Omega f(t)\right) \\ &\quad -\tfrac{1}{2}m'\Omega^{-1}m - iT^{\frac{1}{2}}\textstyle\sum_j \mu_jt_j. \end{aligned} \quad \dots(31)$$

By differentiating (31) with respect to the elements of t and evaluating these derivatives at the origin we can extract the cumulants of $\sqrt{T}q$. The resulting expressions are detailed in Appendix C. These expressions are closely related to the formulae for the expectation of products of quadratic forms in normal variables. Neudecker (1968) obtained the appropriate formulae in the case of the product of three quadratic forms and Kumar (1973) gives a procedure for extracting the required expression in the general case but does not give explicit formulae. In fact, the recent results of Carlson (1972) (see also Exton (1976)) enable us to write down such expectations explicitly in terms of the Lauricella multiple hypergeometric function (Exton (1976)). An alternative solution to the same problem based on the formulae for the cumulants has been given by Magnus (1978).

The cumulant expressions given in Appendix C were programmed for calculation so that the results could be directly combined with the derivative calculations discussed in Section 5 to yield the coefficients in the Edgeworth approximation up to $O(T^{-1})$. The cumulants depend not only on the values of the underlying parameters α, β and σ_u^2 but also on the vectors b_4 and b_5 which contain the sample period values of the exogenous variable I_t, $(t=1, \ldots, T)$ as well as the mean vector m whose components m_t, $(t=0, 1, \ldots, T)$ depend on the whole past history of the exogenous variable, as is clear from (23) above. Thus, to calculate the Edgeworth approximations to the distribution of the OLS and 2SLS estimators of α and β in (21) we need first to specify a series for I_t. Clearly, a wide choice is available, including real data series. To keep this study within manageable bounds and at the same time enable us to measure the effect of different types of exogenous series on the approximations the following two processes were selected for generation of the I_t series:

$$I_t=\rho I_{t-1}+v_t; \qquad v_t \text{ i.i.d. } N(\mathrm{O}, \sigma_v^2) \qquad \ldots(32)$$

and

$$I_t=\rho_1 I_{t-1}+\rho_2 I_{t-2}+v_t; \qquad v_t \text{ i.i.d. } N(\mathrm{O}, \sigma_v^2). \qquad \ldots(33)$$

Data on I_t $(t=-50, \ldots, -1, 0, 1, \ldots, T)$ were generated once and for all from (32) and (33) with parameter values (i) $\rho=0{\cdot}2$, $0{\cdot}5$, $0{\cdot}8$ and $\sigma_v^2=1{\cdot}0$ for (32); and (ii) $\rho_1=0{\cdot}75$, $\rho_2=-0{\cdot}5$ and $\sigma_v^2=1{\cdot}0$ for (33). These values give us in (i) various degrees of correlation in the generated series and in (ii) a series with a cycle of approximately 6 years if the time unit is a year. The remaining parameter σ_v^2 has been fixed at unity and, when we compute the approximations, we consider various values of σ_u^2, the variance of the disturbance on the consumption function (21); this enables us to measure the effect of changes in the signal/noise ratio on the form of the approximations.

Once we have specified the process generating the series I_t we can readily compute the limits in probability of the OLS estimators of α and β. We illustrate some of these computations in Appendix D for the case of (32).

7. NUMERICAL EVALUATION OF THE APPROXIMATE DISTRIBUTIONS

We concentrate on the small sample distributions of the OLS and 2SLS estimators of α and β in (21). Numerical computations of the Edgeworth approximations to these distributions are possible once we have specified values of the underlying parameters α, β and σ_μ^2 in the model (21)–(22) as well as a series for the exogenous variable I_t over the relevant sample period (corresponding to an assumed value for T) and enough of the past history of I_t to accurately compute the components m_t as in (23). Taking into account the additional parameters that occur in the processes (32) and (33) that we have selected for the generation of the I_t series, we have a sizeable parameter space from which to sample values. A selection of numerical results are given in the graph of Appendix E. These graphs illustrate the effect on the distributions of some of the more important parameter changes that have been found but are not in any way exhaustive.

The values of α and β do not seem greatly to influence the shape of the distributions provided we keep within the region of stability. Most of the graphs (Figures 1–5), therefore, refer to the values

$$\alpha=0{\cdot}2 \qquad \beta=0{\cdot}7$$

which give us a stability coefficient of $\delta=0{\cdot}87$ and a long run marginal propensity to consume of $0{\cdot}73$. We also fix the variance of the disturbance on (32) and (33) at a value of $\sigma_v^2=1{\cdot}0$. With these values of the parameters fixed, we then consider the effect of the different exogenous series

$$I_t=\rho I_{t-1}+u_t; \qquad \rho=0{\cdot}2, 0{\cdot}8$$
$$I_t=0{\cdot}75 I_{t-1}-0{\cdot}5 I_{t-2}+u_t$$

and variations

$$\sigma_u^2 = 1{\cdot}0, \qquad \sigma_u^2 = 0{\cdot}5$$

in the variance of the disturbance on the consumption function.

Rather than give a detailed account of individual figures it may be useful to summarize some of the features that seem to emerge from these graphs. As in other cases, some subjective judgement about the importance of errors of different magnitudes is required in making an assessment. The following comments are, therefore, quite brief and readers may wish to form their own views of the main implications of the results.

(i) The central location (measured, for instance by the approximate median) of the 2SLS estimator is closer to the true value of the parameter than OLS. In many cases, the apparent bias of OLS is substantial, representing as much as 100 per cent of the true parameter value (e.g. Figure 1(a)).

(ii) The bias of the OLS estimator seems to increase noticeably as σ_u^2/σ_v^2 increases in value and also as $\sigma_u^2/\text{var}\,(I_t)$ increases.

(iii) The OLS estimator of α is biased upwards (as in the case of the non-dynamic consumption function treated by Bergstrom (1962)); while the OLS estimator of β is biased downwards (as in the case of an autoregressive equation (Phillips (1977*a*)).

(iv) The magnitude and direction of the bias apparent in the small sample distribution of the OLS estimator is compatible with the asymptotic bias given by the calculations in Appendix D.

(v) The relative locations of the distributions of the OLS and 2SLS estimators seem compatible with the numerical differences between these estimates that have been observed in empirical work. For example, Klein (1969) gives the following set of results for OLS and 2SLS estimates of the consumption function for non-durable goods in the Klein–Goldberger model (with $T = 31$):

$$C_{nt} = \begin{bmatrix} 0{\cdot}332 \\ (6{\cdot}5) \\ 0{\cdot}250 \\ (4{\cdot}0) \end{bmatrix} Y_t + \begin{bmatrix} 0{\cdot}616 \\ (9{\cdot}2) \\ 0{\cdot}723 \\ (8{\cdot}9) \end{bmatrix} C_{nt-1} - \begin{bmatrix} 0{\cdot}378 \\ (0{\cdot}3) \\ -1{\cdot}17 \\ (0{\cdot}8) \end{bmatrix} \begin{matrix} \text{OLS} \\ \\ \text{2SLS}^{\cdot} \\ \\ \end{matrix}$$

(vi) The sampling dispersion of the 2SLS estimators seems to be larger than that of OLS, although for certain of the exogenous series the differences are not great.

(vii) In most cases, the smaller sampling dispersion of the OLS estimator does not offset the bias of the estimator in terms of concentration about the true value of the parameter. The graphs seem to the author to support the conclusion that 2SLS is still the better general purpose estimator in this model. However, the probability of outliers (including negative short run marginal propensities) appears to be substantial, particularly for the small sample size $T = 10$.

(viii) An important factor in determining the shape of the distribution appears to be the process generating the exogenous series I_t. For data from the second order process (33), the distributions display a greater degree of concentration and the bias of the OLS estimator is reduced. This is true of estimates of both α and β and it seems particularly relevant in the case $T = 10$.

(ix) The distributions show differing rates of convergence as T increases and this feature seems to be related to the exogenous series being used. In general, the convergence to a degenerate distribution seems more uniform when the data are generated by the second order process.

(x) For the small sample size $T = 10$, there is some evidence that the approximations will be unreliable in the tails. In a number of cases for the 2SLS estimator, the approximation overshoots unity and is non-monotonic in the domain considered (c.f. Phillips (1977*a*)).

8. FINAL COMMENTS

The analysis in Sections 5 and 6 can be applied directly to t-ratio statistics in the same model. The cumulants as given in Appendix C remain unchanged but the form of the error function, denoted $e_T(q)$ in Section 3 of the paper, must be adapted to correspond to the t-statistic. The function is now a ratio with a surd in the denominator, but this can be handled with no difficulty in the analytic differentiation routine during the computation of the approximation.[45]

One limitation of the results in Sections 5–7 is the assumption of a normally distributed disturbance on the consumption function. In this, as well as other situations in econometrics, we may often wish to explore the effect of departures from normality on the shape of the small sample distributions of our estimators. One possible procedure for handling such departures is to work with an error distribution that includes the normal as a special case (such as the exponential power distributions or a Gram–Charlier distribution). The advantage of working with a Gram–Charlier error distribution is that the cumulant formulae would be readily calculable along the same lines as the present paper but with extra parameters included that measure the extent of the departure from normality. Some preliminary work dealing with quadratic forms has been done along these lines by Subrahmaniam ((1966) and (1968)). As yet there have been no applications of this idea in econometrics.

Another potential application of the Edgeworth approximation in the present context is to analyse the effects of specification error. It would be of interest, in particular, to see how the recent results of this type of analysis in classical situations discussed in Section 2 are affected by the presence of lagged endogenous variables in the regressor set.

APPENDIX A: SOME EXACT THEORY IN THE CLASSICAL CASE

Our discussion will concentrate on the 2SLS estimator. Other instrumental variable estimators and k-class estimators (non-stochastic k) can be treated in a similar way although in the case of k-class estimators where $0 < k < 1$ the analysis is more complicated.

A(i) *An Equation with Two Endogenous Variables*

We consider a single structural equation such as (10) in which there is one right hand side endogenous variable (i.e. $n = 1$). Then we have

$$y_1 = \beta y_2 + Z_1\gamma + u. \qquad \ldots(\text{A.1})$$

The reduced form equations for y_1 and y_2 are

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

where Z_2 is a $T \times K_2$ matrix of exogenous variables excluded from (A.1). We assume that the usual standardizing transformations (Basmann (1963*a*) and (1974)) 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 assume that the equation (A.1) is identified so that $K_2 \geqq 1$.

The 2SLS estimator of β in (A.1) is given by the ratio

$$\hat{\beta} = y_2'Ry_1/y_2'Ry_2 \qquad \ldots(\text{A.2})$$

where

$$R = Z_2(Z_2'Z_2)^{-1}Z_2' = Z_2Z_2'.$$

The joint Laplace transform of $y_2'Ry_2$ and $y_2'Ry_1$ is[46]

$$L(w_1, w_2) = (1-2w_1-w_2^2)^{-K_2/2} \cdot \exp\left[\frac{\mu^2}{2}\left\{\frac{(1+\beta w_2)^2}{1-2w_1-w_2^2}-1\right\}\right] \qquad \ldots(\text{A}.3)$$

where $\mu^2 = T\pi_{22}'\pi_{22}$ and where w_1 and w_2 satisfy $1-2Re(w_1)-(Re(w_2))^2>0$. The density function of $\hat{\beta}$ is now given by the inversion formula (Cramér (1946))

$$\text{pdf}(r) = \frac{1}{2\pi i}\int_{-i\infty}^{i\infty} \left.\frac{\partial L(u-rw, w)}{\partial u}\right|_{u=0} dw.$$

Following the analysis in Holly and Phillips (1977) this reduces to the integral

$$\frac{1}{2\pi i}\int_{-i\infty}^{i\infty} B(w)\exp\left\{\frac{\mu^2}{2}\psi(w)\right\}dw \qquad \ldots(\text{A}.4)$$

where

$$B(w) = \left[K_2 + \frac{\mu^2(1+\beta w)^2}{1+2rw-w^2}\right](1+2rw-w^2)^{-(K_2+2)/2}$$

and

$$\psi(w) = \frac{w^2(1+\beta^2)+2w(\beta-r)}{1+2rw-w^2}.$$

We note that the integrand in (A.4) is analytic except for those values of w where $1+2rw-w^2=0$ i.e. except for the following two points on the real axis

$$a_1 = r-(1+r^2)^{\frac{1}{2}}, \qquad a_2 = r+(1+r^2)^{\frac{1}{2}}.$$

When K_2 is even, these two points are essential singularities; when K_2 is odd, the points are also branch points because of the fractional power that occurs in $B(w)$. We, therefore, deal with these two cases separately.

(i) *K_2 even*

We consider the contour illustrated in Figure A(a). The integral along the imaginary axis $[-iR, iR]$ together with the integral around the semi-circle C is just $2\pi i$ times the residue of the integrand at the essential singularity $r-(1+r^2)^{\frac{1}{2}}$. Moreover, if we let $R\to\infty$ it is easy to see, from the behaviour of the integrand in (A.4) and the inequality (see, for instance, Miller (1960, p. 74))

$$\left|\int_C f(w)dw\right| \leqq \pi R \max_{w\in C}|f(w)|$$

where $f(w) = B(w)\exp\{\frac{1}{2}\mu^2\psi(w)\}$, that the integral around the semi-circle C tends to zero as $R\to\infty$. Thus, we are left with the simple relationship

$$\text{pdf}(r) = \text{residue}_{w=a_1}\left[B(w)\exp\left\{\frac{\mu^2}{2}\psi(w)\right\}\right]. \qquad \ldots(\text{A}.5)$$

(A.5) is a simple relationship[47] and to extract the analytic form of the density we need to compute the residue at a_1. To do this we need only identify the coefficient of the term $1/(w-a_1)$ in the Laurent series expansion of the integrand $f(w)$ in an annulus around a_1.

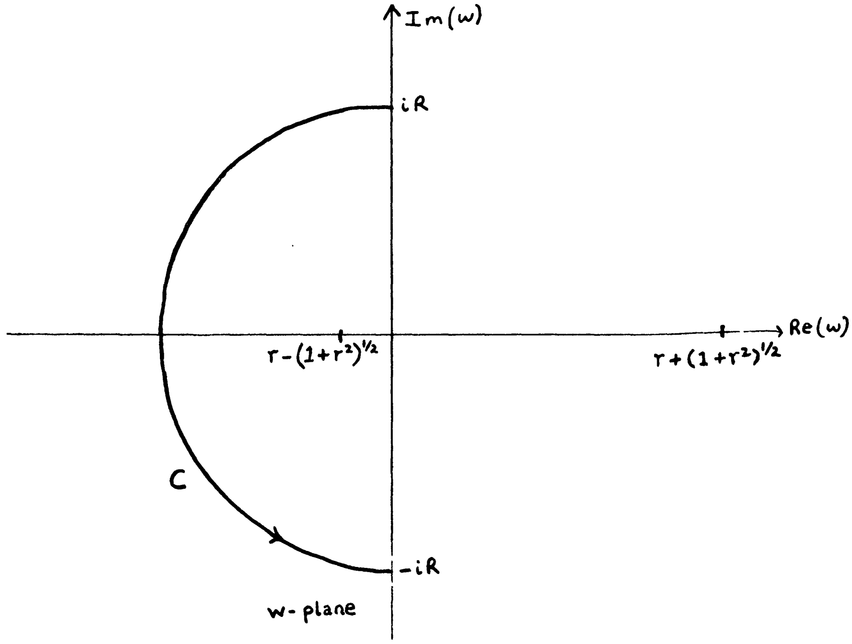


FIGURE A(a)

We can write $1+2rw-w^2=-(w-a_1)(w-a_2)$ and expand $\psi(w)$ and the first factor of $B(w)$ in partial fractions. This gives us the following representation:

$$f(w)=\left[(K_2-\mu^2\beta^2)+\frac{\alpha A_1(r)}{w-a_1}-\frac{\alpha A_2(r)}{w-a_2}\right]\left(\frac{(-1)^l}{(w-a_1)^l(w-a_2)^l}\right)\exp\left\{-\frac{\mu^2}{2}(1+\beta^2)\right\}$$
$$\cdot\exp\left\{-\frac{\mu^2}{2\alpha}\left(\frac{1}{w-a_2}\right)B_2^2(r)\right\}\exp\left\{\frac{\mu^2}{2\alpha}\left(\frac{1}{w-a_1}\right)B_1^2(r)\right\} \quad \ldots(A.6)$$

where

$$l=(K_2+2)/2,\ \alpha=(a_2-a_1)^{-1}=\tfrac{1}{2}(1+r^2)^{-\frac{1}{2}},$$
$$A_1(r)=\mu^2(1+\beta^2)-\frac{2\beta\mu^2(1+\beta r)}{\alpha}+2\beta\mu^2(1+\beta r)a_2,$$
$$A_2(r)=\mu^2(1+\beta^2)+2\beta\mu^2(1+\beta r)a_2,$$
$$B_1(r)=\frac{1+\beta r}{2(1+r^2)^{\frac{1}{2}}}-\frac{\beta}{2} \quad \ldots(A.7)$$

and

$$B_2(r)=\frac{1+\beta r}{2(1+r^2)^{\frac{1}{2}}}+\frac{\beta}{2}. \quad \ldots(A.8)$$

The factors involving $(w-a_2)^{-1}$ taken to some power can now be expanded using the binomial expansion as

$$(w-a_2)^{-m}=(-\alpha)^m\{1-\alpha(w-a_1)\}^{-m}$$
$$=(-\alpha)^m\sum_{k=0}^{\infty}\frac{\Gamma(m+k)}{\Gamma(m)k!}\alpha^k(w-a_1)^k$$

for w in a neighbourhood of a_1. We also expand the final exponential factor of (A.5) in a power series. This gives us the required Laurent series for $f(w)$. We now find the coefficient of $(w-a_1)^{-1}$, and hence the density, to be

$$\exp\left\{-\frac{\mu^2}{2}(1+\beta^2)\right\}\left[(K_2-\mu^2\beta^2)\alpha^{2l-1}\sum_{m=0}^{\infty}\frac{(\mu^2/2)^m B_2^{2m}(r)}{m!}\right.$$

$$\cdot\sum\sum\sum_{p+k=n+l-1}\frac{(l)_p(m)_k}{p!k!n!}\left(\frac{\mu^2}{2}B_1^2(r)\right)^n$$

$$+\alpha^{2l+1}A_1(r)\sum_{m=0}^{\infty}\frac{(\mu^2/2)^m B_2^{2m}(r)}{m!}\sum\sum\sum_{p+k=n+l}\frac{(l)_p(m)_k}{p!k!n!}\left(\frac{\mu^2}{2}B_1^2(r)\right)^n$$

$$\left.+\alpha^{2l+1}A_2(r)\sum_{m=0}^{\infty}\frac{(\mu^2/2)^m B_2^{2m}(r)}{m!}\sum\sum\sum_{p+k=n+l-2}\frac{(l)_p(m)_k}{p!k!n!}\left(\frac{\mu^2}{2}B_1^2(r)\right)^n\right] \quad \ldots(\text{A.9})$$

where $(a)_n$ denotes $\Gamma(a+n)/\Gamma(a)$. No doubt this can be simplified into the form given by Richardson (1968).

(ii) *K_2 odd*

We work from (A.4) above and note that, in this case, $B(w)$ has a branch point at $w=a_1=r-(1+r^2)^{\frac{1}{2}}$. We, therefore, cut the real axis from $-\infty$ to $r-(1+r^2)^{\frac{1}{2}}$ and consider the contour illustrated in Figure A(b) below.

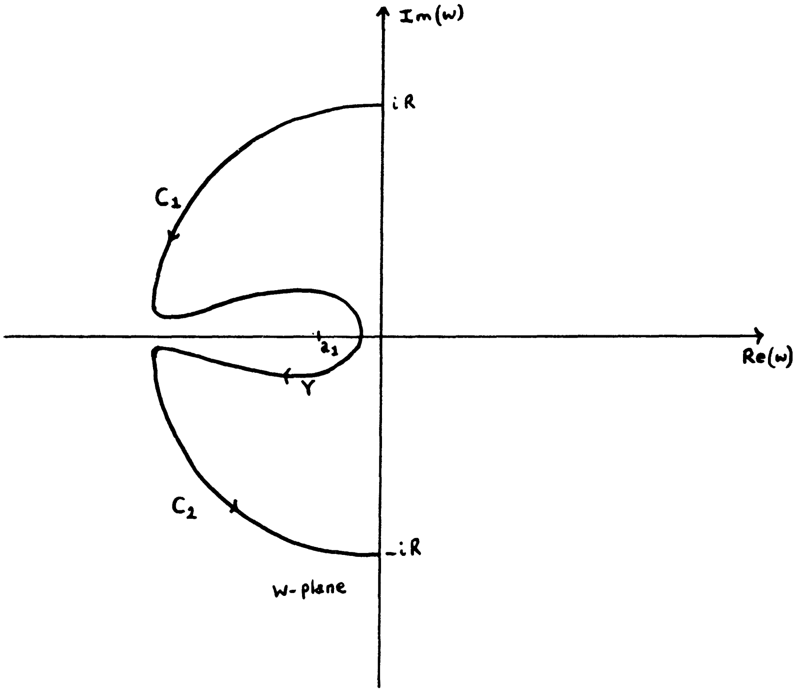


FIGURE A(b)

As in the previous case where K_2 was even, the integrals over C_1 and C_2 tend to zero as the radius $R \to \infty$. Thus

$$\text{pdf}(r) = \frac{1}{2\pi i}\int_{-i\infty}^{i\infty} f(w)dw = \frac{1}{2\pi i}\int_{\gamma^*} f(w)dw \qquad \ldots(\text{A}.10)$$

where

$f(w) = B(w)\exp\{\tfrac{1}{2}\mu^2\psi(w)\}$ as before and γ^* is the contour shown in Figure A(c).

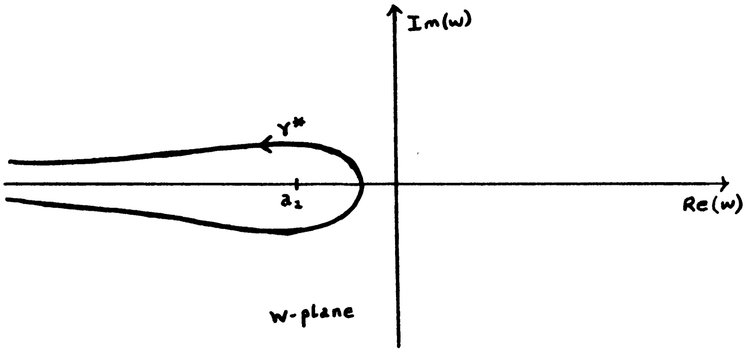


FIGURE A(c)

According to (A.10), the density function is given by a loop integral[48] starting at minus infinity passing around the branch point and receding to minus infinity. Many special functions have integral representations similar to (A.10), of which perhaps the best known is the Hankel integral defining the Gamma function (Erdéyli (1953, p. 13)). We now show how to transform (A.8) into a form which is recognizable in terms of one of the Pochhammer integral representations of the confluent hypergeometric function (Slater (1960, pp. 38–41)).

We start by introducing a new complex variable ζ defined by the equation

$$w - a_1 = c(1-\zeta)/\zeta \qquad \ldots(\text{A}.11)$$

where

$$c = a_1 - a_2 = -2(1+r^2)^{\frac{1}{2}}.$$

This transformation takes $w = a_1$ into $\zeta = 1$ and $w = -\infty$ into $\zeta = 0$ in the extended planes. The contour γ^* in Figure A(c) is mapped into the contour η shown in Figure A(d).

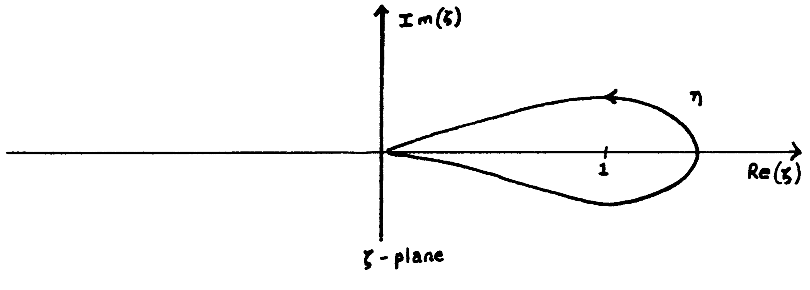


FIGURE A(d)

Under the transformation (A.11) we find that $w-a_2=c/\zeta$, $dw/d\zeta=-c/\zeta^2$ and after a little manipulation

$$\begin{aligned}\psi(w)&=-(1+\beta^2)+\left\{\frac{1+\beta r}{2(1+r^2)^{\frac{1}{2}}}-\frac{\beta}{2}\right\}^2+\zeta\left\{\frac{1+\beta r}{2(1+r)^{\frac{1}{2}}}+\frac{\beta}{2}\right\}^2-\frac{1}{1-\zeta}\left\{\frac{1+\beta r}{2(1+r^2)^{\frac{1}{2}}}-\frac{\beta}{2}\right\}^2\\&=-(1+\beta^2)+B_1^2(r)+\zeta B_2^2(r)-\left(\frac{1}{1-\zeta}\right)B_1^2(r)\end{aligned}$$

in the notation of (A.7) and (A.8). Moreover

$$\left(\frac{1}{1+2rw-w^2}\right)^{(K_2+2)/2}=\frac{(-1)^{(K_2+2)/2}\zeta^{K_2+2}}{C^{K_2+2}(1-\zeta)^{(K_2+2)/2}}$$

and

$$\frac{\mu^2(1+\beta w)^2}{1+2rw-w^2}=-\frac{\mu^2\{\zeta(1+\beta a_2)+\beta c\}^2}{c^2(1-\zeta)}.$$

Now

$$\operatorname{pdf}(r)=\frac{1}{2\pi i}\int_{\gamma^*}f(w)dw=\frac{1}{2\pi i}\int_{\eta}g(\zeta)d\zeta$$

where

$$\begin{aligned}g(\zeta)&=f(w(\zeta))\frac{dw}{d\zeta}\\&=\left[K_2-\frac{\mu^2\{\zeta(1+\beta a_2)+\beta c\}^2}{c^2(1-\zeta)}\right]\frac{(-1)^{(K_2+2)/2}\zeta^{K_2+2}}{c^{K_2+2}(1-\zeta)^{(K_2+2)/2}}\cdot\frac{(-c)}{\zeta^2}\cdot\exp\left\{-\frac{\mu^2}{2}(1+\beta^2+B_1^2(r))\right\}\\&\quad\cdot\exp\left\{\zeta\frac{\mu^2}{2}B_2^2(r)\right\}\exp\left\{-\frac{1}{1-\zeta}\frac{\mu^2}{2}B_1^2(r)\right\}.\end{aligned}$$

Thus

$$\begin{aligned}\operatorname{pdf}(r)&=\frac{(-1)^{K_2/2+2}}{2\pi i c^{K_2+1}}\exp\left\{-\frac{\mu^2}{2}(1+\beta^2+B_1^2(r))\right\}\int_{\eta}\left[K_2-\frac{\mu^2\{\zeta(1+\beta a_2)+\beta c\}^2}{c^2(1-\zeta)}\right]\\&\quad\cdot\exp\left\{\zeta\frac{\mu^2}{2}B_2^2(r)\right\}\sum_{m=0}^{\infty}\frac{(\mu^2/2)^m B_1^{2m}(r)}{m!}\frac{\zeta^{K_2}d\zeta}{(1-\zeta)^{(K_2+2)/2+m}}\end{aligned}$$

$$\begin{aligned}\operatorname{pdf}(r)&=\frac{(-1)^{K_2/2}\exp\left\{-\frac{\mu^2}{2}(1+\beta^2+B_1^2(r))\right\}}{2\pi i c^{K_2+1}}\sum_{m=0}^{\infty}\frac{(\mu^2/2)^m B_1^{2m}(r)}{m!}\\&\quad\cdot\int_{\eta}\left[K_2-\frac{\mu^2\{\zeta^2(1+\beta a_2)^2+2\beta c(1+\beta a_2)\zeta+\beta^2c^2\}}{c^2(1-\zeta)}\right]\\&\quad\cdot\exp\left\{\zeta\frac{\mu^2}{2}B_2^2(r)\right\}\frac{\zeta^{K_2}d\zeta}{(1-\zeta)^{(K_2+2)/2+m}}.\end{aligned}\qquad\text{...(A.12)}$$

We now note the following Pochhammer integral representation of the ${}_1F_1$ function (Slater (1960, p. 38), Erdéyli (1953, p. 272)):

$$\int_\eta e^{x\zeta}\zeta^{a-1}(1-\zeta)^{b-a-1}d\zeta=\frac{\Gamma(a)\Gamma(b-a)[1-e^{2(b-a)\pi i}]}{\Gamma(b)}\,{}_1F_1(a,b;x)$$

if Re $(a)>0$ and $b-a$ is not a positive integer. Since $K_2>0$ and $-(K_2/2+m)$ is not a positive integer we can use the above integral in (A.12). We obtain

$$\begin{aligned}\mathrm{pdf}\,(r)=&\frac{(-1)^{K_2/2}\exp\left\{-\frac{\mu^2}{2}(1+\beta^2+B_1^2(r))\right\}}{2\pi i c^{K_2+1}}\sum_{m=0}^{\infty}\frac{(\mu^2/2)^m B_1^{2m}(r)}{m!}\\
&\times 2\left[K_2\frac{\Gamma(K_2+1)\Gamma\left(-\frac{K_2}{2}-m\right)}{\Gamma\left(\frac{K_2}{2}-m+1\right)}\,{}_1F_1\left(K_2+1,\frac{K_2}{2}-m+1;\frac{\mu^2}{2}B_2^2(r)\right)\right.\\
&-\frac{\mu^2(1+\beta a_2)^2}{c^2}\,\frac{\Gamma(K_2+3)\Gamma\left(-\frac{K_2}{2}-m-1\right)}{\Gamma\left(\frac{K_2}{2}-m+2\right)}\,{}_1F_1\left(K_2+3,\frac{K_2}{2}-m+2;\frac{\mu^2}{2}B_2^2(r)\right)\\
&-\frac{2\beta c(1+\beta a_2)}{c^2}\,\frac{\Gamma(K_2+2)\Gamma\left(-\frac{K_2}{2}-m-1\right)}{\Gamma\left(\frac{K_2}{2}-m+1\right)}\,{}_1F_1\left(K_2+2,\frac{K_2}{2}-m+1;\frac{\mu^2}{2}B_2^2(r)\right)\\
&\left.-\beta^2\frac{\Gamma(K_2+1)\Gamma\left(-\frac{K_2}{2}-m-1\right)}{\Gamma\left(\frac{K_2}{2}-m\right)}\,{}_1F_1\left(K_2+1,\frac{K_2}{2}-m;\frac{\mu^2}{2}B_2^2(r)\right)\right].\end{aligned}$$

A(ii) *An Equation With $n+1$ Endogenous Variables*

We use the structural equation

$$y_1=Y_2\beta+Z_1\gamma+u$$

given in (10) and we write the reduced form of the endogenous variables of this equation as

$$[y_1:Y_2]=[Z_1:Z_2]\begin{bmatrix}\pi_{11} & \Pi_{12}\\ \pi_{21} & \Pi_{22}\end{bmatrix}+[v_1:V_2]$$

where Π_{22} is a $K_2\times n$ matrix of rank n, $(K_2\geqq n)$. We assume that the exogenous variables have been orthogonalized so that $T^{-1}Z'Z=I_K$ where $Z=[Z_1:Z_2]$ and $K=K_1+K_2$ and that the rows of $[v_1:V_2]$ are independent normal with mean vector zero and covariance matrix I_n (i.e. the usual standardizing transformations have been carried out). We also write $\Pi_{22}'\Pi_{22}=\bar{\Pi}_{22}'\bar{\Pi}_{22}$ where $\bar{\Pi}_{22}$ is an $n\times n$ non-singular matrix.

The joint probability density function of the 2SLS estimator (see (11) in the paper) is given by (Phillips (1978*b*))

$$\text{pdf}(r) = \frac{\text{etr}\left\{-\frac{T}{2}(I+\beta\beta')\bar{\Pi}_{22}'\bar{\Pi}_{22}\right\}\Gamma_n\left(\frac{N+n+1}{2}\right)}{\pi^{n/2}[\det(I+rr')]^{(N+n+1)/2}}$$

$$\cdot \sum_{j=0}^{\infty} \frac{\left(\frac{N}{2}\right)_j}{j!\Gamma_n\left(\frac{N+n}{2}+j\right)} \left[\left(\frac{T}{2}\beta'\bar{\Pi}_{22}'\left(\text{adj}\frac{\partial}{\partial W}\right)\bar{\Pi}_{22}\beta\right)^j\right.$$

$$\cdot (\det(I+W))^{(N+n)/2+j-(n+1)}$$

$$\left.\cdot {}_1F_1\left(\frac{N+n+1}{2}, \frac{N+n}{2}+j; \frac{T}{2}(I+W)\bar{\Pi}_{22}(I+\beta r')(I+rr')^{-1}(I+r\beta')\bar{\Pi}_{22}'\right)\right]_{W=0}$$

where $N = K_2 - n$ = degree of overidentification of the structural equation. As pointed out in Section 2 of the paper, the leading term of this density reveals the order to which moments are finite. The term involves the factor $[\det(I+rr')]^{-(N+n+1)/2} = (1+r'r)^{-(N+n+1)/2}$ which is similar in form to the principal factor of a multivariate *t*-density when $N > 0$ and a multivariate Cauchy-density when $N = 0$.

When $n = 1$, the above multivariate density reduces to the univariate density function for the 2SLS estimator in the two endogenous variable case, i.e.

$$\text{pdf}(r) = \frac{\exp\left\{-\frac{\mu^2}{2}(1+\beta^2)\right\}}{B\left(\frac{1}{2}, \frac{N+1}{2}\right)(1+r^2)^{(N+2)/2}} \sum_{j=0}^{\infty} \frac{\left(\frac{N}{2}\right)_j}{j!\left(\frac{N+1}{2}\right)_j}\left(\frac{\mu^2}{2}\beta^2\right)^j$$

$$\cdot {}_1F_1\left(\frac{N+2}{2}, \frac{N+1}{2}+j, \frac{\mu^2}{2}\frac{(1+\beta r)^2}{1+r^2}\right)$$

where $\mu^2 = T\bar{\pi}_{22}^2 = T\pi_{22}'\pi_{22}$ is the concentration parameter and $N = K_2 - 1$ in this case.

An asymptotic expansion of the joint probability density function in the case of general n and N can be derived from the formula given above. Marginal density functions can then be approximated by the numerical integration of this approximate joint density for low values of n. Some numerical computations are reported in Phillips (1978*b*) for the case $n = 2$.

APPENDIX B: DIAGRAMMATIC STRUCTURE OF CERTAIN ASPECTS OF ASYMPTOTIC THEORY

This diagram on p. 208 is not intended to be in any way comprehensive but should indicate the main branches of asymptotic theory which lead to interesting asymptotic expansions. We use the following notation:

$F_T(x)$ = finite sample distribution function of statistic of interest.

$$I(x) = \int_{\infty}^{x} i(y)dy.$$

$f_T(x)$ = finite sample probability density of statistic of interest.

$$i(x) = (1/\sqrt{2\pi})\exp(-x^2/2).$$

Asymptotic Theory

Classical Central Limit Theorems

Standardized sums of i.i.d.r.v.'s $\{X_t\}$

$Z_T = \sqrt{T}\{(X_1 + \ldots + X_T)/T - m\}/\sigma$

$F_T(x) \to I(x)$

$f_T(x) \to i(x)$

Extensions

Cramér (1928, 1937)

Ibragimov and Linnik (1971)

useful summary of results

$F_T(x) = I(x) + i(x)\sum_{j=1}^{k} Q_j(x)T^{-j/2} + o(T^{-k/2})$

$f_T(x) = i(x)\{1 + \sum_{j=1}^{k} P_j(x)T^{-j/2}\} + o(T^{-k/2})$

Refinements of Error Term

Berry–Esseen (1945)

$\sup_x |F_T(x) - I(x)| \leq \dfrac{CE|X_t|^3}{\sqrt{T}\sigma^3}$, $C = 2{\cdot}05$ (an absolute constant)

Petrov (1964)

$|f_T(x) - i(x)\{1 + \sum_{j=1}^{k} P_j(x)T^{-j/2}\}| < \dfrac{\eta(T)}{T^{k/2}(1+|x|^{k+2})}$

where $\eta(T) = o(1)$ as $T \to \infty$

Multidimensional Generalizations

Sazonov (1968)

$\sup_{x\in R^k} |F_T(x) - I(x)| \leq C(k)\sum_t \left(\dfrac{\Lambda_{tt}}{\Lambda}\rho_t\right)\dfrac{1}{\sqrt{T}}$

$\rho_t = E|X_t|^3/\sigma_t^3$

Λ = determinant of matrix of correlation coefficients

Λ_{tt} = leading minors

$C(k)$ = an absolute constant depending only on k.

Dependent Variable Limit Theorems

Diananda (1953), Rozanov (1967), Anderson (1971), Hannan (1970), Schönfeld (1971)

Central limit theorems for standardized sums of stationary sequences such as

$Z_T = \sum_{t=1}^{T} y_t/\sqrt{T}$ $y_t = \sum_{s=-\infty}^{\infty} \gamma_s v_{t-s}, \sum_{s=-\infty}^{\infty} |\gamma_s| < \infty$

where $\{v_t\}$ is a sequence of i.i.d.r.v.'s

$Z_T = d_T^{-1}\sum_{t=1}^{T} x_t y_t$ $d_T = (\sum_{t=1}^{T} x_t^2)^{\frac{1}{2}}$

where $\{x_t\}$ is a non-random sequence and $\{y_t\}$ satisfies a uniform mixing condition and conditions on fourth order cumulants.

Statistics more general than Standardized Means

Wallace (1958), Bhattacharya and Ghosh (1978)

expansions for functions of sample means

Chambers (1967)

$\sqrt{T}q$ with rth cumulants $O(T^{(-r/2)+1})$

algorithms for functions of statistics

Sargan (1975)

error in an econometric estimator (classical models)

$\hat{\beta} - \beta = e_T(p, w)$

where

$p \sim N(0, \Omega/T)$

w independent of p

$\sqrt{T}w$ with rth cumulants $O(T^{(-r/2)+1})$

$e_T(\ ) \in C^{k+1}$

Sargan (1976*a*)

specialization of Chambers algorithm and formulae detailed up to $O(T^{-1})$

Phillips (1977*b*)

generalization of Sargan (1975)

error in estimator (classical and non-classical models)

$\hat{\beta} - \beta = e_T(q)$ where distribution of

$\sqrt{T}q$ has Edgeworth expansion and cumulants $O(T^{(-r/2)+1})$

$e_T(\ ) \in C^{k+1}$

Asymptotic χ^2 statistics

Sargan (1976*b*)

expansion of the distribution of χ^2 type criteria in terms of χ^2 distribution and its derivatives (classical models)

Large Deviation Limit Theory

Standardized sums of i.i.d.r.v.'s $\{X_t\}$

$Z_T = \sqrt{T}\{(x_1 + \ldots + x_T)/T - m\}/\sigma$

Cramér (1938)

$F_T(-x) = I(-x)\exp\left\{-\dfrac{x^3}{\sqrt{T}}\Psi\left(-\dfrac{x}{\sqrt{T}}\right)\right\}\left[1 + O\left(\dfrac{x}{\sqrt{T}}\right)\right]$

$1 - F_T(x) = (1 - I(x)\exp\left\{\dfrac{x^3}{\sqrt{T}}\Psi\left(\dfrac{x}{\sqrt{T}}\right)\right\}\left[1 + O\left(\dfrac{x}{\sqrt{T}}\right)\right]$

for $x = 0(\sqrt{T})$ and where

$\psi(z) = \psi_0 + \psi_1 z + \ldots$

under condition $E(\exp(a|X_t|)) < \infty$.

Richter (1957)

$f_T(x) = i(x)\exp\left\{\dfrac{x^3}{\sqrt{T}}\Psi\left(\dfrac{x}{\sqrt{T}}\right)\right\}\left[1 + O\left(\dfrac{x}{\sqrt{T}}\right)\right]$

Ibragimov and Linnik (1971)

useful summary of results

Extensions

Richter (1957)

local theorem for non-identically distributed summands

Richter (1958)

multidimensional i.i.d. r vectors

Petrov (1968)

integral theorem for non-identically distributed summands

Petrov (1975)

useful summary of results

Statistics more general than standardized means

Phillips (1977*c*)

$\sqrt{T}q$ with rth cumulants $O(T^{(-r/2)+1})$ and tail condition on characteristic function

Phillips (1976)

algorithm for functions of statistics $t = \sqrt{T}e_T(q)$

formula for large deviations up to $t = O(T^{\frac{1}{6}})$

APPENDIX C: CUMULANTS OF $\sqrt{T}q$ UP TO THE FOURTH ORDER

We use the following notation for the derivatives of the second characteristic (or cumulant generating function): subscripts a, b, c and d denote values in $\{1, 2, 3\}$ and subscripts p, q, r and s denote values in $\{4, 5\}$. For instance,

$$\lambda abc = \frac{\partial^3 \lambda(0)}{\partial t_a \partial t_b \partial t_c} \qquad (a, b, c = 1, 2, 3)$$

$$\lambda abp = \frac{\partial^3 \lambda(0)}{\partial t_a \partial t_b \partial t_p} \qquad \begin{matrix}(a, b = 1, 2, 3)\\(p = 4, 5)\end{matrix}$$

and

$$\lambda abpq = \frac{\partial^4 \lambda(0)}{\partial t_a \partial t_b \partial t_p \partial t_q} \qquad \begin{matrix}(a, b = 1, \ldots, 3)\\(p, q = 4, 5)\end{matrix}.$$

All first derivatives are zero, since q is standardized about its mean, and higher order derivatives are as follows:

$$\lambda ab = -\frac{2}{T} \operatorname{tr}(A_a \Omega A_b \Omega) - \frac{4}{T} m' A_a \Omega A_b m$$

$$\lambda ap = -\frac{2}{T} b_p' \Omega A_a m$$

$$\lambda pq = -\frac{1}{T} b_p' \Omega b_p$$

$$\lambda abc = -\frac{4i}{T^{\frac{3}{2}}}[\operatorname{tr}(A_c \Omega A_b \Omega A_a \Omega) + \operatorname{tr}(A_b \Omega A_c \Omega A_a \Omega)]$$
$$-\frac{8i}{T^{\frac{3}{2}}} m'[A_a \Omega A_b \Omega A_c + A_a \Omega A_c \Omega A_b + A_c \Omega A_a \Omega A_b] m$$

$$\lambda abp = -\frac{4i}{T^{\frac{3}{2}}} m'[A_b \Omega A_a \Omega + A_a \Omega A_b \Omega] b_p$$

$$\lambda apq = -\frac{2i}{T^{\frac{3}{2}}} b_p' \Omega A_a \Omega b_q$$

$$\lambda pqr = 0$$

$$\lambda abcd = \frac{8}{T^2}[\operatorname{tr}(A_d \Omega A_c \Omega A_b \Omega A_a \Omega) + \operatorname{tr}(A_c \Omega A_d \Omega A_b \Omega A_a \Omega) + \operatorname{tr}(A_c \Omega A_b \Omega A_d \Omega A_a \Omega)$$
$$+ \operatorname{tr}(A_d \Omega A_b \Omega A_c \Omega A_a) + \operatorname{tr}(A_b \Omega A_d \Omega A_c \Omega A_a \Omega)$$
$$+ \operatorname{tr}(A_b \Omega A_c \Omega A_d \Omega A_a \Omega)] + \frac{16}{T^2} m'[A_d \Omega A_c \Omega A_b \Omega A_a$$
$$+ A_d \Omega A_b \Omega A_c \Omega A_a + A_d \Omega A_b \Omega A_a \Omega A_c + A_d \Omega A_c \Omega A_a \Omega A_b$$
$$+ A_d \Omega A_a \Omega A_c \Omega A_b + A_d \Omega A_a \Omega A_b \Omega A_c] m$$
$$+ \frac{16}{T^2} m'[A_c \Omega A_d \Omega A_b \Omega A_a + A_c \Omega A_b \Omega A_d \Omega A_a$$
$$+ A_b \Omega A_d \Omega A_c \Omega A_a + A_b \Omega A_c \Omega A_d \Omega A_a + A_b \Omega A_d \Omega A_a \Omega A_c$$
$$+ A_b \Omega A_a \Omega A_d \Omega A_c] m$$

$$\lambda abcp = \frac{8}{T^2} m'[A_c\Omega A_b\Omega A_a\Omega + A_b\Omega A_c\Omega A_a\Omega + A_b\Omega A_a\Omega A_c\Omega + A_c\Omega A_a\Omega A_b\Omega$$

$$+ A_a\Omega A_c\Omega A_b\Omega + A_a\Omega A_b\Omega A_c\Omega]b_p$$

$$\lambda abpq = \frac{4}{T^2} b'_p[\Omega A_b\Omega A_a\Omega + \Omega A_a\Omega A_b\Omega]b_q$$

$$\lambda apqr = 0$$

$$\lambda pqrs = 0.$$

APPENDIX D: LIMITS IN PROBABILITY OF THE LEAST SQUARES ESTIMATORS

If we take a simple parametric model as the process determining the exogenous series I_t then we can readily calculate the limits in probability of the least squares estimators of α and β in (21). Suppose we have, as in (32) in the main body of the paper,

$$I_t = \rho I_{t-1} + v_t, \qquad |\rho| < 1$$

with the v_t i.i.d. $N(0, \sigma_v^2)$ and independent of the u_t in (21). Then, conditional on a given realization of this I_t series, we have the following expectations for finite T:

$$\mu_a = T^{-1}E(x'A_a x) = T^{-1}\{\mathrm{tr}\,(A_a\Omega) + m'A_a m\}, \qquad (a = 1, 2, 3)$$

$$\mu_p = T^{-1}E(b'_p x), \qquad (p = 4, 5)$$

$$\mu = T^{-1}d'd.$$

Now, using bars to indicate limits of the above as $T \to \infty$, we extract by routine manipulations the following limits in probability of the least squares estimators:

$$\mathrm{plim}_{T\to\infty}\, \alpha^* = \frac{\bar{\mu}_1(\bar{\mu}_3 + \bar{\mu}_5) - \bar{\mu}_2(\bar{\mu}_2 + \bar{\mu}_4)}{\bar{\mu}_1(\bar{\mu} + \bar{\mu}_3 + 2\bar{\mu}_5) - (\bar{\mu}_2 + \bar{\mu}_4)^2}, \qquad \ldots(\mathrm{D.1})$$

$$\mathrm{plim}_{T\to\infty}\, \beta^* = \frac{\bar{\mu}_2(\bar{\mu} + \bar{\mu}_3 + 2\bar{\mu}_5) - (\bar{\mu}_2 + \bar{\mu}_4)(\bar{\mu}_3 + \bar{\mu}_5)}{\bar{\mu}_1(\bar{\mu} + \bar{\mu}_3 + 2\bar{\mu}_5) - (\bar{\mu}_2 + \bar{\mu}_4)^2}, \qquad \ldots(\mathrm{D.2})$$

where

$$\bar{\mu}_1 = \frac{\sigma_u^2}{(1-\alpha)^2(1-\delta^2)} + \frac{\alpha^2\sigma_v^2}{(1-\alpha)^2}\left(\frac{1+\rho\delta}{1-\rho\delta}\right)\left(\frac{1}{1-\delta^2}\right)\left(\frac{1}{1-\rho^2}\right),$$

$$\bar{\mu}_2 = \frac{\sigma_u^2}{(1-\alpha)^2(1-\delta^2)} + \frac{\alpha^2\sigma_v^2(\rho+\delta)}{(1-\alpha)^2(1-\rho\delta)(1-\delta^2)(1-\rho^2)},$$

$$\bar{\mu}_3 = \bar{\mu}_1, \qquad \bar{\mu} = \sigma_v^2/(1-\rho^2), \qquad \bar{\mu}_4 = \frac{\alpha\rho\sigma_v^2}{(1-\alpha)(1-\rho^2)}\left(\frac{1}{1-\rho\delta}\right),$$

$$\bar{\mu}_5 = \frac{\alpha\sigma_v^2}{(1-\alpha)(1-\rho^2)}\left(\frac{1}{1-\rho\delta}\right).$$

The following Tables show the numerical values of the above probability limits (D.1) and (D.2) in certain cases. The values in the Table approximate moderately well the general location of the approximate small sample distributions of the least squares estimators illustrated in Figures 1–6 of Appendix E. These results support the view held by Hendry ((1973), (1974), (1976)) based on experimental methods concerning the relevance of asymptotic theory in helping to characterize finite sample distributions in dynamic models.

Both Tables D(a) and D(b) make it clear that a critical parameter in determining the magnitude of the asymptotic bias of the estimators is the signal to noise ratio. When var $(I_t)/\text{var}\ (u_t) = \sigma_v^2/\{(1-\rho^2)\sigma_u^2\} = 10{\cdot}416$ $(\sigma_u^2 = 0{\cdot}1, \sigma_v^2 = 1{\cdot}0, \rho = 0{\cdot}2)$ we have a 21 per cent asymptotic bias for α^* and a 12 per cent bias for β^*; when var $(I_t)/\text{var}\ (u_t) = 1{\cdot}0416$ the corresponding figures are a 114·5 per cent bias for α^* and a 52·7 per cent bias for β^*. The values in the third column of Tables D(a) and D(b) help to give us some idea of the effect of the system dynamics on the asymptotic bias. (We change the value of α to bring it closer to that of the long run marginal propensity to consume $\alpha/(1-\beta)$ in the cases where $\beta \neq 0$). When $\sigma_u^2 = 0{\cdot}1, \sigma_v^2 = 1{\cdot}0$ and $\rho = 0{\cdot}2$, for instance, we now find an asymptotic bias for α^* of 6 per cent.

TABLE D

(a) $\rho = 0{\cdot}2$

	$\alpha = 0{\cdot}3$, $\beta = 0{\cdot}4$	$\alpha = 0{\cdot}2$, $\beta = 0{\cdot}7$	$\alpha = 0{\cdot}6$, $\beta = 0{\cdot}0$	True parameter values	
plim α^*	0·3622	0·2703	0·6361	$\sigma_u^2 = 0{\cdot}1$	$\sigma_v^2 = 1{\cdot}0$
	0·5278	0·4596	0·7312	$\sigma_u^2 = 0{\cdot}5$	$\sigma_v^2 = 1{\cdot}0$
	0·6434	0·5919	0·7970	$\sigma_u^2 = 1{\cdot}0$	$\sigma_v^2 = 1{\cdot}0$
plim β^*	0·3509	0·6333	−0·0095	$\sigma_u^2 = 0{\cdot}1$	$\sigma_v^2 = 1{\cdot}0$
	0·2523	0·4673	−0·0188	$\sigma_u^2 = 0{\cdot}5$	$\sigma_v^2 = 1{\cdot}0$
	0·1892	0·3527	−0·0179	$\sigma_u^2 = 1{\cdot}0$	$\sigma_v^2 = 1{\cdot}0$

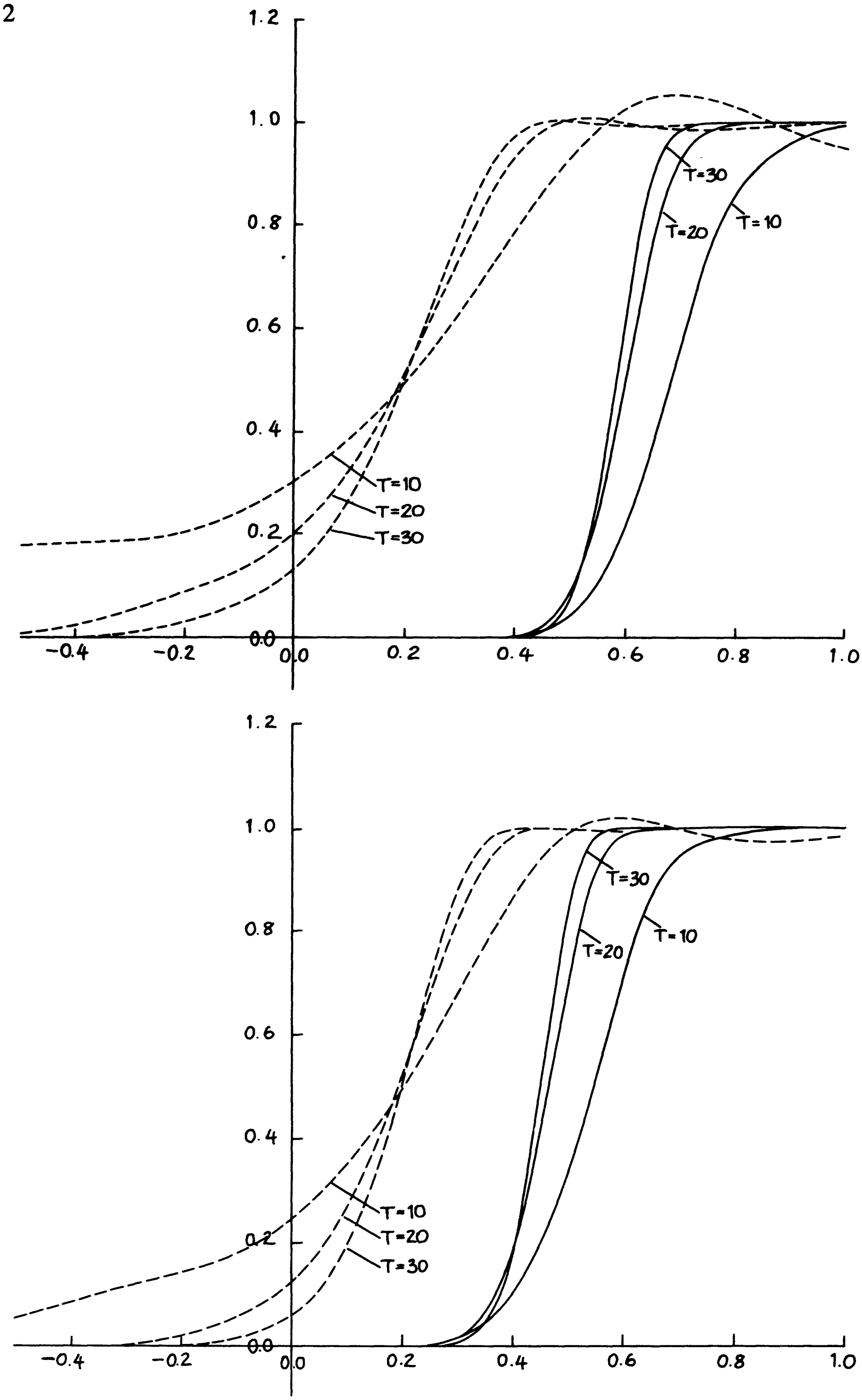
(b) $\rho = 0{\cdot}8$

	$\alpha = 0{\cdot}3$, $\beta = 0{\cdot}4$	$\alpha = 0{\cdot}2$, $\beta = 0{\cdot}7$	$\alpha = 0{\cdot}06$, $\beta = 0{\cdot}00$	True parameter values	
plim α^*	0·3449	0·2368	0·6317	$\sigma_u^2 = 0{\cdot}1$	$\sigma_v^2 = 1{\cdot}0$
	0·4452	0·3417	0·6956	$\sigma_u^2 = 0{\cdot}5$	$\sigma_v^2 = 1{\cdot}0$
	0·5208	0·4311	0·7385	$\sigma_u^2 = 1{\cdot}0$	$\sigma_v^2 = 1{\cdot}0$
plim β^*	0·3263	0·6503	−0·0384	$\sigma_u^2 = 0{\cdot}1$	$\sigma_v^2 = 1{\cdot}0$
	0·2145	0·5324	−0·0850	$\sigma_u^2 = 0{\cdot}5$	$\sigma_v^2 = 1{\cdot}0$
	0·1645	0·4486	−0·0923	$\sigma_u^2 = 1{\cdot}0$	$\sigma_v^2 = 1{\cdot}0$

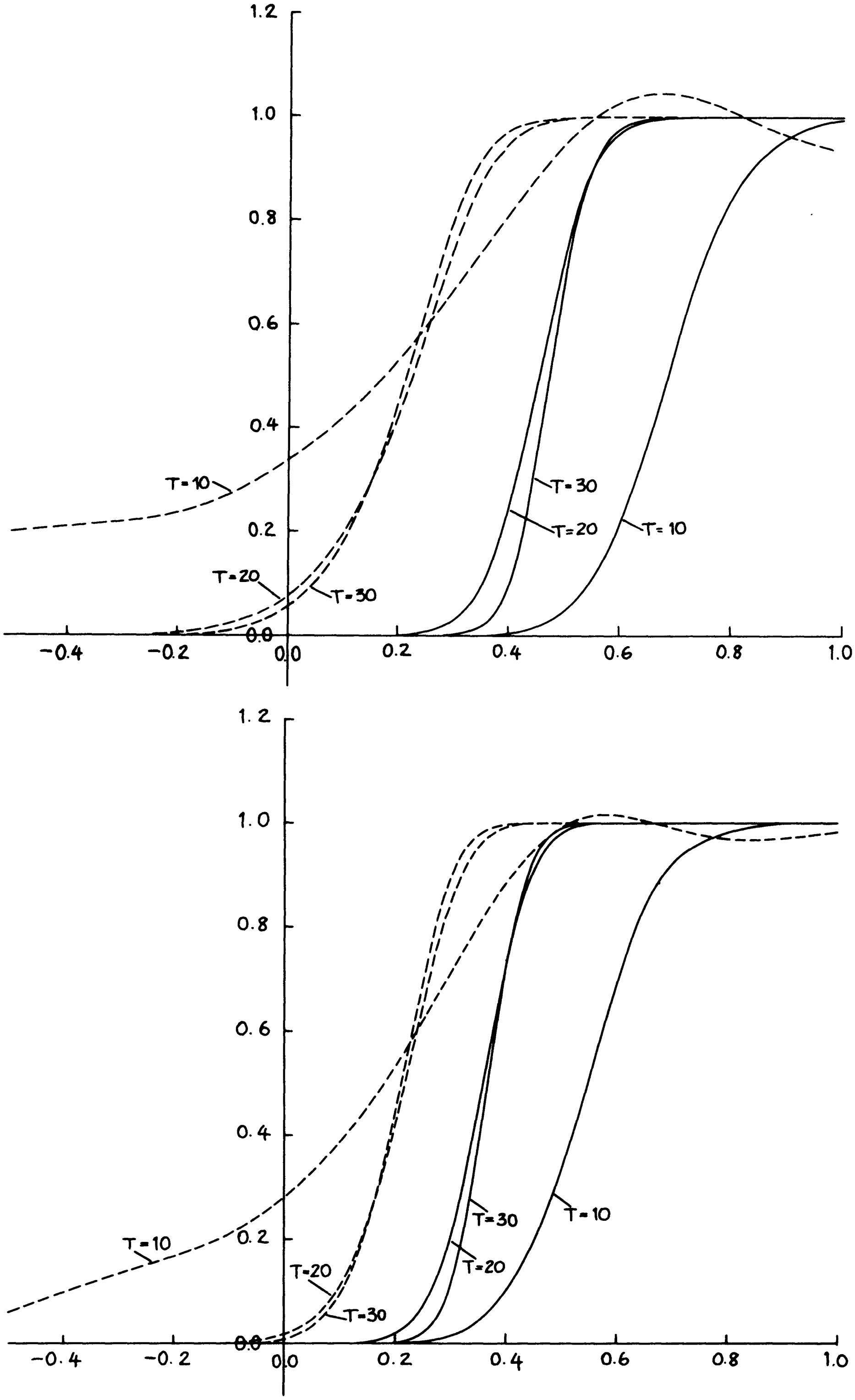
APPENDIX E: FIGURES 1–6

Approximate distributions are graphed for 2SLS (displayed by the broken line: - - - -) and OLS (displayed by the unbroken line: ———) estimators of α (Figures 1–3) and β (Figures 4–6) for various values of the sample size T and the true values of the parameters as well as the two types of exogenous series given by (32) and (33) in the paper. The sub-classifications (a) and (b) in the numbering of the Figures refer to the different values of σ_u^2, the variance of the disturbance on the consumption function (21). Thus (a) is associated with $\sigma_u^2 = 1{\cdot}0$ and (b) with $\sigma_u^2 = 0{\cdot}5$.

To facilitate reading of the figures the parameter values and type of exogenous series are listed at the foot of each figure.



FIGURES 1A and 1B
Distributions of OLS and 2 SLS estimators of α
(A) $\alpha = 0{\cdot}2 \quad \beta = 0{\cdot}7 \quad \sigma_u^2 = 1{\cdot}0 \quad I_t = \rho I_{t-1} + \nu_t \quad \rho = 0{\cdot}2$
(B) $\alpha = 0{\cdot}2 \quad \beta = 0{\cdot}7 \quad \sigma_u^2 = 0{\cdot}5 \quad I_t = \rho I_{t-1} + \nu_t \quad \rho = 0{\cdot}2$

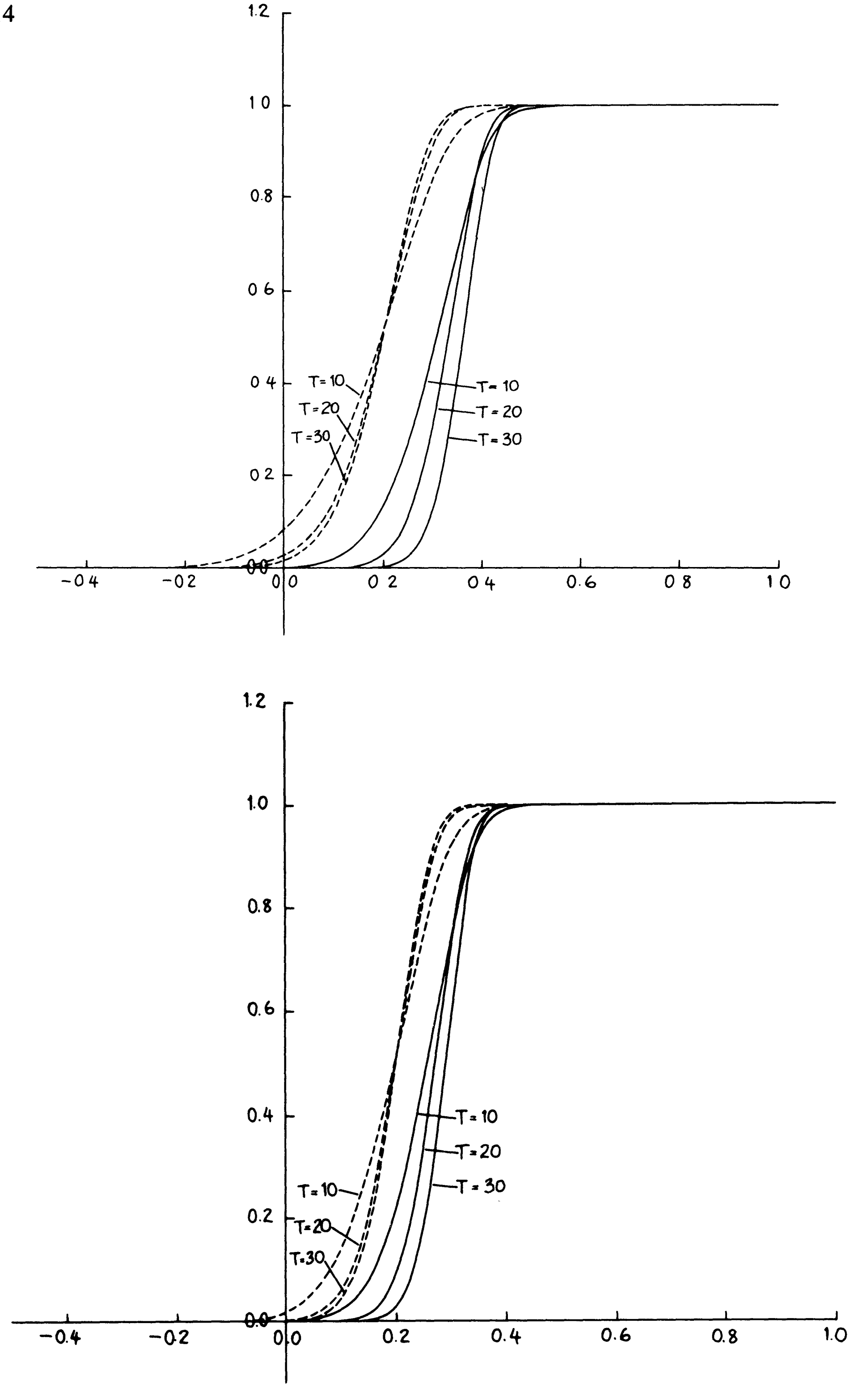


FIGURES 2A and 2B

Distributions of OLS and 2SLS estimators of α

(A) $\alpha = 0{\cdot}2 \quad \beta = 0{\cdot}7 \quad \sigma_u^2 = 1{\cdot}0 \quad I_t = \rho I_{t-1} + v_t \quad \rho = 0{\cdot}8$

(B) $\alpha = 0{\cdot}2 \quad \beta = 0{\cdot}7 \quad \sigma_u^2 = 0{\cdot}5 \quad I_t = \rho I_{t-1} + v_t \quad \rho = 0{\cdot}8$

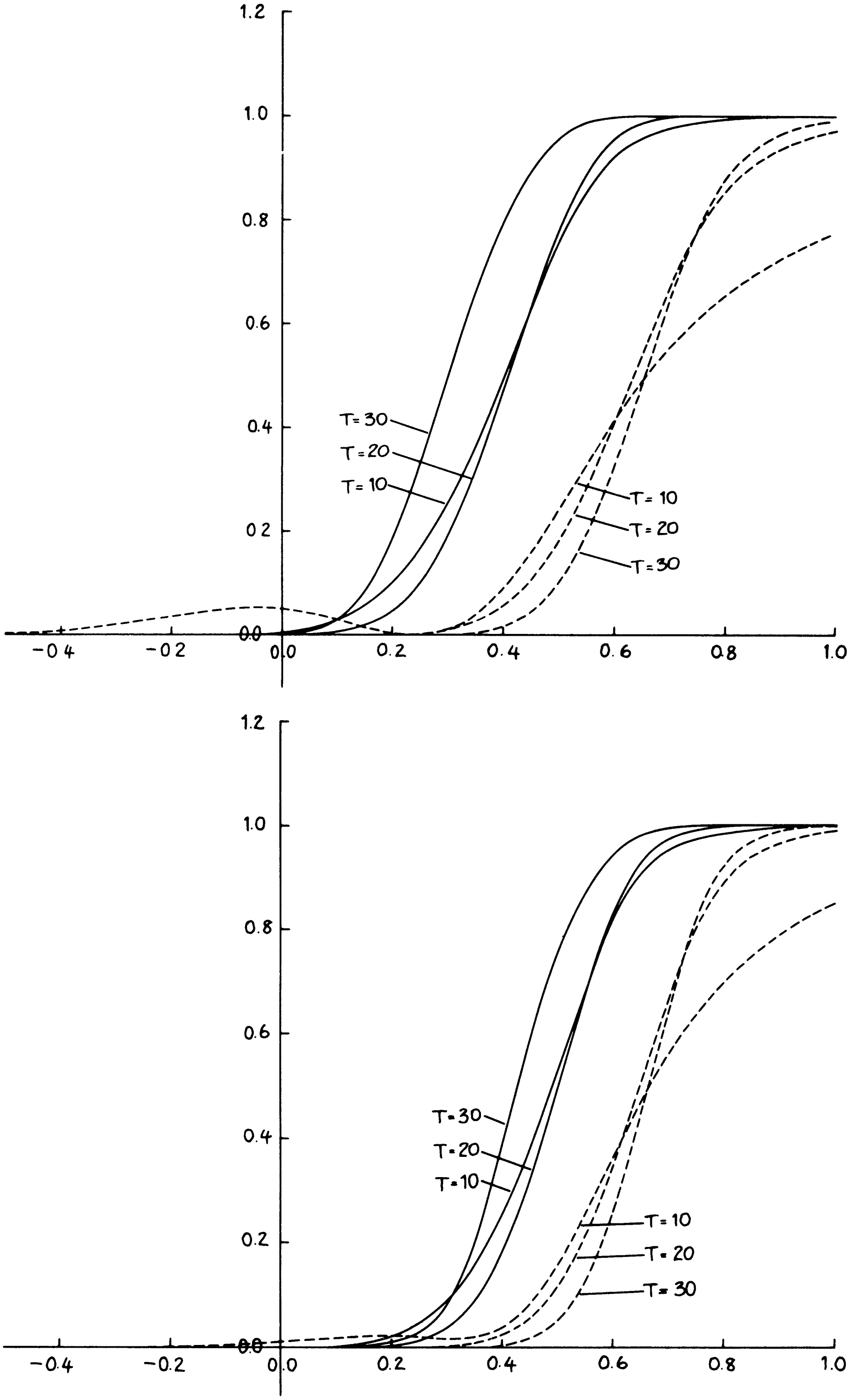


FIGURES 3A and 3B

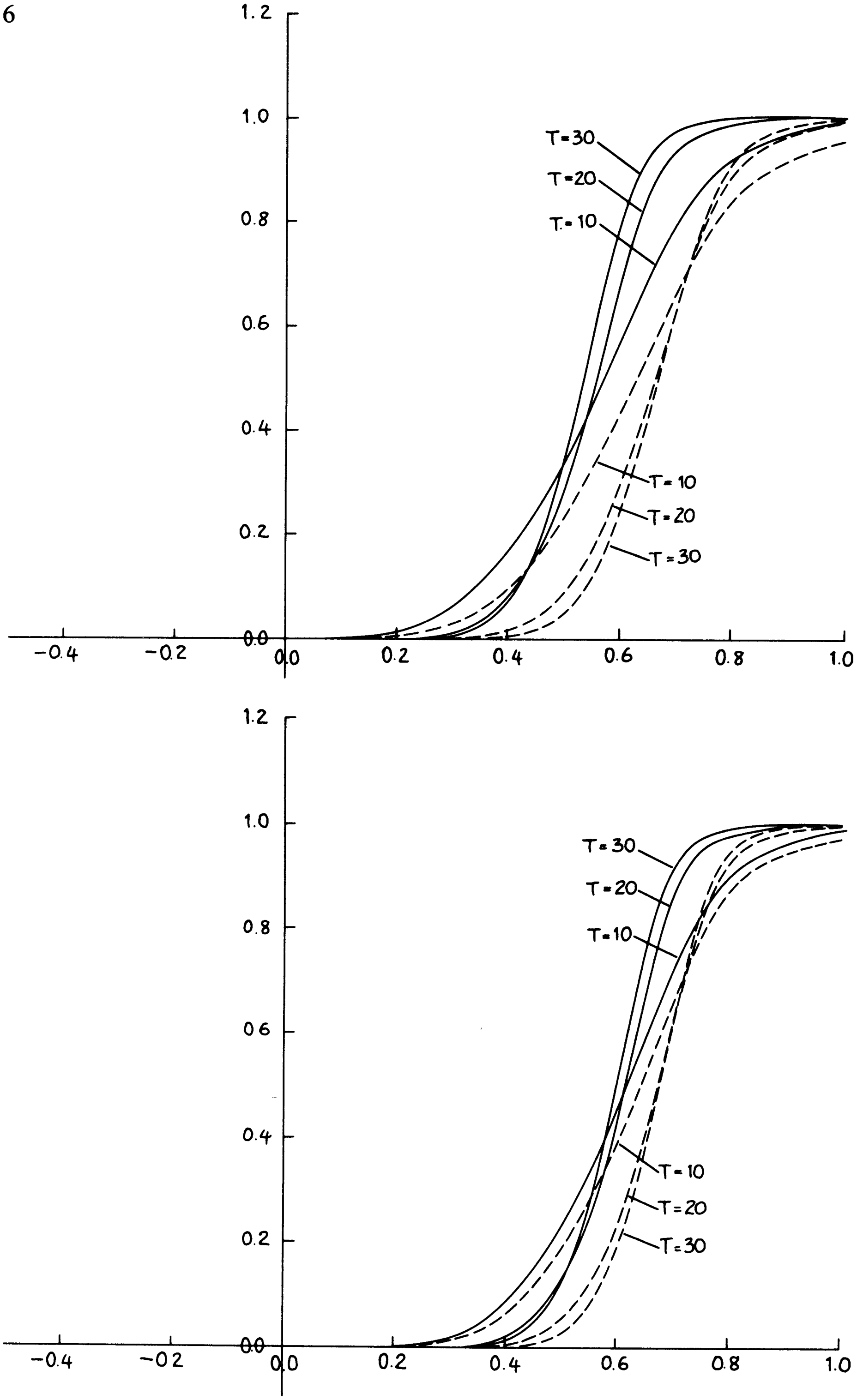
Distributions of OLS and 2SLS estimators of α

(A) $\alpha = 0{\cdot}2$ $\beta = 0{\cdot}7$ $\sigma_u^2 = 1{\cdot}0$ $I_t = \rho_1 I_{t-1} + \rho_2 I_{t-2} + v_t$ $\rho_1 = 0{\cdot}75$ $\rho_2 = -0{\cdot}5$

(B) $\alpha = 0{\cdot}2$ $\beta = 0{\cdot}7$ $\sigma_u^2 = 0{\cdot}5$ $I_t = \rho_1 I_{t-1} + \rho_2 I_{t-2} + v_t$ $\rho_1 = 0{\cdot}75$ $\rho_2 = -0{\cdot}5$



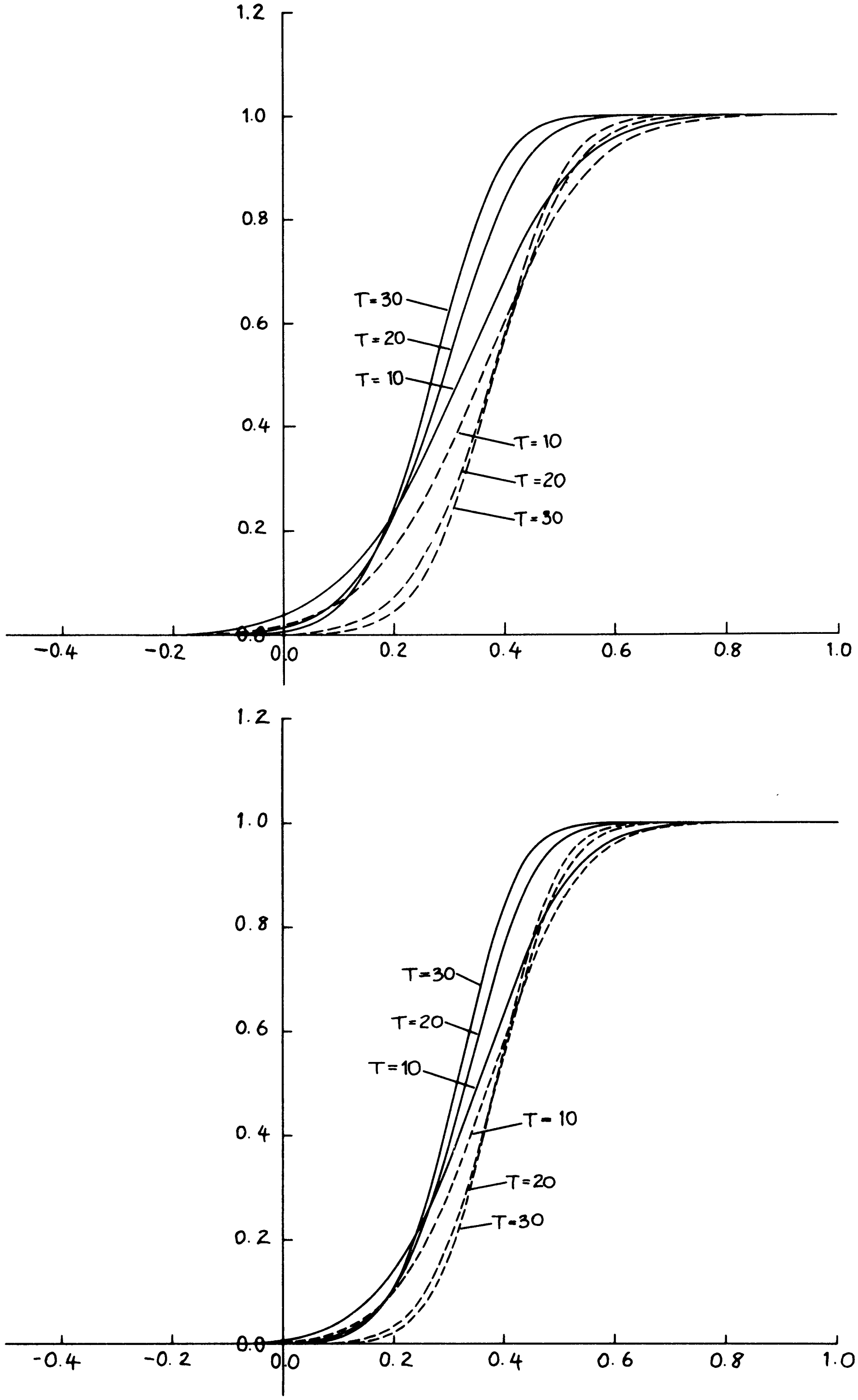
FIGURES 4A and 4B
Distributions of OLS and 2SLS estimators of β
(A) $\alpha=0{\cdot}2 \quad \beta=0{\cdot}7 \quad \sigma_u^2=1{\cdot}0 \quad I_t=\rho I_{t-1}+v_t \quad \rho=0{\cdot}8$
(B) $\alpha=0{\cdot}2 \quad \beta=0{\cdot}7 \quad \sigma_u^2=0{\cdot}5 \quad I_t=\rho I_{t-1}+v_t \quad \rho=0{\cdot}8$



FIGURES 5A and 5B
Distributions of OLS and 2SLS estimators of β

(A) $\alpha = 0{\cdot}2$ $\beta = 0{\cdot}7$ $\sigma_u^2 = 1{\cdot}0$ $I_t = \rho_1 I_{t-1} + \rho_2 I_{t-2} + v_t$ $\rho_1 = 0{\cdot}75$ $\rho_2 = -0{\cdot}50$

(B) $\alpha = 0{\cdot}2$ $\beta = 0{\cdot}7$ $\sigma_u^2 = 0{\cdot}5$ $I_t = \rho_1 I_{t-1} + \rho_2 I_{t-2} + v_t$ $\rho_1 = 0{\cdot}75$ $\rho_2 = -0{\cdot}50$



FIGURES 6A and 6B

Distributions of OLS and 2SLS estimators of β

(A) $\alpha=0{\cdot}3$ $\beta=0{\cdot}4$ $\sigma_u^2=1{\cdot}0$ $I_t=\rho_1 I_{t-1}+\rho_2 I_{t-2}+v_t$ $\rho_1=0{\cdot}75$ $\rho_2=-0{\cdot}50$

(B) $\alpha=0{\cdot}3$ $\beta=0{\cdot}4$ $\sigma_u^2=0{\cdot}5$ $I_t=\rho_1 I_{t-1}+\rho_2 I_{t-2}+v_t$ $\rho_1=0{\cdot}75$ $\rho_2=-0{\cdot}50$

A preliminary version of this paper was first prepared in early 1976 and has formed the basis of a number of seminars that I have given over the last two years. Many people have commented on the paper and/or the seminars and I have benefited particularly from discussions with W. C. Brainard, H. E. Daniels, J. Hartigan, A. Holly, E. Maasoumi and J. D. Sargan. Mervyn King made some helpful comments on the presentation of the paper, for which I am grateful. I am particularly indebted to Ralph Bailey for extensive work in programming the computations in the paper; without his help the numerical results reported in Section 7 of the paper would not have been obtained. I also thank Michael Prior for some programming help in the early stages of this project and Roger Bloxsom for some research assistance. The research was supported by the SSRC under Grant Number HR 3432 and the paper was prepared, in part, while I was a visitor at the Cowles Foundation for Research in Economics, Yale University.

NOTES

1. Bergstrom's paper can be taken with the important and fundamental paper by Basmann (1961) as initiating the mathematical study of the small sample behaviour of various estimators of the parameters in simultaneous equations systems. The other important landmark in the emergence of studies in this area was the work of Nagar (1959) on moment approximations.

2. Basmann (1963*b*) derived the exact density function and mean of the 2SLS estimator of the marginal propensity to consume in an overidentified case.

3. As Bergstrom remarks in his paper (p. 480) a number of useful Monte Carlo studies were available but the natural disadvantages of these studies made an exact, mathematical study appealing.

4. A more complete survey of exact theory up to the early 1970's is contained in Basmann (1974).

5. It may be relevant to mention here that the exact distributions of least squares estimators in simple time series models such as the non-circular first order autoregression are still not known in analytic form.

6. Just as we have discussed earlier that personal judgement is often necessary in assessing the relative merits of different estimators, so too is an element of personal judgement involved in the decision as to what constitutes a *good* approximation. Partly, this decision will rest on an investigator's view of what amounts to an acceptable error of approximation and, partly, his decision will be influenced by the region of the distribution over which he is relying on the errors of the approximation being small. Thus, a certain approximation may involve acceptable errors in the body of the distribution amounting to as much as 95 per cent but involve quite unacceptable errors in the remaining tail areas.

7. Working with a similar model, Richardson and Wu (1971) also carry out a comparison of OLS and 2SLS estimators. Their comparison is based on the analytic expressions for the bias and mean squared errors (when these exist) of the two estimators. They consider all coefficients in the equation under study and a wide range of different parameter values. Expressions for the exact finite sample moments of OLS and 2SLS estimators in the same situation have also been given by Takeuchi (1970).

8. Working independently and a little earlier Richardson (1968) had derived the exact density of the 2SLS estimator in the same context.

9. However Sawa's numerical results revealed that, in certain cases, the superiority of 2SLS over OLS may be slight. This is particularly relevant when the degree of overidentification (N) becomes large and the distribution of the 2SLS estimator is markedly skewed towards that of the OLS estimator. This feature has led to the development of certain combined estimators (based on a linear combination of OLS and 2SLS) designed to improve the location of the estimator when N is large (see, in particular, Sawa (1973*a*, 1973*b*). A comparison of the degree of concentration of 2SLS and LIML estimators recently carried out by Anderson (1974) (see also Phillips and Wickens (1978, ch. 6, Question 6.19)) and based on the first three terms of the Edgeworth expansions of the finite sample distributions suggests that the distribution of the LIML estimator is the more concentrated in such cases (i.e. when N is large), provided we exclude most of the tail areas of the distribution when we measure concentration. The latter condition is related to the fact that LIML estimators possess no integral small sample moments, whereas 2SLS estimators possess moments up to order N (see the ensuing discussion in the paper).

10. This statement refers to the case where there are no exogenous variables in the equation other than a constant. If there are K_1 exogenous variables included in the equation as well as a constant then moments of the OLS estimator of order less than $T-1-K_1$ exist; while, in this case, moments of the 2SLS estimator of order less than K_2 still exist (i.e. integral moments up to order K_2-1 exist, as before, in this case). For detailed analysis in the case of an equation with included exogenous variables, see Richardson and Wu (1971), whose numerical results appear consistent in their general implication with those of Sawa.

11. If there are n equations in the model $n-1$ independent restrictions are necessary for identification. The actual number of restrictions is, from the form of the equation (with two endogenous variables and K_2 excluded exogenous variables), $n-2+K_2=(n-1)+(K_2-1)$. The degree of overidentification is, therefore, taken as K_2-1.

12. The same result has been verified by Sargan (1970) for full information maximum likelihood (FIML) estimators and conjectured by Basmann (1974) to apply generally to LIML and FIML estimators of structural coefficients.

13. In addition to the above exact theory in cases where $n>1$, Rhodes (1977) has obtained a multiple series representation of the density function of the likelihood ratio identifiability test statistic for an equation with $n+1$ endogenous variables. This generalizes the result of McDonald (1972) for the case $n=1$.

14. Muirhead (1975) quotes Sugiyama (1972) as requiring over 100 terms of this type of series before achieving adequate convergence.

15. Of course many other types of approximations can be developed, and some of these are more closely linked to the exact density functions than the Edgeworth approximation (see, for example, Sargan (1976*a*) and Holly and Phillips (1977)). An important area of research now and in the future is the development of alternative methods of approximation and the determination of the conditions under which the different approximations can be expected to do well.

16. This section does not attempt to provide a comprehensive treatment and the interested reader is strongly recommended to consult the survey paper of Wallace (1958), which contains an excellent introduction to this area and a valuable list of references. The recent survey paper by Bhattacharya (1977), the book by Bhattacharya and Rao (1976) and the review paper by Bickel (1974) also contain some useful discussion of the material in this section.

17. We write $f_T(x)$ because it is implicitly assumed that the statistic is computed from a sample of observations of size T, so that T is a parameter of the resulting distribution.

18. More details of the history of the series expansion (1) and the related Edgeworth series are given by Cramér (1972, 1976).

19. See, for instance, Cramér (1946, p. 223).

20. For a complete discussion of asymptotic series the reader is referred to the books by Copson (1967) and De Bruijn (1958).

21. Since $E(Z_T)=0$ and $E(Z_T^2)=1$, and the first three Hermite polynomials are $H_0(x)=1$, $H_1(x)=x$, $H_2(x)=x^2-1$, we have for the first three coefficients in (1)

$$a_0=\int_{-\infty}^{\infty} f_T(x)dx=1$$

$$a_1=-\int_{-\infty}^{\infty} xf_T(x)dx=0$$

and

$$a_2=\int_{-\infty}^{\infty} (x^2-1)f_T(x)dx=1-1=0.$$

The remaining terms in (6) are most easily obtained by considering the characteristic function of Z_T. A full discussion is given by Cramér (1946, pp. 224–227).

22. A somewhat weaker result had been given by Cramér in 1923 (see Cramér (1976, pp. 514–515)).

23. This value is quoted in Petrov (1975, p. 321). A value of 2·05 was obtained by Wallace (1958). The value of 3·00 is given by Feller (1970, p. 542) in a simple proof of (9).

24. However, Bhattacharya (1977) gives a result similar to (9) for statistics which can be represented as smooth functions of standardized sums of independent identically distributed random variables.

25. I have recently mentioned elsewhere (Phillips (1978*a*)) that the idea of such general expansions was even considered in the second part of the seminal paper by Edgeworth (1905). This fact is of some relevance to the reappraisal of Edgeworth's contribution to mathematical statistics that is currently in progress (see, for example, Stigler (1978) and Seal (1967)). Many economists may not be aware that Edgeworth published more than 70 articles on probability and statistics. A review of his contribution to mathematical statistics and a list of publications are given by Bowley (1928).

26. It is assumed that the equation is identified and the matrices Z and Z_1 have full rank.

27. In this simple case we have $h'(\hat{\beta}-\beta)=e_T(p)$ and we need not introduce the extra variates w that were mentioned earlier.

28. For an example see Phillips (1977*a*). Even in cases where we can do this (when, for instance, there are first order sample moments from a normally distributed population), we cannot always go on to assert the statistical independence of these moments and the higher order sample moments (which are not, in such cases, normally distributed).

29. These conditions enable us to transform, subject to an error which is small as $T\to\infty$, probability statements involving $\hat{\beta}-\beta$ into probability statements involving one of the components of q. The remaining conditions ensure that the distribution of q is itself sufficiently well behaved (with moments of a high enough order and the right order of magnitude) to admit a valid series expansion. The latter expansion is obtained essentially in terms of the cumulants of q (c.f. (7) above). These terms are then combined with the terms of the Taylor series expansion of $e_T(q)$ about the origin to obtain the corresponding series expansion for $\sqrt{T}(\hat{\beta}-\beta)$. Details of the formulae are given in Sargan (1976*a*) and Phillips (1977*a*) (the latter in a slightly different notation).

30. One line of approach in verifying these conditions would be to take the general form of the joint characteristic function of linear and quadratic forms that arises in autoregressive models (see equation (30)) and use this form to check the conditions on the distribution of $\sqrt{T}q$.

31. Bhattacharya (1977) gives a theorem similar to that of Chibisov but under slightly weaker conditions.

32. The problem is briefly discussed by Pfanzagl (1973), but seems otherwise to be seldom mentioned.

33. Some further work on this type of expansion of the underlying statistic has been done by Gusev (1976). An asymptotic expansion for the distribution of maximum likelihood estimators is given by Michel (1975).

34. The main works to date (dealing with classical models) are Sargan and Mikhail (1971) and Anderson and Sawa ((1973), (1977)). One interesting application to non-linear regression models (but which does not provide tabulations of the exact and approximate distributions) has recently been published by Ivanov (1976).

35. The best single reference work in this area is the treatise by Ibragimov and Linnik (1971) which contains an extensive survey of research on large deviations up to the late sixties. The seminal paper by Richter (1957) is also a useful reference.

36. This theory is sometimes distinguished from the theory of very large deviations where the argument x is not restricted to a zone of $o(\sqrt{T})$. Ibragimov and Linnik report some results on very large deviations for standardized sums showing how the limiting tails can be represented as the sum of a rational function in x and the normal integral. This type of expansion seems likely to be useful only on the extreme tail of a distribution. It is interesting to note that Sargan and Mikhail (1971) derive from the integral defining the exact distribution of an instrumental variables estimate in a simultaneous equations model an expansion in powers of $1/x$ which is closely related to this.

37. The first two coefficients in (17) are: $\psi_0 = k_3/6\sigma^3$ and $\psi_1 = (\sigma^2 k_4 - 3k_3^2)/24\sigma^6$ where k_j represents the jth cumulant of X_t (see, for instance, Feller (1970, p. 553)).

38. See, for example, the numerical computations in Phillips (1978a).

39. This theory then applies to the random vectors $\sqrt{T}q$ discussed in Section 3.

40. In Phillips (1977a) it was discovered that the Edgeworth approximation becomes unreliable, particularly in the tail areas, as the autoregressive coefficient approaches unity.

41. We have also assumed that there is no constant term in the consumption function. This is justified by the hypothesis that in steady state growth the ratio C/Y is approximately constant.

42. I acknowledge with thanks the help of Ralph Bailey and, in the early stages, Michael Prior in writing this programme. Advice from Clifford Wymer on the principles of programming analytic derivatives is also appreciated.

43. Note that the Edgeworth approximation is derived from an expansion of the distribution of a suitably normalized statistic. In the case of the parameter α we are then working with the statistic $\sqrt{T}(\hat{\alpha} - \alpha)$ for the 2SLS estimator; and it is easy to show that α is the value assumed by the right hand side of (28) when $q_i = 0$ $(i = 1, \ldots, 5)$. Thus, when the sample moments take on their expected values, there is no error in the estimator $\hat{\alpha}$. A minor complication occurs in the case of the OLS estimators, for which the same result does not hold. In this case we standardize by considering $\sqrt{T}(\alpha^* - \bar{\alpha})$ where $\bar{\alpha}$ is the value taken by the right side of (26) when the q_i are all zero. This is logical, because $\sqrt{T}(\alpha^* - \bar{\alpha})$ has the same limiting distributions as $\sqrt{T}(\alpha^* - \text{plim } \alpha^*)$. To see this, we need only observe the plim α^* is the limit of $\bar{\alpha}$ as $T \to \infty$ and $\bar{\alpha}$ differs from its limit by a quantity of $O(T^{-1})$.

44. This can be derived after some manipulation from the result given by Lukacs and Laha (1964, p. 55).

45. An Edgeworth approximation to the distribution of a t-ratio statistic in a first order autoregression is derived in Phillips (1977a).

46. It is easy to derive (A.3) from the general formula (30) I have given in the paper: An explicit derivation for this special case is given by Sawa (1972).

47. In the case of k-class estimators with $0 < k < 1$ we have two singularities in the left half plane and the density function is then just the sum of the residues of the integrand at these points.

48. See, for instance, Miller (1960, p. 161).

REFERENCES

ANDERSON, T. W. (1971) *The Statistical Analysis of Time Series* (New York: Wiley).

ANDERSON, T. W. (1974), "An Asymptotic Expansion of the Distribution of the Limited Information Maximum Likelihood Estimate of a Coefficient in a Simultaneous Equation System", *Journal of the American Statistical Association*, **69**, 565–573.

ANDERSON, T. W. and SAWA, T. (1973), "Distribution of Estimates of Coefficients of a Single Equation in a Simultaneous System and their Asymptotic Expansions", *Econometrica*, **41**, 683–714.

ANDERSON, T. W. and SAWA, T. (1977), "Numerical Evaluation of the Exact and Approximate Distribution Functions of the Two Stage Least Squares Estimate" (Stanford Economic Series Technical Report, No. 239).

BASMANN, R. L. (1961), "A Note on the Exact Finite Sample Frequency Functions of Generalized Classical Linear Estimators in Two Leading Overidentified Cases", *Journal of the American Statistical Association*, **56**, 619–636.

BASMANN, R. L. (1963*b*), "A Note on the exact Finite Sample Frequency Functions of Generalized Classical Linear Estimators in a Leading Three Equation Case", *Journal of the American Statistical Association*, **58**, 161–171.

BASMANN, R. L. (1963*b*), "Remarks Concerning the Application of Exact Sample Distribution Functions of GCL Estimators in Econometric Statistical Inference", *Journal of the American Statistical Association*, **58**, 943–976.

BASMANN, R. L. (1974), "Exact Finite Sample Distributions and Test Statistics: A Survey and Appraisal", chapter 4 in Intriligator, M. D. and Kendrick, D. A. (ed.) *Frontiers of Quantitative Economics*, Vol. II (Amsterdam: North-Holland).

BASMANN, R. L., BROWN, F. L., DAWES, W. S. and SCHOEPFLE, G. K. (1972), "Exact Finite Sample Density Functions of GCL Estimators of Structural Coefficients in a Leading Exactly Identifiable Case", *Journal of the American Statistical Association*, **67**, 252.

BERGSTROM, A. R. (1962), "The Exact Sampling Distributions at Least Squares and Maximum Likelihood Estimators of the Marginal Propensity to Consume", *Econometrica*, **30**, 480–490.

BERRY, A. C. (1941), "The Accuracy of the Gaussian Approximation to the Sum of Independent Variates", *Transactions of the American Mathematical Society*, **48**, 122–136.

BHATTACHARYA, R. N. (1975), "On Errors of Normal Approximation", *Annals of Probability*, **3**, 815–828.

BHATTACHARYA, R. N. (1977), "Refinements of the Multidimensional Central Limit Theorem and Applications", *Annals of Probability*, **5**, 1–27.

BHATTACHARYA, R. N. and GHOSH, J. K. (1978), "On the Validity of the Formal Edgeworth Expansion", *Annals of Statistics*, **6**, 434–451.

BHATTACHARYA, R. N. and RAO, R. R. (1976) *Normal Approximation and Asymptotic Expansions* (New York: Wiley).

BICKEL, P. J. (1974), "Edgeworth Expansions in Nonparametric Statistics", *Annals of Statistics*, **2**, 1–20.

BOWLEY, A. L. (1928) *F. Y. Edgeworth's Contributions to Mathematical Statistics* (London: Royal Statistical Society).

CARLSON, B. C. (1972), "Intégrandes à Deux Formes Quadratiques", *Comptes Rendu Academie Science Paris*, **274**, 1458–1461.

CHAMBERS, J. M. (1967), "On Methods of Asymptotic Approximation for Multivariate Distributions", *Biometrika*, **54**, 367–383.

CHARLIER, C. V. L. (1905), "Über das Fehlergesetz", *Ark. Mat. Astr. Fiz.*, **2**, No. 8.

CHERNOFF, H. (1956), "Large Sample Theory: Parametric Case", *Annals of Mathematical Statistics*, **27**, 1–22.

CHIBISOV, D. M. (1972), "An Asymptotic Expansion for the Distribution of a Statistic Admitting an Asymptotic Expansion", *Theory of Probability and Applications*, **17**, 620–630.

CHIBISOV, D. M. (1973), "An Asymptotic Expansion for a Class of Estimators containing Maximum Likelihood Estimators", *Theory of Probability and Its Applications*, **18**, 295–303.

CONSTANTINE, A. G. (1963), "Some Noncentral Distribution Problems in Multivariate Analysis", *Annals of Mathematical Statistics*, **34**, 1270–1285.

COPSON, E. T. (1967) *Asymptotic Expansions* (Cambridge: Cambridge University Press).

CRAMÉR, H. (1925), "On Some Classes of Series Used in Mathematical Statistics", *Trans. 6th Congr. Scand. Math.*, 399–425.

CRAMÉR, H. (1928), "On the Composition of Elementary Errors", *Skand. Aktnarietidskrift*, **11**, 13–74 and 141–180.

CRAMÉR, H. (1937) *Random Variables and Probability Distributions* (Cambridge University Press).

CRAMÉR, H. (1938), "Sur un nouveau théorème-limite de la Théorie des Probabilités", *Actualites Sci. et Ind.*, **736**, 5–23.

CRAMÉR, H. (1946) *Mathematical Methods of Statistics* (Princeton University Press).

CRAMÉR, H. (1972), "On the History of Certain Expansions used in Mathematical Statistics", *Biometrika*, **59**, 205–207.

CRAMÉR, H. (1976), "Half a Century with Probability Theory: Some Personal Recollections", *Annals of Probability*, **4**, 509–546.

DE BRUIJN, N. G. (1958) *Asymptotic Methods in Analysis* (Amsterdam: North Holland).

DIANANDA, P. H. (1954), "The Central Limit Theorem for *m*-dependent Variables Asymptotically Stationary to Second Order", *Proceedings of the Cambridge Philosophical Society*, **50**, 287–292.

EDGEWORTH, F. Y. (1905), "The Law of Error", *Transactions of the Cambridge Philosophical Society*, **20**, 36–65 and 113–141.

ERDÉYLI, A. (1953) *Higher Transcendental Functions* Vol. 1 (New York: McGraw-Hill).

ESSEEN, C. G. (1945), "Fourier Analysis of Distribution Functions", *Acta Mathematica*, **77**, 1–125.

EXTON, H. (1976) *Multiple Hypergeometric Functions and Applications* (New York: Wiley).

FELLER, W. (1970) *An Introduction to Probability Theory and Its Applications* Vol. II (Wiley: New York).

GRAM, J. P. (1879), "Om Raekkeudviklinger bestempte ved Hjaelp av de mindste Kvadraters Methode" (Kobenhavn).

GUSEV, S. I. (1976), "Asymptotic Expansions Associated with Some Statistical Estimators in the Smooth Case II Expansions of Moments and Distributions", *Theory of Probability and Applications*, **21**, 14–33.

HALE, C., MARIANO, R. S. and RAMAGE, J. (1978), "Finite Sample Analysis of Misspecification in Simultaneous Equations Models" (Pennsylvania Discussion Paper, No. 357).

HANNAN, E. J. (1970) *Multiple Time Series* (New York: Wiley).

HATANAKA, M. (1973), "On the Existence and Approximation Formulae for the Moments of the *k*-class Estimators", *Economic Studies Quarterly*, **24**, 1–15.

HENDRY, D. F. (1973), "On Asymptotic Theory and Finite Sample Experiments", *Econometrica*, **160**, 210–217.

HENDRY, D. F. (1976), "On the Behaviour of Inconsistent Instrumental Variables Estimators in Dynamic Systems with Autocorrelated Errors" (Cowles Foundation Discussion Paper).

HENDRY, D. F. and HARRISON, R. W. (1974), "Monte Carlo Methodology and the Finite Sample Behaviour of Ordinary and Two-Stage Least Squares", *Journal of Econometrics*, **2**, 151–174.

HERZ, C. S. (1955), "Bessel Functions of Matrix Argument", *Annals of Mathematics*, **61**, 474–523.

HOLLY, A. and PHILLIPS, P. C. B. (1977), "A Saddlepoint Approximation to the Distribution of the k-Class Estimator of a Coefficient in a Simultaneous System" (École Polytechnique Discussion Paper No. A170 1277) forthcoming in *Econometrica*, 1979.

IBRAGIMOV, I. A. (1967), "On the Chebyshev-Cramér Asymptotic Expansions", *Theory of Probability and its Applications*, **12**, 455–469.

IBRAGIMOV, I. A. and LINNIK, V. V. (1971) *Independent and Stationary Sequences of Random Variables* (Groningen: Wolter-Noordhoff).

IVANOV, A. V. (1976), "An Asymptotic Expansion for the Distribution of the Least Squares Estimator of the Nonlinear Regression Parameter", *Theory of Probability and Its Applications*, **21**, 557–570.

JAMES, A. T. (1964), "Distributions of Matrix Variates and Latent Roots Derived from Normal Samples", *Annals of Mathematical Statistics*, **35**, 475–450.

JAMES, A. T. (1975), "Special Functions of Matrix and Single Argument in Statistics", in *Theory and Application of Special Functions* (Madison).

JOHNSON, N. L. and KOTZ, S. (1972) *Distributions in Statistics: Continuous Multivariate Distributions* (New York: Wiley).

KLEIN, L. R. (1969), "Estimation of Interdependent Systems in Macroeconometrics", *Econometrica*, **37**, 171–192.

KMENTA, J. (1974), "Comments on R. L. Basmann (1974)", pp. 272–275 in Intriligator, M. D. and Kendrick, D. A. (ed.) *Frontiers of Quantitative Economics*, Vol. II (Amsterdam: North Holland).

KUMAR, A. (1973), "Expectation of Product of Quadratic Forms", *Sankhya Series B*, **35**, 359–362.

LEBEDEV, N. N. (1965) *Special Functions and Their Application* (Englewood Cliffs, N.J.: Prentice Hall).

LUKACS, E. and LAHA, R. G. (1964) *Applications of Characteristic Functions* (London: Griffin and Co).

LUKACS, E. (1970) *Characteristic Functions* (Second Edition) (London: Griffin).

MCDONALD, J. B. (1972), "The Exact Finite Sample Distribution Function of the Limited Information Maximum Likelihood Identifiability Test Statistic", *Econometrica*, **40**, 1109–1119.

MAGNUS, J. R. (1978), "The Moments of Products of Quadratic Forms in Normal Variables", *Statistica Neerlandica*, **22**, 69–82.

MARIANO, R. S. (1972), "The Existence of Moments of the Ordinary Least Squares and Two Stage Least Squares Estimators", *Econometrica*, **40**, 643–652.

MARIANO, R. S. (1973), "Approximations to the Distribution Functions of the Ordinary Least-Squares and Two-Stage Estimators in the Case of Two Included Endogeneous Variables", *Econometrica*, **41**, 67–77.

MARIANO, R. S. (1977), "Finite Sample Properties of Instrumental Variable Estimators of Structural Coefficients", *Econometrica*, **45**, 487–495.

MARIANO, R. S. and SAWA, T. (1972), "Exact Finite Sample Distribution of the Limited Information Maximum Likelihood Estimator in the Case of Two Included Endogenous Variables", *Journal of the American Statistical Association*, **67**, 159–163.

MICHEL, R. (1975), "An Asymptotic Expansion for the Distribution of Asymptotic Maximum Likelihood Estimators of Vector Parameters", *Journal of Multivariate Analysis*, **5**, 67–82.

MILLER, K. S. (1960) *Advanced Complex Calculus* (New York: Harper).

MUIRHEAD, R. J. (1975), "Expressions for Some Hypergeometric Functions of Matrix Argument with Applications", *Journal of Multivariate Analysis*, **5**, 283–293.

NAGAR, A. L. (1959), "The Bias and Moment Matrix of the General k-Class Estimators of the Parameters in Structural Equations", *Econometrica*, **27**, 575–595.

NEUDECKER, H. (1968), "The Kronecker Matrix Product and Some of its Applications in Econometrics", *Statistica Neerlandica*, **22**, 69–82.

ORCUTT, G. H. and WINOKUR, H. S. (1969), "First Order Autoregression: Inference Estimation and Prediction", *Econometrica*, **37**, 1–14.

PETROV, V. V. (1964), "On Local Limit Theorems for Sums of Independent Random Variables", *Theory of Probability and its Applications*, **9**, 312–320.

PETROV, V. V. (1968), "Asymptotic Behaviour of Probabilities of Large Deviations", *Theory of Probability and its Applications*, **13**, 408–420.

PETROV, V. V. (1975) *Sums of Independent Random Variables* (Berlin: Springer-Verlag).

PFANZAGL, J. (1973), "Asymptotic Expansions related to Minimum Contrast Estimators", *Annals of Statistics*, **1**, 993–1026.

PHILLIPS, P. C. B. (1976), "Large Deviation Expansions Applied to Tail Probabilities in a First Order Stochastic Difference Equation" (Birmingham Discussion Paper (A), No. 203).

PHILLIPS, P. C. B. (1977*a*), "Approximations to Some Finite Sample Distributions Associated with a First Order Stochastic Difference Equation", *Econometrica*, **45**, 463–486.

PHILLIPS, P. C. B. (1977*b*), "A General Theorem in the Theory of Asymptotic Expansions as Approximations to the Finite Sample Distributions of Econometric Estimators", *Econometrica*, **45**, 1517–1534.

PHILLIPS, P. C. B. (1977*c*), "A Large Deviation Limit Theorem for Multivariate Distributions", *Journal of Multivariate Analysis*, **7**, 50–62.

PHILLIPS, P. C. B. (1978*a*), "Edgeworth and Saddlepoint Approximations in the First-Order Non Circular Autoregression", *Biometrika*, **65**, 91–98.

PHILLIPS, P. C. B. (1978*b*), "The Exact Finite Sample Density of Instrumental Variable Estimators in an Equation Containing $n+1$ Endogenous Variables", forthcoming in *Econometrica*, 1980.

PHILLIPS, P. C. B. and WICKENS, M. R. (1978) *Exercises in Econometrics* (Ballinger and Philip Allan).

RHODES, G. F. (1977), "The Exact Density Functions for Identifiability Test Statistics: The G_Δ Endogenous Variable Case" (mimeographed, Ohio State University).

RHODES, G. F. and WESTBROOK, M. D. (1977), "An Analysis and Comparison of OLS and 2SLS Exact Density Functions Under Misspecification: The Two Included Endogenous Variable Case" (Ohio Working Paper No. 27).

RICHARDSON, D. H. and WU, DE MIN (1971), "A Note on the Comparison of Ordinary and Two Stage Least Squares Estimators", *Econometrica*, **39**, 973–981.

RICHARDSON, D. H. (1968), "The Exact Distribution of a Structural Coefficient Estimator", *Journal of the American Statistical Association*, **63**, 1214–1226.

RICHTER, W. (1957), "Local Limit Theorems for Large Deviations", *Theory of Probability and its Application*, **2**, 206–219.

RICHTER, W. (1958), "Multi-dimensional Local Limit Theorems for Large Deviations", *Theory of Probability and its Application*, **3**, 100–106.

ROZANOV, Y. A. (1967) *Stationary Random Processes* (San Francisco: Holden-Day).

SARGAN, J. D. (1964), "An Approximate Distribution of the Two Stage Least Squares Estimators", Abstract in *Econometrica*, **32**, 660.

SARGAN, J. D. (1970), "The Finite Sample Distribution of FIML Estimators" (Paper presented to The World Congress of the Econometric Society, Cambridge).

SARGAN, J. D. (1974), "The Validity of Nagar's Expansion for the Moments of Econometric Estimators", *Econometrica*, **42**, 169–176.

SARGAN, J. D. (1975), "Gram-Charlier Approximations Applied to *t*-ratios of *k*-class Estimators", *Econometrica*, **43**, 326–346.

SARGAN, J. D. (1976*a*), "Econometric Estimators and the Edgeworth Approximation", *Econometrica*, **44**, 421–428; and (1977) "Erratum", *Econometrica*, **15**, 272.

SARGAN, J. D. (1967*b*), "Some Approximations to the Distribution of Econometric Criteria which are Asymptotically Distributed as Chi-Squared" (mimeographed, London School of Economics).

SARGAN, J. D. and MIKHAIL, W. M. (1971), "A General Approximation to the Distribution of Instrumental Variable Estimates", *Econometrica*, **39**, 131–169.

SAWA, T. (1969), "The Exact Finite Sampling Distribution of Ordinary Least Squares and Two Stage Least Squares Estimator", *Journal of the American Statistical Association*, **64**, 923–936.

SAWA, T. (1972), "Finite Sample Properties of the *k*-Class Estimators", *Econometrica*, **40**, 653–680.

SAWA, T. (1973*a*), "Almost Unbiased Estimator in Simultaneous Equations Systems", *International Economic Review*, **14**, 97–106.

SAWA, T. (1973*b*), "The Mean Square Error of a Combined Estimator and Numerical Comparison with the TSLS Estimator", *Journal of Econometrics*, **1**, 115–132.

SAZONOV, V. V. (1968), "On the Speed of Convergence of the Multidimensional Central Limit Theorem", *Theory of Probability and its Applications*, **13**, 188–191.

SCHÖNFELD, P. (1971), "A Useful Central Limit Theorem for *m*-Dependent Variables", *Metrika*, **15**, 116–128.

SEAL, H. L. (1967), "The Historical Development of the Gauss Linear Model", *Biometrika*, **54**, 1–24.

SLATER, L. J. (1960) *Confluent Hypergeometric Functions* (Cambridge: Cambridge University Press).

STIGLER, S. M. (1978), "Francis Ysidro Edgeworth, Statistician" (University of Wisconsin Technical Report No. 510) and *Journal of the Royal Statistical Society* (forthcoming).

SUBRAHMANIAM, K. (1966), "Some Contributions to the Theory of Non-Normality-I (Univariate Case)", *Sankhya Series A*, **28**, 389–406.

SUBRAHMANIAM, K. (1968), "Some Contributions to the Theory of Non-Normality–II", *Sankhya, Series A*, **30**, 411–432.

SUBRAHMANIAM, K. (1976), "Recent Trends in Multivariate Normal Distribution Theory: on the Zonal Polynomials and Other Functions of Matrix Argument", *Sankhya, Series A*, **38**, 221–258.

SUGIYAMA, T. (1972), "Percentile Points of the Largest Latent Root of a Matrix and Power Calculations for Testing Hypothesis $\Sigma = I$", *Journal of Japanese Statistical Society*, **3**, 1–8.

SWEETING, T. J. (1977), "Speeds of Convergence for the Multidimensional Central Limit Theorem", *Annals of Probability*, **5**, 28–41.

TAKEUCHI, K. (1970), "Exact Sampling Moments of the Ordinary Least Squares, Instrumental Variable and Two-Stage Least Squares Estimators", *International Economic Review*, **11**, 1–12.

TCHEBYCHEFF, P. L. (1860), "Sur le Développement des Functions à une Seule Variable", *Bull. de l'Acad. Imp des Sci de St. Petersbourg*, (3), **1**, 193–200.

TCHEBYCHEFF, P. L. (1890), "Sur Deux Théorèmes Relatifs aux Probabilités", *Acta Math.*, **14**, 305–315.

ULLAH, A. and NAGAR, A. L. (1974), "The Exact Mean of the Two Stage Least Squares Estimator of the Structural Parameters in an Equation containing three Endogenous Variables", *Econometrica*, **42**, 749–758.

WALLACE, D. L. (1958), "Asymptotic Approximations to Distributions", *Annals of Mathematical Statistics*, **29**, 635–654.