Chapter 8

EXACT SMALL SAMPLE THEORY
IN THE SIMULTANEOUS EQUATIONS MODEL

P C. B PHILLIPS*

Yale University

Contents

Little experience is sufficient to show that the traditional machinery of statistical processes is wholly unsuited to the needs of practical research Not only does it take a cannon to shoot a sparrow, but it misses the sparrow[1] The elaborate mechanism built on the theory of infinitely large samples is not accurate enough for simple laboratory data Only by systematically tackling small sample problems on their merits does it seem possible to apply accurate tests to practical data Such at least has been the aim of this book [From the Preface to the First Edition of R A Fisher (1925).]

1. Introduction

Statistical procedures of estimation and inference are most frequently justified in econometric work on the basis of certain desirable asymptotic properties. One estimation procedure may, for example, be selected over another because it is known to provide consistent and asymptotically efficient parameter estimates under certain stochastic environments. Or, a statistical test may be preferred because it is known to be asymptotically most powerful for certain local alternative hypotheses.[1] Empirical investigators have, in particular, relied heavily on asymptotic theory to guide their choice of estimator, provide standard errors of their estimates and construct critical regions for their statistical tests. Such a heavy reliance on asymptotic theory can and does lead to serious problems of bias and low levels of inferential accuracy when sample sizes are small and asymptotic formulae poorly represent sampling behavior. This has been acknowledged in mathematical statistics since the seminal work of R. A. Fisher,[2] who recognized very early the limitations of asymptotic machinery, as the above quotation attests, and who provided the first systematic study of the exact small sample distributions of important and commonly used statistics.

The first step towards a small sample distribution theory in econometrics was taken during the 1960s with the derivation of exact density functions for the two stage least squares (2SLS) and ordinary least squares (OLS) estimators in simple simultaneous equations models (SEMs). Without doubt, the mainspring for this research was the pioneering work of Basmann (1961), Bergstrom (1962), and Kabe (1963, 1964). In turn, their work reflected earlier influential investigations in econometrics: by Haavelmo (1947) who constructed exact confidence regions for structural parameter estimates from corresponding results on OLS reduced form coefficient estimates; and by the Cowles Commission researchers, notably Anderson and Rubin (1949), who also constructed confidence regions for structural coefficients based on a small sample theory, and Hurwicz (1950) who effectively studied and illustrated the small sample bias of the OLS estimator in a first order autoregression.

---

[1] The nature of local alternative hypotheses is discussed in Chapter 13 of this Handbook by Engle

[2] See, for example, Fisher (1921, 1922, 1924, 1928a, 1928b, 1935) and the treatment of exact sampling distributions by Cramer (1946)

The mission of these early researchers is not significantly different from our own today: ultimately to relieve the empirical worker from the reliance he has otherwise to place on asymptotic theory in estimation and inference. Ideally, we would like to know and be able to compute the exact sampling distributions relevant to our statistical procedures under a variety of stochastic environments. Such knowledge would enable us to make a better assessment of the relative merits of competing estimators and to appropriately correct (from their asymptotic values) the size or critical region of statistical tests. We would also be able to measure the effect on these sampling distributions of certain departures in the underlying stochastic environment from normally distributed errors. The early researchers clearly recognized these goals, although the specialized nature of their results created an impression[3] that there would be no substantial payoff to their research in terms of applied econometric practice. However, their findings have recently given way to general theories and a powerful technical machinery which will make it easier to transmit results and methods to the applied econometrician in the precise setting of the model and the data set with which he is working. Moreover, improvements in computing now make it feasible to incorporate into existing regression software subroutines which will provide the essential vehicle for this transmission. Two parallel current developments in the subject are an integral part of this process. The first of these is concerned with the derivation of direct approximations to the sampling distributions of interest in an applied study. These approximations can then be utilized in the decisions that have to be made by an investigator concerning, for instance, the choice of an estimator or the specification of a critical region in a statistical test. The second relevant development involves advancements in the mathematical task of extracting the form of exact sampling distributions in econometrics. In the context of simultaneous equations, the literature published during the 1960s and 1970s concentrated heavily on the sampling distributions of estimators and test statistics in single structural equations involving only two or at most three endogenous variables. Recent theoretical work has now extended this to the general single equation case.

The aim of the present chapter is to acquaint the reader with the main strands of thought in the literature leading up to these recent advancements. Our discussion will attempt to foster an awareness of the methods that have been used or that are currently being developed to solve problems in distribution theory, and we will consider their suitability and scope in transmitting results to empirical researchers. In the exposition we will endeavor to make the material accessible to readers with a working knowledge of econometrics at the level of the leading textbooks. A cursory look through the journal literature in this area may give the impression that the range of mathematical techniques employed is quite diverse, with the method and final form of the solution to one problem being very different from the next. This diversity is often more apparent than real and it is

---

[3]The discussions of the review article by Basmann (1974) in Intriligator and Kendrick (1974) illustrate this impression in a striking way. The achievements in the field are applauded, but the reader is left with a skeptical view of the usefulness of the results

hoped that the approach we take to the subject in the present review will make the methods more coherent and the form of the solutions easier to relate.

Our review will not be fully comprehensive in coverage but will report the principal findings of the various research schools in the area. Additionally, our focus will be directed explicitly towards the SEM and we will emphasize exact distribution theory in this context. Corresponding results from asymptotic theory are surveyed in Chapter 7 of this Handbook by Hausman; and the refinements of asymptotic theory that are provided by Edgeworth expansions together with their application to the statistical analysis of second-order efficiency are reviewed in Chapter 15 of this Handbook by Rothenberg. In addition, and largely in parallel to the analytical research that we will review, are the experimental investigations involving Monte Carlo methods. These latter investigations have continued traditions established in the 1950s and 1960s with an attempt to improve certain features of the design and efficiency of the experiments, together with the means by which the results of the experiments are characterized. These methods are described in Chapter 16 of this Handbook by Hendry. An alternative approach to the utilization of soft quantitative information of the Monte Carlo variety is based on constructive functional approximants of the relevant sampling distributions themselves and will be discussed in Section 4 of this chapter.

The plan of the chapter is as follows. Section 2 provides a general framework for the distribution problem and details formulae that are frequently useful in the derivation of sampling distributions and moments. This section also provides a brief account of the genesis of the Edgeworth, Nagar, and saddlepoint approximations, all of which have recently attracted substantial attention in the literature. In addition, we discuss the Wishart distribution and some related issues which are central to modern multivariate analysis and on which much of the current development of exact small sample theory depends. Section 3 deals with the exact theory of single equation estimators, commencing with a general discussion of the standardizing transformations, which provide research economy in the derivation of exact distribution theory in this context and which simplify the presentation of final results without loss of generality. This section then provides an analysis of known distributional results for the most common estimators, starting with certain leading cases and working up to the most general cases for which results are available. We also cover what is presently known about the exact small sample behavior of structural variance estimators, test statistics, systems methods, reduced-form coefficient estimators, and estimation under misspecification. Section 4 outlines the essential features of a new approach to small sample theory that seems promising for future research. The concluding remarks are given in Section 5 and include some reflections on the limitations of traditional asymptotic methods in econometric modeling.

Finally, we should remark that our treatment of the material in this chapter is necessarily of a summary nature, as dictated by practical requirements of space. A more complete exposition of the research in this area and its attendant algebraic detail is given in Phillips (1982e). This longer work will be referenced for a fuller discussion of the material whenever it is appropriate in the present chapter.

2. Simple mechanics of distribution theory

2.1. Primitive exact relations and useful inversion formulae

To set up a general framework we assume a model which uniquely determines the joint probability distribution of a vector of n endogenous variables at each point in time ($t = 1, \ldots, T$), namely $\{y_1, \ldots, y_T\}$, conditional on certain fixed exogenous variables $\{x_1, \ldots, x_T\}$ and possibly on certain initial values $\{y_{-k}, \ldots, y_0\}$. This distribution can be completely represented by its distribution function (d.f.), $\mathrm{df}(y|x, y_{-1}; \theta)$ or its probability density function (p.d.f.), $\mathrm{pdf}(y|x, y_{-} ; \theta)$, both of which depend on an unknown vector of parameters θ and where we have set $y' = (y_1', \ldots, y_T')$, $x' = (x_1', \ldots, x_T')$, and $y_-' = (y_{-k}', \ldots, y_0')$. In the models we will be discussing in this chapter the relevant distributions will not be conditional on initial values, and we will suppress the vector y_- in these representations. However, in other contexts, especially certain time-series models, it may become necessary to revert to the more general conditional representation. We will also frequently suppress the conditioning x and parameter θ in the representation $\mathrm{pdf}(y|x; \theta)$, when the meaning is clear from the context. Estimation of θ or a subvector of θ or the use of a test statistic based on an estimator of θ leads in all cases to a function of the available data. Therefore we write in general $\theta_T = \theta_T(y, x)$. This function will determine the numerical value of the estimate or test statistic.

The small sample distribution problem with which we are faced is to find the distribution of θ_T from our knowledge of the distribution of the endogenous variables and the form of the function which defines θ_T. We can write down directly a general expression for the distribution function of θ_T as

$$\mathrm{df}(r) = \mathrm{P}(\theta_T \leqslant r) = \int_{y \in \Theta(r)} \mathrm{pdf}(y)\, dy,$$

$$\Theta(r) = \{y: \theta_T(y, x) \leqslant r\}. \tag{2.1}$$

This is an nT-dimensional integral over the domain of values $\Theta(r)$ for which $\theta_T \leqslant r$.

The distribution of θ_T is also uniquely determined by its characteristic function (c.f.), which we write as

$$\mathrm{cf}(s) = E(e^{\mathrm{i}s\theta_T}) = \int e^{\mathrm{i}s\theta_T(y, x)} \mathrm{pdf}(y)\, dy, \tag{2.2}$$

where the integration is now over the entire y-space. By inversion, the p.d.f. of θ_T is given by

$$\mathrm{pdf}(r) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-\mathrm{i}sr} \mathrm{cf}(s)\, ds, \tag{2.3}$$

and this inversion formula is valid provided cf(s) is absolutely integrable in the Lebesgue sense [see, for example, Feller (1971, p. 509)]. The following two inversion formulae give the d.f. of θ_T directly from (2.2):

$$\mathrm{df}(r) - \mathrm{df}(0) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{1 - e^{-\iota s r}}{\iota s} \, \mathrm{cf}(s) \, \mathrm{d}s \tag{2.4}$$

and

$$\mathrm{df}(r) = \frac{1}{2} + \frac{1}{2\pi} \int_{0}^{\infty} \frac{e^{\iota s r} \mathrm{cf}(-s) - e^{-\iota s r} \mathrm{cf}(s)}{\iota s} \, \mathrm{d}s. \tag{2.5}$$

The first of these formulae is valid whenever the integrand on the right-hand side of (2.4) is integrable [otherwise a symmetric limit is taken in defining the improper integral – see, for example, Cramér (1946, pp. 93–94)]. It is useful in computing first differences in df(r) or the proportion of the distribution that lies in an interval (a, b) because, by subtraction, we have

$$\mathrm{df}(b) - \mathrm{df}(a) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{e^{-\iota s a} - e^{-\iota s b}}{\iota s} \, \mathrm{cf}(s) \, \mathrm{d}s. \tag{2.6}$$

The second formula (2.5) gives the d.f. directly and was established by Gil-Pelaez (1951).

When the above inversion formulae based on the characteristic function cannot be completed analytically, the integrals may be evaluated by numerical integration. For this purpose, the Gil-Pelaez formula (2.5) or variants thereof have most frequently been used. A general discussion of the problem, which provides bounds on the integration and truncation errors, is given by Davies (1973). Methods which are directly applicable in the case of ratios of quadratic forms are given by Imhof (1961) and Pan Jie Jian (1968). The methods provided in the latter two articles have often been used in econometric studies to compute exact probabilities in cases such as the serial correlation coefficient [see, for example. Phillips (1977a)] and the Durbin–Watson statistic [see Durbin and Watson (1971)].

2.2. Approach via sample moments of the data

Most econometric estimators and test statistics we work with are relatively simple functions of the sample moments of the data (y, x). Frequently, these functions are rational functions of the first and second sample moments of the data. More specifically, these moments are usually well-defined linear combinations and matrix quadratic forms in the observations of the endogenous variables and with

the weights being determined by the exogenous series. Inspection of the relevant formulae makes this clear: for example, the usual two-step estimators in the linear model and the instrumental variable (IV) family in the SEM. In the case of limited information and full information maximum likelihood (LIML, FIML), these estimators are determined as implicit functions of the sample moments of the data through a system of implicit equations. In all of these cases, we can proceed to write $\theta_T = \theta_T(y, x)$ in the alternative form $\theta_T = \theta_T^*(m)$, where m is a vector of the relevant sample moments.

In many econometric problems we can write down directly the p.d.f. of the sample moments, i.e. pdf(m), using established results from multivariate distribution theory. This permits a convenient resolution of the distribution of θ_T. In particular, we achieve a useful reduction in the dimension of the integration involved in the primitive forms (2.1) and (2.2). Thus, the analytic integration required in the representation

$$\text{pdf}(m) = \int_{a \in \mathcal{Q}} \text{pdf}(y) \left| \frac{\partial y}{\partial(m, a)} \right| \, da \tag{2.7}$$

has already been reduced. In (2.7) a is a vector of auxiliary variates defined over the space $\mathcal{Q}$ and is such that the transformation $y \rightarrow (m, a)$ is $1:1$.

The next step in reducing the distribution to the density of θ_T is to select a suitable additional set of auxiliary variates b for which the transformation $m \rightarrow (\theta_T, b)$ is $1:1$. Upon changing variates, the density of θ_T is given by the integral

$$\text{pdf}(r) = \int_{b \in \mathcal{B}} \text{pdf}(m) \left| \frac{\partial m}{\partial(r, b)} \right| \, db, \tag{2.8}$$

where $\mathcal{B}$ is the space of definition of b. The simplicity of the representation (2.8) often belies the major analytic difficulties that are involved in the practical execution of this step.[4] These difficulties center on the selection of a suitable set of auxiliary variates b for which the integration in (2.8) can be performed analytically. In part, this process depends on the convenience of the space, $\mathcal{B}$, over which the variates b are to be integrated, and whether or not the final integral has a recognizable form in terms of presently known functions or infinite series.

All of the presently known exact small sample distributions of single equation estimators in the SEM can be obtained by following the above steps. When reduced, the final integral (2.8) is most frequently expressed in terms of infinite

---

[4]See, for example, Sargan (1976a, Appendix B) and Phillips (1980a) These issues will be taken up further in Section 3.5

series involving some of the special functions of applied mathematics, which themselves admit series representations. These special functions are often referred to as higher transcendental functions. An excellent introduction to them is provided in the books by Whittaker and Watson (1927), Rainville (1963), and Lebedev (1972); and a comprehensive treatment is contained in the three volumes by Erdéyli (1953). At least in the simpler cases, these series representations can be used for numerical computations of the densities.

2.3. Asymptotic expansions and approximations

An alternative to searching for an exact mathematical solution to the problem of integration in (2.8) is to take the density pdf(m) of the sample moments as a starting point in the derivation of a suitable approximation to the distribution of θ_T. Two of the most popular methods in current use are the Edgeworth and saddlepoint approximations. For a full account of the genesis of these methods and the constructive algebra leading to their respective asymptotic expansions, the reader may refer to Phillips (1982e). For our present purpose, the following intuitive ideas may help to briefly explain the principles that underlie these methods.

Let us suppose, for the sake of convenience, that the vector of sample moments m is already appropriately centered about its mean value or limit in probability. Let us also assume that $\sqrt{T}m \xrightarrow{\mathcal{D}} N(0, \mathcal{V})$ as $T \to \infty$, where $\xrightarrow{\mathcal{D}}$ denotes "tends in distribution". Then, if $\theta_T = f(m)$ is a continuously differentiable function to the second order, we can readily deduce from a Taylor series representation of $f(m)$ in a neighborhood of $m = 0$ that $\sqrt{T}\{f(m) - f(0)\} \xrightarrow{\mathcal{D}} N(0, \mathcal{W})$, where $\mathcal{W} = (\partial f(0)/\partial m')\mathcal{V}\partial f'(0)/\partial m$. In this example, the asymptotic behavior of the statistic $\sqrt{T}\{f(m) - f(0)\}$ is determined by that of the linear function $\sqrt{T}(\partial f(0)/\partial m')m$ of the basic sample moments. Of course, as $T \to \infty$, $m \to 0$ in probability, so that the behavior of $f(m)$ in the immediate locality of $m = 0$ becomes increasingly important in influencing the distribution of this statistic as T becomes large.

The simple idea that underlies the principle of the Edgeworth approximation is to bridge the gap between the small sample distribution (with T finite) and the asymptotic distribution by means of correction terms which capture higher order features of the behavior of $f(m)$ in the locality of $m = 0$. We thereby hope to improve the approximation to the sampling distribution of $f(m)$ that is provided by the crude asymptotic. Put another way, the statistic $\sqrt{T}\{f(m) - f(0)\}$ is approximated by a polynomial representation in m of higher order than the linear representation used in deducing the asymptotic result. In this sense, Edgeworth approximations provide refinements of the associated limit theorems which give us the asymptotic distributions of our commonly used statistics. The reader may usefully consult Cramér (1946, 1972), Wallace (1958), Bhattacharya and Rao

(1976), and the review by Phillips (1980b) for further discussion, references, and historical background.

The concept of using a polynomial approximation of θ_T in terms of the elements of m to produce an approximate distribution for θ_T can also be used to approximate the moments of θ_T, where these exist, or to produce pseudo-moments (of an approximating distribution) where they do not.[5] The idea underlies the work by Nagar (1959) in which such approximate moments and pseudo-moments were developed for k-class estimators in the SEM. In popular parlance these moment approximations are called Nagar approximations to the moments. The constructive process by which they are derived in the general case is given in Phillips (1982e).

An alternative approach to the development of asymptotic series approximations for probability densities is the saddlepoint (SP) method. This is a powerful technique for approximating integrals in asymptotic analysis and has long been used in applied mathematics. A highly readable account of the technique and a geometric interpretation of it are given in De Bruijn (1958). The method was first used systematically in mathematical statistics in two pathbreaking papers by Daniels (1954, 1956) and has recently been the subject of considerable renewed interest.[6]

The conventional approach to the SP method has its starting point in inversion formulae for the probability density like those discussed in Section 2.1. The inversion formula can commonly be rewritten as a complex integral and yields the p.d.f. of θ_T from knowledge of the Laplace transform (or moment-generating function). Cauchy's theorem in complex function theory [see, for example, Miller (1960)] tells us that we may well be able to deform the path of integration to a large extent without changing the value of the integral. The general idea behind the SP method is to employ an allowable deformation of the given contour, which is along the imaginary axis, in such a way that the major contribution to the value of the integral comes from the neighborhood of a point at which the contour actually crosses a saddlepoint of the modulus of the integrand (or at least its dominant factor). In crude terms, this is rather akin to a mountaineer attempting to cross a mountain range by means of a pass, in order to control the maximum

---

[5] This process involves a stochastic approximation to the statistic θ_T by means of polynomials in the elements of m which are grouped into terms of like powers of $T^{-1/2}$. The approximating statistic then yields the "moment" approximations for θ_T Similar "moment" approximations are obtained by developing alternative stochastic approximations in terms of another parameter Kadane (1971) derived such alternative approximations by using an expansion of θ_T (in the case of the k-class estimator) in terms of increasing powers of σ, where σ^2 is a scalar multiple of the covariance matrix of the errors in the model and the asymptotics apply as $\sigma \to 0$ Anderson (1977) has recently discussed the relationship between these alternative parameter sequences in the context of the SEM

[6] See. for example, Phillips (1978), Holly and Phillips (1979), Daniels (1980), Durbin (1980a, 1980b), and Barndorff-Nielson and Cox (1979).

altitude he has to climb. This particular physical analogy is developed at some length by De Bruijn (1958).

A new and elegant approach to the extraction of SP approximations has recently been developed by Durbin (1980a). This method applies in cases where we wish to approximate the p.d.f. of a sufficient statistic and has the great advantage that we need only know the p.d.f. of the underlying data pdf(y; θ) and the limiting mean information matrix $\lim_{T \to \infty} E\{-T^{-1}\partial^2 \ln[\text{pdf}(y; \theta)]/\partial\theta\partial\theta'\}$ in order to construct the approximation. This is, in any event, the information we need to extract the maximum likelihood estimator of θ and write down its asymptotic covariance matrix. Durbin's approach is based on two simple but compelling steps. The first is the fundamental factorization relation for sufficient statistics, which yields a powerful representation of the required p.d.f. for a parametric family of densities. The second utilizes the Edgeworth expansion of the required p.d.f. but at a parametric value (of θ) for which this expansion has its best asymptotic accuracy. This parametric recentering of the Edgeworth expansion increases the rate of convergence in the asymptotic series and thereby can be expected to provide greater accuracy at least for large enough T. Algebraic details, further discussion and examples of the method are given in Phillips (1982e).

2.4. The Wishart distribution and related issues

If $X = [x_1, \ldots, x_T]$ is an $n \times T$ matrix variate (i.e. matrix of random variates) whose columns are independent $N(0, \Omega)$ then the $n \times n$ symmetric matrix $A = XX' = \sum_{t=1}^{T} x_t x_t'$ has a Wishart distribution with p.d.f. given by

$$\text{pdf}(A) = \frac{1}{(2)^{nT/2}\Gamma_n\left(\frac{T}{2}\right)(\det \Omega)^{T/2}}\text{etr}\left(-\tfrac{1}{2}\Omega^{-1}A\right)(\det A)^{(T-n-1)/2}. \quad (2.9)$$

Since A is symmetric $n \times n$, this density has $N = \tfrac{1}{2}n(n+1)$ independent arguments and is supported on the subset (a natural cone) of N dimensional Euclidean space for which A is positive definite (which we write as $A > 0$). It is a simple and useful convention to use the matrix A as the argument of the density in (2.9), although in transforming the distribution we must recognize the correct number of independent arguments.

In (2.9) above $\Gamma_n(z)$ is the multivariate gamma function defined by the integral

$$\Gamma_n(z) = \int_{S > 0} \text{etr}(-S)(\det S)^{z-(1/2)(n+1)}dS.$$

This integral is a (matrix variate) Laplace transform [see, for example, Herz

(1955) and Constantine (1963)] which converges absolutely for $\text{Re}(z) > \frac{1}{2}(n-1)$ and the domain of integration is the set of all positive definite matrices. It can be evaluated in terms of univariate gamma functions as

$$\Gamma_n(z) = \pi^{(1/4)n(n-1)} \prod_{i=1}^{n} \Gamma\left(z - \tfrac{1}{2}(i-1)\right)$$

[see James (1964)]. In (2.9) we also use the abbreviated operator representation $\text{etr}(\cdot) = \exp\{\text{tr}(\cdot)\}$.

The parameters of the Wishart distribution (2.9) are: (i) the order of the symmetric matrix A. namely n; (ii) the degrees of freedom T, of the component variates x_t in the summation $A = XX' = \sum_{t=1}^{T} x_t x_t'$; and (iii) the covariance matrix, Ω, of the normally distributed columns x_t in X. A common notation for the Wishart distribution (2.9) is then $\mathfrak{W}_n(T, \Omega)$ [see, for example, Rao (1973, p. 534)]. This distribution is said to be central (in the same sense as the central χ^2 distribution) since the component variates x_t have common mean $E(x_t) = 0$. In fact, when $n = 1$, $\Omega = 1$, and $A = a$ is a scalar, the density (2.9) reduces to $(2)^{-T/2} \Gamma(T/2)^{-1} a^{T/2-1} e^{-(1/2)a}$, the density of a central χ^2 with T degrees of freedom.

If the component variates x_t in the summation are not restricted to have a common mean of zero but are instead independently distributed as $N(m_t, \Omega)$, then the joint distribution of the matrix $A = XX' = \sum_{t=1}^{T} x_t x_t'$ is said to be (non-central) Wishart with non-centrality matrix $\overline{M} = MM'$, where $M = [m_1, \ldots, m_T]$. This is frequently denoted $\mathfrak{W}_n(T, \Omega, \overline{M})$, although M is sometimes used in place of $\overline{M}$ [as in Rao (1973), for example]. The latter is a more appropriate parameter in the matrix case as a convenient generalization of the non-centrality parameter that is used in the case of the non-central χ^2 distribution – a special case of $\mathfrak{W}_n(T, \Omega, \overline{M})$ in which $n = 1$, $\Omega = 1$, and $\overline{M} = \sum_{t=1}^{T} m_t^2$.

The p.d.f. of the non-central Wishart matrix $A = XX' = \sum_{t=1}^{T} x_t x_t'$, where the x_t are independent $N(m_t, \Omega)$, $M = [m_1, \ldots, m_T] = E(X)$, and $\overline{M} = MM'$ is given by

$$\text{pdf}(A) = \frac{\text{etr}\left(-\frac{1}{2}\Omega^{-1}\overline{M}\right)}{2^{nT/2} \Gamma_n\left(\dfrac{T}{2}\right)(\det \Omega)^{T/2}} \, {}_0F_1\left(\frac{T}{2}; \frac{1}{4}\Omega^{-1}\overline{M}\Omega^{-1}A\right)$$

$$\times \text{etr}\left(-\tfrac{1}{2}\Omega^{-1}A\right)(\det A)^{(T-n-1)/2}. \tag{2.10}$$

In (2.10) the function ${}_0F_1(;)$ is a matrix argument hypergeometric function, closely related to the Bessel function of matrix argument discussed by Herz (1955). Herz extended the classical hypergeometric functions of scalar argument [see, for example. Erdéyli (1953)] to matrix argument functions by using multidimensional Laplace transforms and inverse transforms. Constantine (1963) discovered that hypergeometric functions ${}_pF_q$ of a matrix argument have a general

series representation in terms of zonal polynomials as follows:

$$
{}_pF_q(a_1,\ldots,a_p; b_1,\ldots,b_q; S) = \sum_{j=0}^{\infty} \sum_J \frac{(a_1)_J \cdots (a_p)_J}{(b_1)_J \cdots (b_q)_J} \frac{C_J(S)}{j!}.
$$
(2.11)

In (2.11) J indicates a partition of the integer j into not more than n parts, where S is an $n \times n$ matrix. A partition J of weight r is a set of r positive integers $\langle j_1,\ldots,j_r \rangle$ such that $\sum_{i=1}^{r} j_i = j$. For example $\langle 2, 1 \rangle$ and $\langle 1, 1, 1 \rangle$ are partitions of 3 and are conventionally written (21) and (1^3). The coefficients $(a)_J$ and $(b)_J$ in (2.11) are multivariate hypergeometric coefficients defined by

$$
(a)_J = \prod_{i=1}^{n} \left(a - \tfrac{1}{2}(i-1) \right)_{j_i}, \quad \text{for } J = \langle j_1,\ldots,j_n \rangle,
$$

and where

$$
(\lambda)_j = \lambda(\lambda+1)\ldots(\lambda+j-1) = \Gamma(\lambda+j)/\Gamma(\lambda).
$$

The factor $C_J(S)$ in (2.11) is a zonal polynomial and can be represented as a symmetric homogeneous polynomial of degree j of the latent roots of S. General formulae for these polynomials are presently known only for the case $m = 2$ or when the partition of j has only one part, $J = (j)$ [see James (1964)]. Tabulations are available for low values of j and are reported in James (1964). These can be conveniently expressed in terms of the elementary symmetric functions of the latent roots of S [Constantine (1963)] or in terms of the quantities:

s_m = sum of the mth powers of the latent roots of S.

Thus, the first few zonal polynomials take the form:

degree j	partition J	zonal polynomial $C_J(S)$
1	1	s_1
2	1^2	$\frac{2}{3}(s_1^2 - s_2)$
	2	$\frac{1}{3}(s_1^2 + 2s_2)$
3	1^3	$\frac{1}{3}(s_1^3 - 3s_1 s_2 + 2s_3)$
	21	$\frac{2}{5}(s_1^3 + s_1 s_2 - 2s_3)$
	3	$\frac{1}{15}(s_1^3 + 6s_1 s_2 + 8s_3)$

[see, for example, Johnson and Kotz (1972, p. 171)]. Algorithms for the extraction of the coefficients in these polynomials have been written [see James (1968) and McLaren (1976)] and a complete computer program for their evaluation has recently been developed and made available by Nagel (1981). This is an important development and will in due course enhance what is at present our very limited ability to numerically compute and readily interpret multiple infinite series such as (2.11). However, certain special cases of (2.11) are already recognizable in terms of simpler functions: when $n = 1$ we have the classical hypergeometric functions

$$_pF_q(a_1,\ldots,a_p; b_1,\ldots,b_q; s) = \sum_{j=0}^{\infty} \frac{(a_1)_j\ldots(a_p)_j s^j}{(b_1)_j\ldots(b_q)_j j!}$$

[see, for example, Lebedev (1965, ch. 9)]; and when $p = q = 0$ we have

$$_0F_0(S) = \sum_{j=0}^{\infty} \sum_J C_J(S)/j! = \text{etr}(S),$$

which generalizes the exponential series and which is proved in James (1961); and when $p = 1$ and $q = 0$ we have

$$_1F_0(a; S) = \sum_{j=0}^{\infty} \sum_J \frac{(a)_j}{j!} C_J(S) = (\det(I - S))^{-a},$$

which generalizes the binomial series [Constantine (1963)]. The series $_0F_1(:)$ in the non-central Wishart density (2.10) generalizes the classical Bessel function. [The reader may recall that the non-central χ^2 density can be expressed in terms of the modified Bessel function of the first kind – see, for example, Johnson and Kotz (1970, p. 133).] In particular, when $n = 1$, $\Omega = 1$, $\overline{M} = \lambda$, and $A = a$ is a scalar, we have

$$\text{pdf}(a) = \frac{\exp\{-\frac{1}{2}(a + \lambda)\}}{2^{T/2}\Gamma(T/2)} a^{T/2-1} \sum_{j=0}^{\infty} \frac{(\frac{1}{4}\lambda a)^j}{(T/2)_j j!}$$

$$= \frac{\exp\{-\frac{1}{2}(a + \lambda)\}}{2^{T/2}} \sum_{j=0}^{\infty} \frac{\lambda^j a^{T/2+j-1}}{\Gamma(T/2 + j) j! 2^{2j}}. \tag{2.12}$$

This is the usual form of the p.d.f. of a non-central χ^2 variate.

3. Exact theory in the simultaneous equations model

3.1. The model and notation

We write the structural form of a system of G contemporaneous simultaneous stochastic equations as

$$YB + ZC = U, \tag{3.1}$$

and its reduced form as

$$Y = Z\Pi + V, \tag{3.2}$$

where $Y' = [y_1, \ldots, y_T]$ is a $G \times T$ matrix of T observations of G endogenous variables, $Z' = [z_1, \ldots, z_T]$ is a $K \times T$ matrix of T observations of K non-random exogenous variables, and $U' = [u_1, \ldots, u_T]$ is a $G \times T$ matrix of the structural disturbances of the system. The coefficient matrices B $(G \times G)$ and C $(K \times G)$ comprise parameters that are to be estimated from the data and about which some *a priori* economic knowledge is assumed; usually this takes the form of simple (and frequently zero exclusion type) restrictions upon certain of the coefficients together with conventional normalization restrictions. As is usual in this contemporaneous version of the SEM (see Chapter 4 and Chapter 7 in this Handbook by Hsiao and Hausman, respectively), it is also assumed that the u_t $(t = 1, \ldots, T)$ are serially independent random vectors distributed with zero mean vector and (non-singular) covariance matrix Σ. The coefficient matrix B is assumed to be non-singular and these conditions imply that the rows, v_t', of V in (3.2) are independent random vectors with zero mean vector and covariance matrix $\Omega = B'^{-1}\Sigma B^{-1}$. To permit the development of a distribution theory for finite sample sizes we will, unless otherwise explicitly stated, extend these conventional assumptions by requiring v_t $(t = 1, \ldots, T)$ to be i.i.d. $N(0, \Omega)$. Extensions to non-normal errors are possible [see Phillips (1980b), Satchell (1981), and Knight (1981)] but involve further complications.

We will frequently be working with a single structural equation of (3.1) which we write in the following explicit form that already incorporates exclusion type restrictions:

$$y_1 = Y_2\beta + Z_1\gamma + u \tag{3.3}$$

or

$$y_1 = W_1\delta + u, \qquad W_1 = \left[Y_2 \vdots Z_1\right], \qquad \delta' = (\beta', \gamma'), \tag{3.4}$$

where y_1 $(T \times 1)$ and Y_2 $(T \times n)$ contain T observations of $n + 1$ included

endogenous variables. Z_1 is a $T \times K_1$ matrix of included exogenous variables, and u is the vector of random disturbances on this equation. Thus, (3.3) explicitly represents one column of the full model (3.1). The reduced form of (3.3) is written

$$\left[y_1 \vdots Y_2 \right] = \left[Z_1 \vdots Z_2 \right] \begin{bmatrix} \pi_{11} & \Pi_{12} \\ \pi_{21} & \Pi_{22} \end{bmatrix} + \left[v_1 \vdots V_2 \right] \tag{3.5}$$

or

$$X = Z\Pi^s + V^s, \qquad X = \left[y_1 \vdots Y_2 \right], \qquad Z = \left[Z_1 \vdots Z_2 \right], \tag{3.5'}$$

where Z_2 is a $T \times K_2$ matrix of exogenous variables excluded from (3.3). To simplify notation the selection superscripts in (3.5') will be omitted in what follows. The system (3.5) represents $n + 1$ columns of the complete reduced form (containing $G \geqslant n + 1$ columns) given in (3.2). The total number of exogenous variables in (3.5) is $K = K_1 + K_2$ and the observation matrix Z is assumed to have full rank, K. We also assume that $K_2 \geqslant n$ and the submatrix Π_{22} ($K_2 \times n$) in (3.4) has full rank ($= n$) so that the structural equation is identified. Note that (3.3) can be obtained by postmultiplication of (3.5) by $(1, -\beta')'$ which yields the relations

$$\pi_{11} - \Pi_{12}\beta = \gamma, \qquad \pi_{21} - \Pi_{22}\beta = 0. \tag{3.6}$$

We will sometimes use the parameter $N = K_2 - n$ to measure the degree by which the structural relation (3.3) is overidentified.

3.2. Generic statistical forms of common single equation estimators

As argued in Section 2.2, most econometric estimators and test statistics can be expressed as simple functions of the sample moments of the data. In the case of the commonly used single equation estimators applied to (3.3) we obtain relatively simple generic statistical expressions for these estimators in terms of the elements of moment matrices which have Wishart distributions of various degrees of freedom and with various non-centrality parameter matrices. This approach enables us to characterize the distribution problem in a simple but powerful way for each case. It has the advantage that the characterization clarifies those cases for which the estimator distributions will have the same mathematical forms but for different values of certain key parameters and it provides a convenient first base for the mathematics of extracting the exact distributions. Historically the approach was first used by Kabe (1963, 1964) in the econometrics context and has since been systematically employed by most authors working in this field. An excellent recent discussion is given by Mariano (1982).

We will start by examining the IV estimator. δ_{IV}, of the coefficient vector $\delta' = (\beta', \gamma')$ in (3.3)–(3.4) based on the instrument matrix H. δ_{IV} minimizes the quantity

$$(y - W_1\delta)'H(H'H)^{-1}H'(y - W_1\delta),\tag{3.7}$$

and writing

$$P_D = D(D'D)^{-1}D', \qquad Q_D = I - P_D.\tag{3.8}$$

we obtain by stepwise minimization of (3.7) the following explicit expressions for the IV estimators of the subvectors β and γ:

$$\gamma_{IV} = (Z_1'P_HZ_1)^{-1}Z_1'P_H(y_1 - Y_2\beta_{IV}),\tag{3.9}$$

$$\beta_{IV} = \left\{Y_2'\left[P_H - P_HZ_1(Z_1'P_HZ_1)^{-1}Z_1'P_H\right]Y_2\right\}^{-1}$$
$$\times \left\{Y_2'\left[P_H - P_HZ_1(Z_1'P_HZ_1)^{-1}Z_1'P_H\right]y_1\right\}.\tag{3.10}$$

In the usual case where H includes Z_1 as a subset of its instruments and $P_HZ_1 = Z_1$ we have the simple formulae:

$$\gamma_{IV} = (Z_1'Z_1)^{-1}Z_1'(y_1 - Y_2\beta_{IV}).\tag{3.11}$$

$$\beta_{IV} = \left[Y_2'(P_H - P_{Z_1})Y_2\right]^{-1}\left[Y_2'(P_H - P_{Z_1})y_1\right].\tag{3.12}$$

We define the moment matrix

$$A(P_H) = \begin{bmatrix} a_{11}(P_H) & a_{21}'(P_H) \\ a_{21}(P_H) & A_{22}(P_H) \end{bmatrix} = \begin{bmatrix} y_1'(P_H - P_{Z_1})y_1 & y_1'(P_H - P_{Z_1})Y_2 \\ Y_2'(P_H - P_{Z_1})y_1 & Y_2'(P_H - P_{Z_1})Y_2 \end{bmatrix}$$
$$= X'(P_H - P_{Z_1})X.\tag{3.13}$$

The generic statistical form for the estimator β_{IV} in (3.12) is then

$$\beta_{IV} = A_{22}^{-1}(P_H)a_{21}(P_H).\tag{3.14}$$

This specializes to the cases of OLS and 2SLS where we have, respectively,

$$\beta_{OLS} = \left[Y_2'Q_{Z_1}Y_2\right]^{-1}\left[Y_2'Q_{Z_1}y_1\right] = A_{22}^{-1}(I)a_{21}(I),\tag{3.15}$$

$$\beta_{2SLS} = \left[Y_2'(P_Z - P_{Z_1})Y_2\right]^{-1}\left[Y_2'(P_Z - P_{Z_1})y_1\right] = A_{22}^{-1}(P_Z)a_{21}(P_Z).\tag{3.16}$$

In a similar way we find that the k-class estimator $\beta_{(k)}$ of β has the generic form

$$\beta_{(k)} = \left\{ Y_2'\left[k(P_Z - P_{Z_1}) + (1-k)Q_{Z_1} \right] Y_2 \right\}^{-1}$$

$$\times \left\{ Y_2'\left[k(P_Z - P_{Z_1}) + (1-k)Q_{Z_1} \right] y_1 \right\}$$

$$= \left[kA_{22}(P_Z) + (1-k)A_{22}(I) \right]^{-1} \left[ka_{21}(P_Z) + (1-k)a_{21}(I) \right]. \quad (3.17)$$

The LIML estimator, β_{LIML}, of β minimizes the ratio

$$\frac{\beta_\Delta' A(I)\beta_\Delta}{\beta_\Delta'\left[A(I) - A(P_Z)\right]\beta_\Delta} = 1 + \frac{\beta_\Delta' A(P_Z)\beta_\Delta}{\beta_\Delta'\left[A(I) - A(P_Z)\right]\beta_\Delta} = 1 + \frac{\beta_\Delta' W\beta_\Delta}{\beta_\Delta' S\beta_\Delta}, \quad \text{say,}$$

$$(3.18)$$

where $\beta_\Delta' = (1, -\beta')$ and β_{LIML} satisfies the system

$$\left\{ A(I) - \lambda\left[A(I) - A(P_Z)\right] \right\}\beta_\Delta = 0, \quad (3.19)$$

where λ is the minimum of the variance ratio in (3.18). Thus, β_{LIML} is given by the generic form

$$\beta_{\text{LIML}} = \left[\lambda A_{22}(P_Z) + (1-\lambda)A_{22}(I) \right]^{-1} \left[\lambda a_{21}(P_Z) + (1-\lambda)a_{21}(I) \right],$$

$$(3.20)$$

that is, the k-class estimator (3.17) with $k = \lambda$.

The above formulae show that the main single equation estimators depend in a very similar way on the elements of an underlying moment matrix of the basic form (3.13) with some differences in the projection matrices relevant to the various cases. The starting point in the derivation of the p.d.f. of these estimators of β is to write down the joint distribution of the matrix A in (3.13). To obtain the p.d.f. of the estimator we then transform variates so that we are working directly with the relevant function $A_{22}^{-1}a_{21}$. The final step in the derivation is to integrate over the space of the auxiliary variates, as prescribed in the general case of (2.8) above, which in this case amounts essentially to (a_{11}, A_{22}). This leaves us with the required density function of the estimator.

The mathematical process outlined in the previous section is simplified, without loss of generality, by the implementation of standardizing transformations. These transformations were first used and discussed by Basmann (1963, 1974). They reduce the sample second moment matrix of the exogenous variables to the identity matrix (orthonormalization) and transform the covariance matrix of the endogenous variables to the identity matrix (canonical form). Such transformations help to reduce the parameter space to an essential set and identify the

critical parameter functions which influence the shape of the distributions.[7] They are fully discussed in Phillips (1982e) and are briefly reviewed in the following section.

3.3. The standardizing transformations

We first partition the covariance matrix Ω conformably with $[y_1 \vdots Y_2]$ as

$$\Omega = \begin{bmatrix} \omega_{11} & \omega'_{21} \\ \omega_{21} & \Omega_{22} \end{bmatrix}.$$ (3.21)

Then the following result [proved in Phillips (1982e)] summarizes the effect of the standardizing transformations on the model.

Theorem 3.3.1

There exist transformations of the variables and parameters of the model given by (3.3) and (3.5) which transform it into one in which

$$T^{-1}Z'Z = I_K \quad \text{and} \quad \Omega = I_{n+1}.$$ (3.22)

Under these transformations (3.3) and (3.5) can be written in the form

$$y_1^* = Y_2^* \beta^* + \bar{Z}_1 \bar{\gamma}^* + u^*$$ (3.23)

and

$$\left[y_1^* \vdots Y_2^* \right] = \bar{Z} \bar{\Pi}^* + \bar{V},$$ (3.24)

where $T^{-1}\bar{Z}'\bar{Z} = I_K$ and the rows of $[y_1^* \vdots Y_2^*]$ are uncorrelated with covariance matrix given by I_{n+1}. Explicit formulae for the new coefficients in (3.23) are

$$\beta^* = \left(\omega_{11} - \omega'_{21}\Omega_{22}^{-1}\omega_{21} \right)^{-1/2} \Omega_{22}^{1/2} \left(\beta - \Omega_{22}^{-1}\omega_{21} \right)$$ (3.25)

and

$$\bar{\gamma}^* = \left(\frac{Z_1'Z_1}{T} \right)^{1/2} \left(\omega_{11} - \omega'_{21}\Omega_{22}^{-1}\omega_{21} \right)^{-1/2} \gamma.$$ (3.26)

[7]As argued recently by Mariano (1982), these reductions also provide important guidelines for the design of Monte Carlo experiments (at least in the context of SEMs) by indicating the canonical parameter space which is instrumental in influencing the shape of the relevant small sample distributions and from which a representative sample of points can be taken to help reduce the usual specificity of simulation findings

These transformations preserve the number of excluded exogenous variables in the structural equation and the rank condition for its identifiability.   □

It turns out that the commonly used econometric estimators of the standardized coefficients β^* and $\bar{\gamma}^*$ in (3.23) are related to the unstandardized coefficient estimators by the same relations which define the standard coefficients, namely (3.25) and (3.26). Thus, we have the following results for the 2SLS estimator [see Phillips (1982e) once again for proofs].

Theorem 3.3.2

The 2SLS estimator, β_{2SLS}, of the coefficients of the endogenous variables in (3.3) are invariant under the transformation by which the exogenous variables are orthornormalized. The 2SLS estimator, γ_{2SLS}, is not, in general, invariant under this transformation. The new exogenous variable coefficients are related to the original coefficients under the transformation $\bar{\gamma} = J_{11}\gamma$ and to the estimators by the corresponding equation $\bar{\gamma}_{2SLS} = J_{11}\gamma_{2SLS}$, where $J_{11} = (Z_1'Z_1/T)^{1/2}$.   □

Theorem 3.3.3

The 2SLS estimators of β^* and $\bar{\gamma}^*$ in the standardized model (3.23) are related to the corresponding estimators of β and γ in the unstandardized model (3.3) by the equations:

$$\beta_{2SLS}^* = \left(\omega_{11} - \omega_{21}'\Omega_{22}^{-1}\omega_{21}\right)^{-1/2}\Omega_{22}^{1/2}\left(\beta_{2SLS} - \Omega_{22}^{-1}\omega_{21}\right) \tag{3.27}$$

and

$$\bar{\gamma}_{2SLS}^* = \left(\frac{Z_1'Z_1}{T\left(\omega_{11} - \omega_{21}'\Omega_{22}^{-1}\omega_{21}\right)}\right)^{1/2}\gamma_{2SLS}. \tag{3.28}$$

               □

Results that correspond to these for 2SLS can be derived similarly for other estimators such as IV and LIML [see Phillips (1982e) for details].

The canonical transformation induces a change in the coordinates by which the variables are measured and therefore (deliberately) affects their covariance structure. Some further properties of the transformed structural equation (3.23) are worth examining. Let us first write (3.23) in individual observation form as

$$y_{1t}^* = y_{2t}^{*'}\beta^* + \bar{z}_{1t}'\bar{\gamma}^* + u_t^*. \tag{3.29}$$

Then, by simple manipulations we find that

$$\text{cov}(y_{2t}^*, u_t^*) = -\beta^*, \tag{3.30}$$

$$\text{var}(u_t^*) = 1 + \beta^{*'}\beta^*, \tag{3.31}$$

and

$$\text{corr}(y_{2t}^*, u_t^*) = -\beta^* / (1 + \beta^{*\prime}\beta^*)^{1/2}. \tag{3.32}$$

These relations show that the transformed coefficient vector, β^*, in the standardized model contains the key parameters which determine the correlation pattern between the included variables and the errors. In particular, when the elements of β^* become large the included endogenous variables and the error on the equation become more highly correlated. In these conditions, estimators of the IV type will normally require larger samples of data to effectively purge the included variables of their correlation with the errors. We may therefore expect these estimators to display greater dispersion in small samples and slower convergence to their asymptotic distributions under these conditions than otherwise. These intuitively based conjectures have recently been substantiated by the extensive computations of exact densities by Anderson and Sawa (1979)[8] and the graphical analyses by Phillips (1980a, 1982a) in the general case.

The vector of correlations corresponding to (3.32) in the unstandardized model is given by

$$\text{corr}(y_{2t}, u_t) = \frac{\Omega_{22}^{1/2}(\Omega_{22}^{-1}\omega_{21} - \beta)}{(\omega_{11} - 2\beta'\omega_{21} + \beta'\Omega_{22}\beta)^{1/2}} = \frac{-\beta^*}{(1 + \beta^{*\prime}\beta^*)^{1/2}}, \tag{3.33}$$

so that for a fixed reduced-form error covariance matrix, Ω, similar conditions persist as the elements of β grow large. Moreover, as we see from (3.33), the transformed structural coefficient β^* is itself determined by the correlation pattern between regressors and error in the unstandardized model. The latter (like β^*) can therefore be regarded as one of the critical sets of parameters that influence the shape of the distribution of the common estimators of the coefficient β.

3.4. The analysis of leading cases

There are two special categories of models in which the exact density functions of the common SEM estimators can be extracted with relative ease. In the first category are the just identified structural models in which the commonly used consistent estimators all reduce to indirect least squares (ILS) and take the form

$$\beta_{\text{ILS}} = \left[\overline{Z}_2'Y_2\right]^{-1}\left[\overline{Z}_2'y_1\right] \tag{3.34}$$

---

[8]See also the useful discussion and graphical plots in Anderson (1982)

of a matrix ratio of normal variates. In the two endogenous variable case (where $n = 1$) this reduces to a simple ratio of normal variates whose p.d.f. was first derived by Fieiller (1932) and in the present case takes the form[9]

$$\text{pdf}(r) = \frac{\exp\left\{-\frac{\mu^2}{2}(1+\beta^2)\right\}}{\pi(1+r^2)}\, {}_1F_1\left(1, \frac{1}{2}; \frac{\mu^2}{2}\frac{(1+\beta r)^2}{1+r^2}\right), \tag{3.35}$$

where $\mu^2 = T\Pi'_{22}\Pi_{22}$ is the scalar concentration parameter.[10] In the general case of $n + 1$ included endogenous variables the density (3.35) is replaced by a multivariate analogue in which the ${}_1F_1$ function has a matrix argument [see (3.46) below]. The category of estimators that take the generic form of a matrix ratio of normal variates, as in (3.34), also include the general IV estimator in the overidentified case provided the instruments are non-stochastic: that is, if $\beta_{IV} = [W'Y_2]^{-1}[W'y_1]$ and the matrix W is non-stochastic, as distinct from its usual stochastic form in the case of estimators like 2SLS in overidentified equations. This latter case has been discussed by Mariano (1977). A further application of matrix ratios of normal variates related to (3.34) occurs in random coefficient SEMs where the reduced-form errors are a matrix quotient of the form $A^{-1}a$ where both a and the columns of A are normally distributed. Existing theoretical work in this area has proceeded essentially under the hypothesis that $\det A$ is non-random [see Kelejian (1974)] and can be generalized by extending (3.35) to the multivariate case in much the same way as the exact distribution theory of (3.34), which we will detail in Section 3.5 below.

The second category of special models that facilitate the development of an exact distribution theory are often described as leading cases of the fully para-meterized SEM.[11] In these leading cases, certain of the critical parameters are set equal to zero and the distribution theory is developed under this null hypothesis. In the most typical case, this hypothesis prescribes an absence of simultaneity and a specialized reduced form which ensures that the sample moments of the data on which the estimator depends have central rather than (as is typically the case) non-central distributions.[12] The adjective "leading" is used advisedly since the distributions that arise from this analysis typically provide the leading term in the multiple series representation of the true density that applies when the null

---

[9] This density is given, for example, in Mariano and McDonald (1979)

[10] This parameter is so called because as $\mu^2 \to \infty$ the commonly used single equation estimators all tend in probability to the true parameter Thus, the distributions of these estimators all "concentrate" as $\mu^2 \to \infty$, even if the sample size T remains fixed See Basmann (1963) and Mariano (1975) for further discussion of this point

[11] See Basmann (1963) and Kabe (1963, 1964)

[12] Some other specialized SEM models in which the distributions of commonly used estimators depend only on central Wishart matrices are discussed by Wegge (1971).

hypothesis itself no longer holds. As such the leading term provides important information about the shape of the distribution by defining a primitive member of the class to which the true density belongs in the more general case. In the discussion that follows, we will illustrate the use of this technique in the case of IV and LIML estimators.[13]

We set $\beta = 0$ in the structural equation (3.3) and $\Pi_{22} = 0$ in the reduced form so that y_1 and y_2 (taken to be a vector of observations on the included endogenous variable now that $n = 1$) are determined by the system[14]

$$y_1 = Z_1\gamma + u, \qquad y_2 = Z_1\Pi_{12} + v_2. \tag{3.36}$$

The IV estimator of β is

$$\beta_{IV} = (y_2'Z_3Z_3'y_2)^{-1}(y_2'Z_3Z_3'y_1). \tag{3.37}$$

under the assumption that standardizing transformations have already been performed. Let Z_3 be $T \times K_3$ with $K_3 \geq 1$ so that the total number of instruments is $K_1 + K_3$. Simple manipulations now confirm that the p.d.f. of β_{IV} is given by [see Phillips (1982e)]

$$\text{pdf}(r) = \left[B\left(\frac{1}{2}, \frac{K_3}{2} \right) \right]^{-1} (1 + r^2)^{-(K_3 + 1)/2}, \tag{3.38}$$

where $B(\frac{1}{2}, K_3/2)$ is the beta function. This density specializes to the case of 2SLS when $K_3 = K_2$ and OLS when $K_3 = T - K_1$. [In the latter case we may use (3.15) and write $Q_{Z_1} = I - T^{-1}Z_1Z_1' = C_1C_1'$, where C_1 is a $T \times (T - K_1)$ matrix whose columns are the orthogonal latent vectors of Q_{Z_1} corresponding to unit latent roots.] The density (3.38) shows that integral moments of the distribution exist up to order $K_3 - 1$: that is, in the case of 2SLS, $K_2 - 1$ (or the degree of overidentification) and, in the case of OLS, $T - K_1 - 1$.

The result corresponding to (3.38) for the case of the LIML estimator is [see Phillips (1982e) for the derivation]

$$\text{pdf}(r) = \left[\pi(1 + r^2) \right]^{-1}, \qquad -\infty < r < \infty. \tag{3.39}$$

---

[13]An example of this type of analysis for structural variance estimators is given in Section 3 7

[14]In what follows it will often not be essential that both $\beta = 0$ and $\Pi_{22} = 0$ for the development of the "leading case" theory What is essential is that $\Pi_{22} = 0$, so that the structural coefficients are, in fact, unidentifiable Note that the reduced-form equations take the form

$$y_1 = Z_1\pi_{11} + v_1, \qquad Y_2 = Z_1\Pi_{12} + V_2,$$

when $\Pi_{22} = 0$ The first of these equations corresponds to (3.36) in the text when $\beta = 0$

Thus, the exact sampling distribution of the β_{LIML} is Cauchy in this leading case. In fact, (3.39) provides the leading term in the series expansion of the density of LIML derived by Mariano and Sawa (1972) in the general case where $\beta \neq 0$ and $\Pi_{22} \neq 0$. We may also deduce from (3.39) that β_{LIML} has no finite moments of integral order, as was shown by Mariano and Sawa (1972) and Sargan (1970). This analytic property of the exact distribution of β_{LIML} is associated with the fact that the distribution displays thicker tails than that of β_{IV} when $K_3 > 1$. Thus, the probability of extreme outliers is in general greater for β_{LIML} than for β_{IV}. This and other properties of the distributions of the two estimators will be considered in greater detail in Sections 3.5 and 3.6.

3.5. The exact distribution of the IV estimator in the general single equation case

In the general case of a structural equation such as (3.3) with $n + 1$ endogenous variables and an arbitrary number of degrees of overidentification, we can write the IV estimator β_{IV} of β in the form

$$\beta_{\text{IV}} = (Y_2' Z_3 Z_3' Y_2)^{-1}(Y_2' Z_3 Z_3' y_1), \tag{3.40}$$

where the standardizing transformations are assumed to have been carried out. This is the case where $H = [Z_1 \vdots Z_3]$ is a matrix of $K_1 + K_3$ instruments used in the estimation of the equation. To find the p.d.f. of β_{IV} we start with the density of the matrix:

$$A = \begin{bmatrix} a_{11} & a_{21}' \\ a_{21} & A_{22} \end{bmatrix} = T^{-1} \begin{bmatrix} y_1' Z_3 Z_3' y_1 & y_1' Z_3 Z_3' Y_2 \\ Y_2' Z_3 Z_3' y_1 & Y_2' Z_3 Z_3' Y_2 \end{bmatrix}.$$

In general this will be non-central Wishart with a p.d.f. of the form

$$\text{pdf}(A) = \frac{\text{etr}\left(-\tfrac{1}{2}MM'\right)}{2^{(1/2)K_3(n+1)}\Gamma_{n+1}\left(\dfrac{K_3}{2}\right)} {}_0F_1\left(\frac{K_3}{2}; \frac{1}{4}MM'A\right)$$

$$\times \text{etr}\left(-\tfrac{1}{2}A\right)(\det A)^{(1/2)(K_3-n-2)}$$

[see (2.10) above] where $M = E(T^{-1/2}X'Z_3) = T^{-1/2}\Pi'Z'Z_3$.

We now introduce a matrix S which selects those columns of Z_2 which appear in Z_3, so that $Z_3 = Z_2 S$. Then, using the orthogonality of the exogenous variables, we have

$$M' = T^{-1/2}Z_3'Z\Pi = T^{1/2}[OS']\Pi = T^{1/2}S'[\pi_{21}\Pi_{22}] = T^{1/2}S'\Pi_{22}[\beta, I]$$

in view of the relations (3.6) given above. Writing $\Pi'_{22}SS'\Pi_{22}$ as $\bar{\Pi}'_{22}\bar{\Pi}_{22}$, where $\bar{\Pi}_{22}$ is an $n \times n$ matrix (which is non-singular since the structural equation (3.3) is assumed to be identified), we find that

$$\text{etr}(-\tfrac{1}{2}MM') = \text{etr}\left\{ -\frac{T}{2}(I + \beta\beta')\bar{\Pi}'_{22}\bar{\Pi}_{22} \right\}.$$

Moreover, since the non-zero latent roots of $MM'A$ are the latent roots of

$$T\bar{\Pi}_{22}[\beta, I]A\begin{bmatrix} \beta' \\ I \end{bmatrix}\bar{\Pi}'_{22},$$

(3.41) becomes

$$\frac{\text{etr}\left\{ -\frac{T}{2}(I + \beta\beta')\bar{\Pi}'_{22}\bar{\Pi}_{22} \right\}}{2^{(1/2)K_3(n+1)}\Gamma_{n+1}\left(\frac{K_3}{2}\right)}\, {}_0F_1\left(\frac{K_3}{2}; \frac{T}{4}\bar{\Pi}_{22}[\beta, I]A\begin{bmatrix} \beta' \\ I \end{bmatrix}\bar{\Pi}'_{22} \right)$$

$$\times \text{etr}(-\tfrac{1}{2}A)(\det A)^{(1/2)(K_3 - n - 2)}.$$

We now transform variables from the matrix variate A to $w = a_{11} - r'A_{22}r$, $r = A_{22}^{-1}a_{21}$, and $A_{22} = A_{22}$. The Jacobian of the transformation is $\det A_{22}$ and we have

$$\text{pdf}(w, r, A_{22})$$

$$= \frac{\text{etr}\left\{ -\frac{T}{2}(I + \beta\beta')\bar{\Pi}'_{22}\bar{\Pi}_{22} \right\}}{2^{(1/2)K_3(n+1)}\Gamma_{n+1}\left(\frac{K_3}{2}\right)}$$

$$\times {}_0F_1\left(\frac{K_3}{2}; \frac{T}{4}\{w\bar{\Pi}_{22}\beta\beta'\bar{\Pi}'_{22} + \bar{\Pi}_{22}(1 + \beta r')A_{22}(I + r\beta')\bar{\Pi}'_{22}\} \right)$$

$$\times \exp(-\tfrac{1}{2}(w + r'A_{22}r))\text{etr}(-\tfrac{1}{2}A_{22})\{w\det A_{22})^{(1/2)(K_3 - n - 2)}\det A_{22}$$

$$= \frac{\text{etr}\left\{ -\frac{T}{2}(I + \beta\beta')\bar{\Pi}'_{22}\bar{\Pi}_{22} \right\}}{2^{(1/2)K_3(n+1)}\Gamma_{n+1}\left(\frac{K_3}{2}\right)}$$

$$\times {}_0F_1\left(\frac{K_3}{2}; \left\{ \frac{T}{4}w\bar{\Pi}_{22}\beta\beta'\bar{\Pi}'_{22} + \bar{\Pi}_{22}(I + \beta r')A_{22}(I + r\beta')\bar{\Pi}'_{22} \right\} \right)$$

$$\times \exp(-\tfrac{1}{2}w)\text{etr}(-\tfrac{1}{2}(I + rr')A_{22})w^{(1/2)(K_3 - n - 2)}(\det A_{22})^{(1/2)(K_3 - n)}.$$

Define $L = K_3 - n$ and introduce the new matrix variate $B = (I + rr')^{1/2} A_{22}$ $(I + rr')^{1/2}$. The Jacobian of this transformation is $[\det(I + rr')]^{-(n+1)/2}$ and we have

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

$$\times {}_0F_1\left[\frac{L+n}{2} ; \frac{T}{4}\left\{ w\bar{\Pi}_{22}\beta\beta'\bar{\Pi}_{22}' + \bar{\Pi}_{22}(I + \beta r')(I + rr')^{-1/2} \right.\right.$$

$$\left.\left. \times B(I + rr')^{-1/2}(I + \beta r')\bar{\Pi}_{22}' \right\} \right] \exp(-\tfrac{1}{2}w)$$

$$\times \text{etr}(-\tfrac{1}{2}B)w^{L/2-1}(\det B)^{L/2}. \tag{3.42}$$

As in the general scheme of development outlined in Section 2.2 we need to integrate out the auxiliary variates (w, B) in (3.42) in order to find the analytic form of the density of β_{IV}. This problem was the main obstacle in the development of an exact distribution theory for single equation estimators in the general case for over a decade following the work of Richardson (1968) and Sawa (1969) that dealt explicitly with the two endogenous variable case $(n = 1)$. In this latter case the ${}_0F_1$ function in (3.42) can be replaced by a power series in the argument:

$$\frac{T}{4}\left\{ w\bar{\Pi}_{22}\beta\beta'\bar{\Pi}_{22}' + \bar{\Pi}_{22}(I + \beta r')(I + rr')^{-1/2}B(I + rr')^{-1/2}(I + r\beta')\bar{\Pi}_{22}' \right\} \tag{3.43}$$

which, when $n = 1$, is a scalar. Powers of this variable may now be expanded in binomial series and inspection of (3.42) shows that terms of this double series may then be integrated simply as gamma functions. When $n > 1$, (3.43) is a matrix and the series development of the ${}_0F_1$ function is in terms of zonal polynomials of this matrix. In the absence of an algebra to develop a binomial type expansion for zonal polynomials of the sum of two matrices, integration of the auxiliary variables (w, B) in (3.42) appeared impossible. However, a solution to this difficulty was found by Phillips (1980a). The idea behind the method developed in this article is to use an alternative representation of the ${}_0F_1$ function in which the argument matrix (3.43) is thrown up into an exponent. The two elements of the binomial matrix sum (3.43) can then effectively be separated and integrated out. (We will not give the full argument here but refer the reader to the article for details.)[15] In short, the process leads to the following analytic form for the exact

[15]An alternative approach to the extraction of the exact density of β_{IV} from (3.42) is given in Phillips (1980a, appendix B) and directly involves the algebra of expanding the zonal polynomial of a sum of two matrices into a sum of more basic polynomials in the constituent matrices. This algebra was developed by Davis (1980a, 1980b) and has recently been extended by Chikuse (1981) to matrix multinomial expansions of zonal polynomials of the sum of several matrices.

finite sample density of β_{IV}:

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

$$\times \sum_{j=0}^{\infty} \frac{\left(\frac{L}{2}\right)_j}{j!\,\Gamma_n\left(\frac{L+n}{2}+j\right)}$$

$$\times\left[\left(\frac{T}{2}\beta'\overline{\Pi}'_{22}\left(\text{adj}\frac{\partial}{\partial W}\right)\overline{\Pi}_{22}\beta\right)^j (\det(I+W))^{(L-1)/2+j}\right.$$

$$\times {}_1F_1\left(\frac{L+n+1}{2},\frac{L+n}{2}+j;\frac{T}{2}(I+W)\overline{\Pi}_{22}\right.$$

$$\left.\left.\times(I+\beta r')(I+rr')^{-1}(I+r\beta')\overline{\Pi}'_{22}\right)\right]_{W=0}. \qquad (3.44)$$

In (3.44) $L = K_3 - n$ is the number of surplus instruments used in the estimation of β. That is, $K_1 + K_3$ instruments are used and at least $K_1 + n$ are needed to perform the estimation by the traditional IV procedure. Thus, when $K_3 = K_2$ and $L = K_2 - n$, (3.44) gives the p.d.f. of the 2SLS estimator of β; and when $K_1 + K_3 = T$, so that $K_3 = T - K_1$ and $L = T - K_1 - n$, (3.44) gives the p.d.f. of the OLS estimator of β.

The matrix $W(n \times n)$ in (3.44) contains auxiliary variables that are useful in reducing the integral from which (3.44) is derived and $\text{adj}(\partial/\partial W)$ denotes the adjoint of the matrix differential operator $\partial/\partial W$. We note that when $n = 1$, W is a scalar, $\text{adj}(\partial/\partial W) = 1$, and (3.44) becomes

$$\text{pdf}(r) = \frac{\exp\left\{-\frac{\mu^2}{2}(1+\beta^2)\right\}\Gamma\left(\frac{L+2}{2}\right)}{\pi^{1/2}\Gamma\left(\frac{L+1}{2}\right)(1+r^2)^{(L+2)/2}}$$

$$\times \sum_{j=0}^{\infty} \frac{\left(\frac{L}{2}\right)_j}{j!\left(\frac{L+1}{2}\right)_j}\left(\frac{\mu^2}{2}\beta^2\right)^j {}_1F_1\left(\frac{L+2}{2},\frac{L+1}{2}+j;\frac{\mu^2}{2}\frac{(1+\beta r)^2}{1+r^2}\right),$$

$$(3.45)$$

in which $\mu^2 = T\overline{\Pi}_{22}^2 = T\Pi'_{22}\Pi_{22}$ is the scalar concentration parameter [recall (3.35) and footnote 10]. The density (3.45) was first derived for 2SLS ($L = K_2 - 1$) and OLS ($L = T - K_1 - 1$) by Richardson (1968) and Sawa (1969).

When $L = 0$ in (3.44) the series corresponding to the suffix j terminates at the first term and we have

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

$$\times {}_1F_1\left(\frac{n+1}{2}; \frac{n}{2}; \frac{T}{2}\bar{\Pi}_{22}(I + \beta r')(I + rr')^{-1}(I + r\beta')\bar{\Pi}'_{22}\right).$$

$$\tag{3.46}$$

That is, a single term involving a matrix argument hypergeometric function as obtained by Sargan (1976a) in this special case.

While (3.44) gives us a general representation of the exact joint density function of instrumental variable estimators in simultaneous equation models, this type of series representation of the density is not as easy to interpret as we would like. It can be said that the leading term in the density reveals the order to which finite sample moments of the estimator exist [cf. Basmann (1974)]. In the present case, we see that when $L = 0$ the leading term involves $[\det(I + rr')]^{-(n+1)/2} = (1 + r'r)^{-(n+1)/2}$, which is proportional to the multivariate Cauchy density [see Johnson and Kotz (1972)]; when $L > 0$ the term involves $[\det(I + rr')]^{-(L+n+1)/2} = (1 + r'r)^{-(L+n+1)/2}$, which is similar to a multivariate t-density. These expressions enable us to verify directly Basmann's conjecture [Basmann (1961, 1963)] that integer moments of the 2SLS estimator $(L = K_2 - n)$ exist up to the degree of overidentification. In other respects, the analytic form of (3.44) is not by itself very revealing. Moreover, series representations such as (3.44) and (3.46) cannot as yet be implemented for numerical calculations as easily as might be expected. The formulae rely on the matrix argument ${}_1F_1$ function and numerical evaluation depends on available tabulations and computer algorithms for the zonal polynomials that appear in the series representation of such matrix argument functions [see (2.11)]. This is an area in which important developments are currently taking place [some discussion and references are given in Section 2 following (2.11)]. Unfortunately, the availability of tabulations and algorithms for zonal-type polynomials[16] will cover only part of the computational difficulty. As noted by Muirhead (1978), the series that involve these polynomials often converge very slowly. This problem arises particularly when the polynomials have large arguments (large latent roots) and it becomes necessary to work deeply into the higher terms of the series in order to achieve convergence. This in turn raises additional

---

[16] This is a generic term that I am using to denote zonal polynomials and more general polynomials of this class but which may involve several argument matrices, as in the work of Davis (1980a, 1980b) and Chikuse (1981)

problems of underflow and overflow in the computer evaluations of the coefficients in the series and the polynomials themselves. To take as a simple example the case of the exact density of the IV estimator in the two endogenous variable case, the author has found that in a crude summation of the double infinite series for the density a thousand or more terms seem to be necessary to achieve adequate convergence when the true coefficient [that is, β in (3.45)] is greater than 5 and the concentration parameter, μ^2, is greater than 10. These are not in any way unrealistic values and the problems increase with the size of the coefficient and concentration parameter. When the density is expressed as a single series involving the $_1F_1$ function of a scalar argument, as in (3.45), these considerations necessitate the computation of the $_1F_1$ function for scalar arguments greater than 225. Use of the conventional asymptotic expansion of the $_1F_1$ function [which is normally recommended when the argument is greater than 10, see Slater (1965)] fails here because one of the parameters of the $_1F_1$ function grows as we enter more deeply into the series and the series itself no longer converges. Undoubtedly, the additional problems encountered in this example quickly become much worse as the dimension of the argument matrices in the special functions and the zonal polynomials increases and as we need to make use of the more general zonal-type polynomials (see footnote 16).

For direct computational work in the case of the IV estimator when there are more than two endogenous variables in the structural equation, the problems reported in the previous section were overcome in Phillips (1980a) by extracting an asymptotic expansion of the exact joint density of the vector coefficient estimator. This involves the use of a multidimensional version of Laplace's method of approximating integrals [see, for example, Bleistein and Handelsman (1976)]. Marginal density expansions were obtained by similar techniques in Phillips (1982a). These results give us direct and readily computable formulae for the joint and marginal densities of the coefficient estimator. The leading terms of these expansions of the joint and marginal densities have an error of $O(T^{-1})$, where T is the sample size and in the univariate (two endogenous variable) case the resulting approximation can be otherwise obtained by the saddlepoint technique as in Holly and Phillips (1979). The latter article demonstrates that the approximation gives high accuracy for some plausible values of the parameters throughout a wide domain of the distribution, including the tails.

The main conclusions about the shape and sensitivity of the p.d.f. of β_{IV} and its components which emerge from the computational work in these articles confirm the results of earlier numerical investigations dealing with the two endogenous variable case by Sawa (1969) and Anderson and Sawa (1979) and the recent experimental investigations by Richardson and Rohr (1982). A full discussion of the two endogenous variable case will be taken up in Section 3.6. In what follows we report briefly the principal results which apply in the multi-endogenous variable cases investigated by Phillips (1980a, 1982a).

(1)   For comparable parameter values the marginal distributions of β_{IV} appear to concentrate more slowly as $T \to \infty$ when the number of endogenous variables $(n + 1)$ in the equation increases.

(2)   The marginal densities are particularly sensitive to the degree of correlation in the concentration parameter matrix $\overline{M} = T \overline{\Pi}_{22}' \overline{\Pi}_{22}$ in (3.44) Setting, for example,

$$\overline{M} = \mu^2 \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$$

in the $n + 1 = 3$ endogenous variable case, the location, dispersion, and skewness of the marginal distributions all seem to be sensitive to ρ. Since $\overline{M}$ approaches singularity as $|\rho| \to 1$ when the equation becomes unidentifiable [Π_{22} in (3.5) and hence $\overline{\Pi}_{22}$ must be of full rank $= n$ for identifiability of the equation] we would expect the dispersion of the marginal distributions of the structural estimator β_{IV} to increase with $|\rho|$. This phenomenon is, in fact, observed in the graphical plots recorded by Phillips (1980a, 1982a) for different values of ρ. The central tendencies of the marginal distributions also seem to be sensitive to the relative signs of ρ and the elements of the true coefficient vector β. We give the following example. When the coefficients β_i and ρ all have the same sign the common set of exogenous variables are compatible as instruments for Y_2 in the regression and the marginal distributions appear to be adequately centered (for small values of L and moderate μ^2); but when β_i and ρ take opposite signs the exogenous variables are less compatible as instruments for the columns of Y_2 and the marginal distributions become less well centered about the true coefficients.

(3)   The effect of increasing the number of endogenous variables, *ceteris paribus*, in a structural equation is a decrease in the precision of estimation. This accords with well-known results for the classical regression model.

(4)   The marginal distribution of β_{IV} displays more bias in finite samples as L, the number of additional instruments used for the n right-hand-side endogenous variables, increases in value. When L becomes small the distribution is more centrally located about the true value of the parameter but also has greater dispersion than when L is large.

3.6.   The case of two endogenous variables (n = 1)

As seen in (3.45) the general form of the joint density (3.44) can be specialized to yield results which apply in the two endogenous variable case. These results were

first established independently by Richardson (1968) and Sawa (1969) for 2SLS and OLS [to which (3.45) applies], by Mariano and Sawa (1972) for LIML, and by Anderson and Sawa (1973) for k-class estimators. Moreover, as demonstrated by Richardson and Wu (1970) and by Anderson (1976) the exact p.d.f.s for 2SLS and LIML directly apply after appropriate changes in notation to the OLS and orthogonal regression estimators of the slope coefficient in the errors in variables model.

Details of the argument leading to the exact density of the 2SLS (or OLS) estimator can be outlined in a few simple steps arising from (3.42) [see Phillips (1982e) for details]. The final result is expression (3.45), obtained above as a specialized case of the more general result in Section 3.5. Expression (3.45) gives the density of β_{2SLS} when $L = K_2 - 1$ and the density of β_{OLS} when $L = T - K_1 - 1$. An alternative method of deriving the density of β_{2SLS} (or β_{OLS}) is given in Phillips (1980b, appendix A), where the Fourier inversion [of the form (2.3)] that yields the density is performed by contour integration.

Similar methods can be used to derive the exact densities of the LIML and k-class estimators, β_{LIML} and $\beta_{(k)}$. In the case of LIML the analysis proceeds as for the leading case but now the joint density of sample moments is non-central [see Phillips (1982e) for details]. This joint density is the product of independent Wishart densities with different degrees of freedom (K_2 and $T - K$, respectively) and a non-centrality parameter matrix closely related to that which applies in the case of the IV estimator analyzed in Section 3.5. The parameterization of the joint density of the sample moments upon which β_{LIML} depends clarifies the key parameters that ultimately influence the shape of the LIML density. These are the (two) degrees of freedom, the non-centrality matrix, and the true coefficient vector. For an equation with two endogenous variables the relevant parameters of the LIML density are then: K_2, $T - K$, μ^2, and β. The mathematical form of the density was first derived for this case by Mariano and Sawa (1972).[17] The parameterization of the LIML density is different from that of the IV density given above. In particular, the relevant parameters of (3.45) are L, μ^2, and β; or in the case of 2SLS, K_2, μ^2, and β. We may note that the IV density depends on the sample size T only through the concentration parameter μ^2, as distinct from the LIML density which depends on the sample size through the degrees of freedom, $T - K$, of one of the underlying Wishart matrices as well as the concentration parameter.

Similar considerations apply with respect to the distribution of the k-class estimator, $\beta_{(k)}$. We see from (3.17) that for $k \neq 0, 1$ the p.d.f. of $\beta_{(k)}$ depends on the joint density of two underlying Wishart matrices. The relevant parameters of the p.d.f. of $\beta_{(k)}$ are then: K_2, $T - K$, k, μ^2, and β. The mathematical form of this

---

[17]See Mariano and McDonald (1979) for a small correction

density for $0 \leqslant k \leqslant 1$ was found by Anderson and Sawa (1973) as a fourth-order infinite series.

Extensive computations are now available for at least some of the exact densities (and associated distribution functions) discussed in this section. Most of this work is due to a series of substantial contributions by T. W. Anderson, T. Sawa, and their associates. An excellent account of their work is contained in Anderson (1982). We summarize below the main features that emerge from their numerical tabulations of the relevant distributions, all of which refer to the two endogenous variable case.

(1) The distribution of β_{2SLS} is asymmetric about the true parameter value, except when $\beta = 0$ [the latter special case is also evident directly from expression (3.45) above]. The asymmetry and skewness of the distribution increase as both β and K_2 increase. For example, when $\beta = 1$, $\mu^2 = 100$, and $K_2 = 30$ the median of the distribution is -1.6 (asymptotic) standard deviations from the true parameter value, whereas at $K_3 = 3$ the median is -0.14 standard deviations from β. As K_2 becomes small the distribution becomes better located about β (as the numbers just given illustrate) but displays greater dispersion. Thus, at $\beta = 1$, $\mu^2 = 100$, and $K_2 = 30$ the interquartile range (measured again in terms of asymptotic standard deviations) is 1.031, whereas at $\beta = 1$, $\mu^2 = 100$, and $K_2 = 3$ the interquartile range is 1.321. Table 3.1 table illustrates how these effects are magnified as β increases:[18]

Table 3.1
Median (MDN) and interquartile range (IQR)
of $\beta_{2SLS} - \beta$ in terms of asymptotic
standard deviations ($\mu^2 = 100$)

K_2	β	1	2	5
3	MDN	−0.140	−0.177	−0.194
	IQR	1.321	1.310	1.304
30	MDN	−1.599	−2 021	−2.215
	IQR	1.031	0.924	0 860

(2) The rate at which the distribution of β_{2SLS} (appropriately centered and standardized) approaches normality depends critically on the values of β and K_2. If either (or both) of these parameters are large, then the approach to normality is quite slow. At $\beta = 1$ and $K_2 = 3$, for example, the value of μ^2 must be at least 100 to hold the maximum error on the asymptotic normal approximation to 0.05; but

---

[18] The numbers in Tables 3 1 and 3.2 have been selected from the extensive tabulations in Anderson and Sawa (1977, 1979) which are recommended to the reader for careful study My thanks to Professors Anderson and Sawa for their permission to quote from their tables

when $K_2 = 10$, μ^2 must be at least 3000 to ensure the same maximum error on the asymptotic distribution.

(3) Since the exact distribution of β_{LIML} involves a triple infinite series, Anderson and Sawa (1977, 1979) tabulated the distribution of a closely related estimator known as LIMLK. This estimator represents what the LIML estimator would be if the covariance matrix of the reduced-form errors were known. In terms of (3.18), β_{LIMLK} minimizes the ratio $\beta_\Delta' W \beta_\Delta / \beta_\Delta' \Omega \beta_\Delta$, where Ω is the reduced-form error covariance matrix and satisfies the system $(W - \lambda_m \Omega)\beta_\Delta = 0$, where λ_m is the smallest latent root of $\Omega^{-1} W$. The exact distribution of β_{LIMLK} can be obtained from the non-central Wishart distribution of W. Anderson and Sawa (1975) give this distribution in the form of a double infinite series that is more amenable to numerical computation than the exact distribution of LIML. In a sampling experiment Anderson et al. (1980) investigated the difference between the LIML and LIMLK distributions and found this difference to be very small except for large values of K_2. Anderson (1977) also showed that expansions of the two distributions are equivalent up to terms of $O(\mu^{-3})$. These considerations led Anderson and Sawa to take LIMLK and a proxy for LIML in analyzing the small sample properties of the latter and in the comparison with 2SLS. They found the central location of LIMLK to be superior to that of 2SLS. In fact, LIMLK is median unbiased for all β and K_2. Moreover, its distribution (appropriately centered and standardized) approaches normality much faster than that of 2SLS. However, LIMLK displays greater dispersion in general than 2SLS and its distribution function approaches unity quite slowly. These latter properties result from the fact that LIMLK, like LIML, has no integral moments regardless of the sample size and its distribution can therefore be expected to have thicker tails than those of 2SLS. Table 3.2 [selected computations from Anderson and Sawa (1979)] illustrates these effects in relation to the corresponding results for 2SLS in Table 3.1.[19]

Table 3.2
Median and interquartile range of $\beta_{\text{LIMLK}} - \beta$
in terms of asymptotic standard deviations ($\mu^2 = 100$)

K_2	β	1	2	5
3	MDN	0	0	0
	IQR	1 360	1.357	1 356
30	MDN	0	0	0
	IQR	1.450	1.394	1.363

[19]We note that since β_{LIMLK} depends only on the non-central Wishart matrix W with degrees of freedom K_2, the distribution of β_{LIMLK} depends on the sample size T only through the concentration parameter μ^2, unlike the distribution of β_{LIML}.

These features of the exact small sample distributions of 2SLS and LIMLK give rise to the following two conclusions reported by Anderson (1982): (a) the distribution of β_{2SLS} may be badly located and skewed unless β and K_2 are small or μ^2 is very large; and (b) the approach to the asymptotic normal distribution is slow for 2SLS and rapid for LIMLK and, apparently, LIML. Thus, in many cases the asymptotic normal theory may be a fairly adequate approximation to the actual distribution of LIML but a less than adequate approximation to the distribution of 2SLS.

These conclusions clearly suggest the use of caution in the application of asymptotic theory and thereby agree with the results of many other studies. One additional point is worthy of mention. The above exact results and reported numerical experience refer to the standardized model as discussed in Section 3.3. When we referred to the true coefficient β above, we therefore meant the true standardized coefficient [as given by β^* in expression (3.25) of Theorem 3.3.1]. But we note that the correlation between the included endogenous regressor, y_{2t}, and the structural error, u_t, in the unstandardized model is a simple function of β^*, namely $\mathrm{corr}(y_t, u_t) = -\beta^*/(1+\beta^{*2})^{1/2}$ as given by (3.33) in the general case. Thus, as the modulus of the standardized coefficient, $|\beta^*|$, increases, the correlation between y_{2t} and u_t increases. We therefore need, *ceteris paribus*, a larger sample of data to effectively purge y_{2t} of its correlation with u_t in estimation by 2SLS (or more generally IV). This correlation is explicitly taken into account when we estimate by LIMLK (or LIML), since we directly utilize the reduced-form error covariance matrix (or an estimate of it) in this procedure. Thus, it may not be too surprising that the finite sample distribution of β_{2SLS} displays a far greater sensitivity to the value (particularly large values) of β^* than does the distribution of LIML, as the computations in Tables 3.1 and 3.2 illustrate.

3.7. Structural variance estimators

In Sections 3.3–3.6 our attention has focused on the distribution of structural coefficient estimators. Structural variance estimators are also of importance, both as measures of residual variation and as components in commonly used test statistics (such as coefficient significance tests of the t ratio and asymptotic χ^2 variety where the metric relies on an estimate of the structural equation error variance). Basmann (1974) has pointed to an additional role that structural variance estimators may play by indicating the demands for accuracy which a model such as (3.3) and (3.5) may place on the measurement of the data.

Structural error variance estimators typically rely on the residuals from an estimated structural equation and their distributions rely, in turn, on those of the structural coefficient estimators. The following quadratic forms define three

alternative classes of estimator for the structural variance, σ^2, of the errors in (3.3):

$$G_1(\beta) = \beta'_\Delta X Q_{Z_1} X \beta_\Delta = \beta'_\Delta A(I) \beta_\Delta, \tag{3.47}$$

$$G_2(\beta) = \beta'_\Delta X' Q_Z X \beta_\Delta = \beta'_\Delta [A(I) - A(P_Z)] \beta_\Delta, \tag{3.48}$$

$$Q(\beta) = G_1(\beta) - G_2(\beta) = \beta'_\Delta X'(P_Z - P_{Z_1}) X \beta_\Delta = \beta'_\Delta A(P_Z) \beta_\Delta.^{[20]} \tag{3.49}$$

Corresponding to (3.47) we have the most common structural error variance estimators, namely

$$\sigma_{IV} = T^{-1} G_1(\beta_{IV}) = T^{-1}(y_1 - Y_2 \beta_{IV})' Q_{Z_1}(y_1 - Y_2 \beta_{IV})$$
$$= T^{-1}(y_1 - Y_2 \beta_{IV} - Z_1 \gamma_{IV})'(y_1 - Y_2 \beta_{IV} - Z_1 \gamma_{IV}) \tag{3.50}$$

and

$$\sigma_{LIML} = T^{-1} G_1(\beta_{LIML}) = T^{-1}(y_1 - Y_2 \beta_{LIML})' Q_{Z_1}(y_1 - Y_2 \beta_{LIML})$$
$$= T^{-1}(y_1 - Y_2 \beta_{LIML} - Z_1 \gamma_{LIML})'(y_1 - Y_2 \beta_{LIML} - Z_1 \gamma_{LIML}) \tag{3.51}$$

for estimation by IV and LIML, respectively.

The exact distributions of σ_{IV} and σ_{LIML} [as well as those of the related estimators based on $G_2(\beta)$ and $Q(\beta)$] can be extracted in steps that reduce the dimensionality of the problem from the essential sample moments and coefficient estimators in (3.50)–(3.51) to the required marginal densities. The analytic derivations relate very closely to those that apply in the distribution theory for the relevant coefficient estimator. The published work on the distribution of structural variance estimators refers to the two endogenous variable case and, in conjunction with other important unpublished material, has been well reviewed by Basmann (1974, pp. 252–254). The essential contributions in this area are by Basmann and Richardson (1969, 1973), who found the exact p.d.f.s of 2SLS variance estimators based on (3.47)–(3.49), and by McDonald (1972), who found the exact p.d.f.s of the LIML variance estimators for the same trinity of quadratic forms. Some of their principal findings can otherwise be obtained by examination of the leading case considered above in Section 3.4.

---

[20] In the case of estimation by IV (with instrument matrix H) it will sometimes be more appropriate to consider the following quadratic form instead of (3 49)

$$Q(\beta) = \beta'_\Delta X'(P_H - P_{Z_1}) X \beta_\Delta = \beta'_\Delta A(P_H) \beta_\Delta$$

In particular, the exact density of σ_{IV} in standardized form and in the leading case is given by [see Phillips (1982e) for derivations]

$$\mathrm{pdf}(\sigma_{\mathrm{IV}}) = \frac{TB\left(\dfrac{1}{2}, \dfrac{T-K_1+K_3}{2}\right)e^{-T\sigma_{\mathrm{IV}}/2}}{2\Gamma\left(\dfrac{T-K_1}{2}\right)B\left(\dfrac{1}{2}, \dfrac{K_3}{2}\right)} \sum_{j=0}^{\infty} \frac{\left(\dfrac{1}{2}\right)_j \left(\dfrac{T\sigma_{\mathrm{IV}}}{2}\right)^{(T-K_1)/2+j-1}}{\left(\dfrac{T-K_1+K_3+1}{2}\right)_j j!}.$$

$$(3.52)$$

Expression (3.52) gives the density of σ_{IV}, where $K_1 + K_3$ is the number of instrumental variables used in the estimation of the equation. When $K_3 = 1$ this corresponds to the just identified case and also to LIML. The latter follows because in the leading case the density of β_{LIML} is Cauchy, as shown in (3.39), and this is just the special case of (3.38) when $K_3 = 1$.

Analysis of (3.52) shows that moments, $E(\sigma_{\mathrm{IV}}^h)$, of σ_{IV} are finite provided $h < K_3/2$. In the case of 2SLS, where $K_3 = K_2$, this corresponds to the results reported by Basmann (1974). And by setting $K_3 = 1$, we deduce that σ_{LIML} has no finite moments of integral order, as shown by McDonald (1972). In this connection we may also note that since $\beta_{2\mathrm{SLS}}$ minimizes $Q(\beta)$ in (3.49) and since $Q(\beta)$ is proportional to a $\chi^2_{K_2}$ variate when β takes on the true value of the coefficient vector, the structural variance estimator $\hat{\sigma}_{2\mathrm{SLS}} = Q(\beta_{2\mathrm{SLS}})/(K_2 - n)$ possesses finite moments of all orders. However, $\hat{\sigma}_{2\mathrm{SLS}}$ (unlike σ_{IV} and σ_{LIML}) is in general an inconsistent estimator of the structural error variance. In fact, Basmann and Richardson (1973) show that the statistic $\hat{\sigma}_{2\mathrm{SLS}}$ has a limiting $\chi^2_{K_2-1}$ distribution as the concentration parameter $\mu^2 \to \infty$. The errors involved in this asymptotic distribution of $\hat{\sigma}_{2\mathrm{SLS}}$ were analyzed for the two endogenous variable case by Ebbeler and McDonald (1973) who found that the errors increased with the size of β^2 and with the number of excluded exogenous variables, K_2.

3.8. Test statistics

The finite sample distributions of certain test statistics as well as structural coefficient and variance estimators have also received attention in the literature. As with the classical linear regression model, knowledge of the distribution of test criteria allows us in principle to construct exact confidence intervals and to carry out significance tests for which the size is exact (or the critical region appropriate for a given test size). However, an important practical difference arises in this context between the classical regression model and the SEM. In the former, the usual t-ratio, F, and χ^2 tests are parameterized only by degrees of freedom which are available upon simple data, parameter, and restriction counts. In the latter,

however, most finite sample distributions depend on a number of parameters some of which figure prominently in the parameterization of the structural and reduced forms. This is evident from the results reported above in sections that pertain to estimators rather than test criteria. It prevents, in particular, our using those results directly to mount significance tests on the coefficients in just the same way as the presence of an unknown error variance prevents our using a normal theory as the basis of an exact significance test for a coefficient in a linear regression. Whereas this problem is simply overcome in the regression model by the use of the t-ratio, it is not so simply resolved in the SEM. Unless we can assume values for the unknown parameters upon which the relevant distribution depends, an exact theory will in most cases be beyond reach.

Two different approaches have been adopted in the literature to assist in resolving this difficulty. The first of these is to develop an Edgeworth expansion of the distribution of the test statistic, then to replace the unknown parameters that arise in the coefficients of this expansion by (consistent) estimates of them. Thus, if a symmetric confidence interval for a parameter based on crude asymptotic theory is corrected by the $O(T^{-1})$ terms in the Edgeworth expansion and the parameters, say ψ, that arise in these terms are replaced by consistent estimates $\hat{\psi}$ for which $\hat{\psi} = \psi + O_p(T^{-1/2})$ the order of magnitude of the error in the Edgeworth correction will be maintained. This approach forms the basis of the work by Sargan (1975, 1976a, 1980) and will be considered more closely in Chapter 15 of this Handbook by Rothenberg.

The second approach to the problem, at least in a sampling theoretic framework, is to use test criteria whose distributions are more parsimoniously parameterized and therefore more useful in providing revisions to asymptotic significance levels and critical regions. The published work in this area is less general than the literature which deals with Edgeworth corrections and the associated distribution theory is more limited than that which has been developed for structural estimators. Nevertheless, some interesting and important results have been obtained which we will now briefly review.

3.8.1. Anderson–Rubin exact confidence intervals and tests

A small sample theory of interval estimation and testing in structural equations such as (3.3) was developed by Anderson and Rubin (1949).[21] Their procedure is applicable when the confidence region or null hypothesis concerns the full vector of endogenous variable coefficients in the structural equation. Thus, if we consider the hypothesis H_0: $\beta = \beta_0$ in (3.3), we may define $y^* = y_1 - Y_2\beta_0$ and rewrite the structural equation under the null hypothesis as $y^* = Z_1\gamma + u$. On the

---

[21]As pointed out by Anderson and Rubin (1949, p 61) their method was independently suggested by Bartlett (1948)

other hand, when H_0 is not true, y^* will [in view of the reduced form (3.5)] be a linear function of both Z_1 and Z_2. Thus, H_0 may be tested by a conventional F-test of the hypothesis that the coefficient vector of Z_2 is zero in the regression of y^* on Z_1 and Z_2. The statistic for this test takes the usual form of

$$F = \frac{T-K}{K_2} \frac{y^{*\prime}(Q_{Z_1} - Q_Z)y^*}{y^{*\prime}Q_Z y^*} \tag{3.53}$$

and has an $F_{K_2, T-K}$ distribution under H_0. When H_0 is false, the denominator of (3.53) is still proportional to a χ^2_{T-K} variate while the numerator becomes non-central $\chi^2_{K_2}$ with the non-centrality dependent upon the vector of coefficient inaccuracy under the null, $\beta - \beta_0$, and a subset of the reduced-form parameters. Thus (3.53) is non-central $F_{K_2, T-K}$ under the alternative hypothesis, $\beta \neq \beta_0$. This test can readily be extended to accommodate hypotheses that involve exogenous variable coefficients and even (under suitable conditions) coefficients from several structural equations. The common requirement in each version of the test is that all structural coefficients pertaining to endogenous variables be specified under the null. This requirement ensures that the model can be rewritten, as above, as a multiple (or multivariate) regression when the null hypothesis holds. The test based on (3.53) is consistent and its power function was considered briefly by Revankar and Mallela (1972). Confidence regions follow from (3.53) in the usual way as the set of all β satisfying the inequality

$$\frac{(y_1 - Y_2\beta)'(Q_{Z_1} - Q_Z)(y_1 - Y_2\beta)}{(y_1 - Y_2\beta)'Q_Z(y - Y_2\beta)} \leqslant \frac{K_2}{T-K} F_{K_2, T-K}(\alpha) \tag{3.54}$$

at the $100(1-\alpha)$ percent confidence level.

3.8.2. An exact structural t-statistic

Richardson and Rohr (1971) studied a structural t-statistic [introduced by Dhrymes (1969)] that can be used to test hypotheses and construct confidence intervals for individual structural coefficients in SEMs. They found the exact distribution of this statistic for an equation with two endogenous variables and showed: (a) that this distribution is, indeed, Student's t-distribution when the true standardized structural coefficient (β) is zero; and (b) that this distribution tends to the t-distribution as $\mu^2 \to \infty$. However, their numerical computations of this exact distribution and its first three moments indicate that the exact distribution will often be poorly approximated by the t-distribution unless β is very small or μ^2 very large. The exact density of their statistic is in fact highly skewed even for large μ^2 and small β. Exact probabilities for intervals symmetric about the origin

are lower than those for the t-distribution (unless $\beta = 0$), so that confidence levels will be overstated and levels of significance will be understated if the t-distribution is used as an approximation in constructing confidence intervals or in two-sided tests.

Their analysis can be illustrated by considering the IV estimator, β_{IV}, in the two endogenous variable case and for the leading null hypothesis $\beta = 0$, $\Pi_{22} = 0$ of Section 3.4. The Richardson-Rohr structural t-statistic is given by

$$t = \left(y_2' Z_3 Z_3 y_2 \right)^{1/2} \beta_{IV}/s, \tag{3.55}$$

where $s^2 = Q(\beta_{IV})/(K_3 - 1)$. Simple manipulations show that this has a Student's t-distribution with $K_3 - 1$ degrees of freedom. In the 2SLS case that is considered by Richardson and Rohr (1971), $K_3 - 1 = K_2 - 1 =$ degree of overidentification of the structural equation and it is assumed that $K_2 - 1 \geqslant 1$.

An interesting experimental investigation that bears on this test has been reported by Maddala (1974). Maddala studied the power functions of the Dhrymes-Richardson-Rohr (DRR) statistic, the Anderson-Rubin (AR) statistic, and the conventional t-ratio statistic (corresponding to what would be justified if the equation were a classical linear regression and estimation were by OLS). For the model and parameter values used by Maddala, he found that the DRR test had very low power in comparison with the AR and conventional test. This outcome is partially explained by Maddala in terms of the different structural variance estimators that are used in the various test statistics. He argues, in particular, that the DRR statistic involves a variance estimator based on $Q(\beta)$ in (3.49). This estimator relies on linear forms in the data such as $Z_3' X$ and does not involve the sample second moments $X'X$ directly, as do the more conventional estimators σ_{IV} and σ_{LIML} in (3.50)–(3.51). To this extent they neglect useful sample information about the error variance and this is reflected in the observed low power of the DRR test in comparison with the conventional tests.

3.8.3. Identifiability test statistics

The structural equation (3.3) may be written in the alternative form

$$y_1 = Y_2 \beta + Z_1 \gamma_1 + Z_2 \gamma_2 + u, \tag{3.56}$$

under what is known [compare Basmann (1960)] as the "identifiability hypothesis":

$$H_0 : \gamma_2 = 0. \tag{3.57}$$

It is usually assumed that $K_2 \geqslant n + 1$ so that attention is focused on the overidentifying restrictions in (3.57). Several tests of these restrictions have been suggested

in the literature and are referred to under the name of *identifiability test statistics*. The most common of these arise naturally in 2SLS and LIML estimation. Their construction relies on the quadratic forms (3.47)–(3.49) studied in connection with structural variance estimators. Explicitly we have

$$\phi_{2SLS} = \frac{G_1(\beta_{2SLS}) - G_2(\beta_{2SLS})}{G_2(\beta_{2SLS})} = \frac{Q(\beta_{2SLS})}{G_2(\beta_{2SLS})}, \tag{3.58}$$

$$\phi_{LIML} = \frac{G_1(\beta_{LIML}) - G_2(\beta_{LIML})}{G_2(\beta_{LIML})} = \frac{Q(\beta_{LIML})}{G_2(\beta_{LIML})}. \tag{3.59}$$

If the identifiability hypothesis (3.57) is valid, the limiting distributions of $T\phi_{2SLS}$ and $T\phi_{LIML}$ are both $\chi^2_{K_2-n}$ as $T \to \infty$ under fairly general conditions. These asymptotic results were derived by Anderson and Rubin (1950), Hood and Koopmans (1953), and Basmann (1960) and are reported in Chapter 7 of this Handbook by Hausman. They lead to the common critical region (of rejection) in a large sample test of identifying restrictions:

$$T\phi \geqslant \chi^2_{K_2-n}(\alpha), \tag{3.60}$$

where α is the chosen significance level and ϕ denotes either ϕ_{2SLS} or ϕ_{LIML}.

As an approximate finite sample test, Anderson and Rubin (1949) suggested the alternative critical region:

$$\frac{T-K}{K_2}\phi_{LIML} \geqslant F_{K_2, T-K}(\alpha). \tag{3.61}$$

This may be justified on the argument that for fixed β in (3.59) the ratio $(T-K)\phi/K_2$ is indeed distributed as $F_{K_2, T-K}$ [compare (3.53) above]. Basmann (1960) criticized this suggestion on the grounds that as $T \to \infty$

$$T\phi_{2SLS}, T\phi_{LIML} \xrightarrow{\mathcal{D}} \chi^2_{K_2-n},$$

whereas

$$K_2 F_{K_2, T-K} \xrightarrow{\mathcal{D}} \chi^2_{K_2}.$$

He also argued that these considerations suggested an adjustment to the numerator degrees of freedom in the F-ratio and, as a result, the alternative critical

regions:

$$\frac{T-K}{K_2-n}\phi_{2SLS} \gtreqless \Bigg\} \quad (3.62)$$

$$\frac{T-K}{K_2-n}\phi_{LIML} \gtreqless \Bigg\} F_{K_2-n, T-K}(\alpha). \quad (3.63)$$

as approximate finite sample tests of identifying restrictions. An experimental study was performed by Basmann to determine the adequacy of the new tests (*inter alia*) and his results give very favorable support to the alternative critical region for the 2SLS statistic in (3.62). An extensive series of experiments recently conducted by Rhodes and Westbrook (1982) adds support to this evidence, indicating that the adequacy of the critical regions (3.62)–(3.63) depends on the degree of overidentification of the equation under the identifiability hypothesis. In particular, the adequacy of these regions in finite samples deteriorates as $K_2 - n$ increases.

The exact finite sample distribution of ϕ_{2SLS} in (3.58) was found by Basmann (1965) for the case in which $n = 1$ and $K_2 - n = 1$ and by Richardson (1969) for $n = 1$ and an arbitrary degree of overidentification $K_2 - 1$. McDonald (1972) discovered the exact p.d.f. of ϕ_{LIML} and showed that it has finite moments of order less than $T - K$. Since ϕ_{2SLS} has finite integer moments of order less than $(T - K)/2$ [from the results of Richardson (1969)] it follows that the tails of the distribution of ϕ_{2SLS} will be thicker than those of ϕ_{LIML}. We notice that this feature of the finite sample distributions of the two statistics is consonant with the observed inequality between the statistics, namely $\phi_{LIML} \leqslant \phi_{2SLS}$, that arises from the construction of the LIML estimator [see (3.18) and (3.59)].

Some of the results in the preceding section have been generalized to the case of an equation containing $n + 1$ endogenous variables in an important contribution by Rhodes (1981). Rhodes found the exact distribution of ϕ_{LIML} and showed that it has finite moments of order less than $(n + 1)(T - K)/2$. The exact distribution depends on the non-centrality parameter matrix which we will denote by $\overline{M}$, as in Section 3.5 above. In our notation:

$$\overline{M} = \tfrac{1}{2}\Sigma^{-1}E(X')(P_Z - P_{Z_1})E(X). \quad (3.64)$$

After standardizing transformations have been carried out and when the null hypothesis is correct, this becomes:

$$\overline{M} = T \begin{bmatrix} \beta' \\ \cdots \\ I_n \end{bmatrix} \Pi_{22}'\Pi_{22}\begin{bmatrix} \beta & \vdots & I_n \end{bmatrix}. \quad (3.65)$$

Thus, under the null $\overline{M}$ has one zero latent root and generally n non-zero roots.

When the null is false, the simpler form of $\overline{M}$ in (3.65) no longer holds and (3.64) normally has rank $n + 1$ rather than n. Thus, the true power functions of tests such as (3.60), (3.61), or (3.63) depend on the values of these non-zero latent roots. Rhodes (1981) investigates the actual size and power of these tests for a selected set of latent roots of $\overline{M}$ and finds that when the non-zero roots are small (less than 10) the true size of each test is very poorly represented by the nominal level of significance. To relate these results to those of Basmann (1960) reported above, Rhodes calculated the non-zero latent roots of the relevant non-centrality matrix for Basmann's experiment and found the roots to be large, explaining in part why (3.63) proved to be quite accurate in those experiments.

Since the exact distribution of ϕ_{LIML} is not amenable to computation, some steps have been taken to provide improvements on the critical regions (3.60), (3.61), and (3.63). McDonald (1974) obtained an approximate F distribution for ϕ_{LIML} by selecting parameters for the former in order that the first two moments of the distributions would be the same. Rhodes (1981) developed an alternative critical region for the test by considering the conditional distribution of ϕ_{LIML} given the other roots of the LIML determinantal equation. In particular, this conditional distribution has a simple asymptotic form as the largest n latent roots of $\overline{M}$ tend to infinity and can be used for the computation of a new critical region for a test based on ϕ_{LIML} and for power function evaluations. It has the advantage (over the conventional asymptotic and other tests we have discussed) of incorporating more sample information, and preliminary experimental results in Rhodes (1981) indicate that it may provide a more accurate critical region for the identifiability test.

3.9. Systems estimators and reduced-form coefficients

In comparison with the analytic results reviewed in previous sections for single equation estimators and test statistics, much less is known about the distribution of full systems estimators, reduced-form coefficient estimators, and their associated test statistics. Most progress has in fact been made in the application of small sample asymptotics by the use of Edgeworth expansions. Here the theory and constructive process detailed in Phillips (1982e) are directly applicable and machine programmable for both structural and reduced-form coefficient estimators. We will consider the analytic results for the two groups of coefficients separately below.

3.9.1. Structural coefficient estimators

Some manageable formulae for the first correction term of $O(T^{-1/2})$ in the Edgeworth expansion have been obtained by Sargan (1976a, appendix C) for

3SLS and FIML systems estimators. But no work is presently available to shed light on the adequacy of these approximations. What we know of their performance in the case of single equation estimators[22] suggests that their adequacy (at least for 3SLS estimation) will deteriorate as certain equations in the system become heavily overidentified. It also seems clear that the size of the system will have an important bearing in this respect, given other relevant factors such as the sample size, reduced-form parameter values, and features of the exogenous series. Some evidence which relates to this issue is available in Phillips (1977c), who developed formulae for the Edgeworth expansion of two-stage Aitken estimators of the parameters in a linear multivariable system subject to general linear cross-equation restrictions.[23] These formulae show that to terms of $O(T^{-1})$ the finite sample distribution is a rescaled version of the exact distribution of the Aitken estimator. This scaling factor depends on the moments of the estimated error covariance matrix and the sample second moments of the exogenous variables. As the number of equations in the system increases, the scale generally changes in such a way that the dispersion of the distribution increases. This corresponds with exact results obtained by Kataoka (1974) for a somewhat simpler version of this model and squares with the intuition that as the precision of our error covariance estimator decreases (through reductions in the effective degrees of freedom) the sampling dispersion of the resulting two-stage coefficient estimator increases. These results for the multivariate linear model furnish interesting conjectures for systems estimation in the SEM. Finally in this connection, we may mention that Nagar-type approximating moments may be deduced from the Edgeworth formulae [see Phillips (1982e)]. Such approximating moments, or pseudo-moments (where this term is appropriate), were derived independently for the 3SLS structural coefficient estimator by Mikhail (1969) in doctoral dissertation work at the London School of Economics.

In addition to the approximate distribution theory discussed above some progress on a leading case analysis for systems estimation along the lines of Section 3.4 is possible. The principles may be illustrated by considering FIML applied to a two-equation system of the form (3.1) with

$$B = \begin{bmatrix} 1 & b_{21} \\ b_{12} & 1 \end{bmatrix}, \tag{3.66}$$

and overidentifying restrictions imposed on each column of the exogenous variable coefficient matrix C. We may consider the null hypothesis in which

[22]See Anderson and Sawa (1979). Holly and Phillips (1979), and Richardson and Rohr (1981) An attempt to tackle this problem by asymptotic expansions in which the degree of overidentification grows large is given by Morimune (1981)

[23]Recent work in the same framework has been published by Maekawa (1980) for *t* ratio type test statistics

$C = 0$ and hence $\Pi = 0$ in the reduced form (3.2). In this case, it is shown in Phillips (1982e) that the joint density of the unrestricted coefficient estimates $(b_{12}^{\text{FIML}}, b_{21}^{\text{FIML}})$ is bivariate Cauchy. This result confirms that the FIML estimator of the structural coefficients (b_{12}, b_{21}) has no finite integral moments. Sargan (1970) originally established the latter result by using the fact that the FIML estimator is independent of the normalization of the structural equation. In the context of (3.66) this argument takes the form that if the implied normalization $(b_{11}, b_{12}) = (1, b_{12})$ were changed to $(b_{11}, 1)$, then the FIML estimates under the alternative normalizations would satisfy the reciprocal relationship $b_{11}^{\text{FIML}} = 1/b_{12}^{\text{FIML}}$. Thus, the FIML estimate of a structural coefficient can be interpreted as the reciprocal of another FIML estimate under a different normalization. This fact would normally imply that the distribution of such an estimator has no integral moments. As in the case of LIML (see Section 3.4) this property of the exact distribution of FIML estimates of the structural coefficients (b_{12}, b_{21}) means that the probability of extreme outliers is generally higher for FIML than for other structural coefficient estimators.

Moments of the 3SLS structural coefficient estimator have been investigated by Sargan (1978). Some difficulty occurs in the treatment of the 2SLS estimated error covariance matrix, $\Sigma_{2\text{SLS}}$, arising in the conventional 3SLS formula. Sargan, therefore, considers two cases. The first case treats $\Sigma_{2\text{SLS}}$ as non-random or, more generally, allows $\Sigma_{2\text{SLS}}$ to be random but bounds the ratio of its largest and smallest latent roots. For this case, Sargan demonstrates that the 3SLS estimator of the coefficients in any equation has finite moments of integral order up to (and including) the degree of overidentification for that equation. Thus, for this case it is proved that 2SLS and 3SLS estimators have finite moments to the same integral order. The second case considered by Sargan allows $\Sigma_{2\text{SLS}}$ to be the conventional estimator of the error covariance matrix. Here it is proved that moments of the 3SLS estimator will be finite provided the order of the moment is less than $(N + 1)/2$, where N is the degree of overidentification. Thus, the mean and variance of the 3SLS estimator will certainly be finite if the degree of overidentification is two and four, respectively. These are sufficient conditions and Sargan conjectures that the earlier result for $\Sigma_{2\text{SLS}}$ non-random also applies in this second case where $\Sigma_{2\text{SLS}}$ is the conventional error covariance matrix estimator.

3.9.2.   Reduced-form coefficients

In an important article, McCarthy (1972) initiated the analytic study of the finite sample properties of restricted reduced-form (RRF) coefficient estimators and associated predictions. The RRF incorporates additional information that is embodied in overidentifying restrictions on the structural equations of the system. To the extent that RRF estimators utilize this information, they were thought for many years to possess higher asymptotic efficiency and, as a result, smaller

variances than unrestricted reduced-form (URF) coefficient estimators.[24] McCarthy demonstrated that if there are overidentifying restrictions on the structural equations the solved reduced-form coefficients from 2SLS will in general possess no integral moments. This property influences the probability in small samples of outliers in the 2SLS reduced-form coefficients and associated predictions. It warns us that RRF estimators may give rise to occasional very poor forecasts and cautions against the use of quadratic loss criteria in estimator evaluations and in Monte Carlo work. Since the publication of McCarthy's article, generalizations of these results have been made to estimators other than 2SLS and various reduced-form estimator modifications have been suggested which attempt to improve on small sample performance. Much of this work is contained in a fundamental paper by Sargan (1976b) and in the doctoral dissertation research of Maasoumi (1977). Regrettably, a good deal of their work has not yet been published. However, with the permission of these authors some of their unpublished results will briefly be reported within the general discussion that follows.

An intuitive explanation for the McCarthy result and its generalizations arises from the transformation which takes the structural system (3.1) into its reduced form (3.2), namely $\Pi = -CB^{-1}$. For structural coefficient estimators $(\hat{B}, \hat{C})$ we deduce the reduced-form estimator $\hat{\Pi} = -\hat{C}(\text{adj}\,\hat{B})/(\det\hat{B})$. Now if there exists a value of $(\hat{B}, \hat{C})$ for which $\det\hat{B} = 0$ while $\hat{C}(\text{adj}\,\hat{B}) \neq 0$ and $\text{pdf}(\hat{B}, \hat{C}) > 0$, then at least some elements of the reduced-form estimator $\hat{\Pi}$ will have no integral moments. This follows because the integral that defines the first moment, namely $\int \hat{\Pi}\,\text{pdf}(\hat{B}, \hat{C})\,d\hat{B}\,d\hat{C}$, does not converge under the stated conditions (just as the integral $\int_a^b |x - a|^{-s}\,dx$ diverges for all values of s in the interval $1 \leq s < \infty$). This intuitive argument underlies the mathematical proof of the 2SLS result by McCarthy (1972) and forms the basis of the following general result first given (in a modified form) by Sargan (1976b) and proved in Phillips (1982e).

Theorem 3.9.1

If $\hat{\beta} = \psi(p)/\phi(p)$, where p is a random n-vector and $\hat{\beta}$ is a scalar function of p and there exists a p_0 in the domain of definition of p such that:

(i) $\psi(p)$ is continuous at p_0 with $\psi(p_0) \neq 0$,

(ii) $\phi(p)$ has continuous first derivatives at p_0, denoted by the vector ϕ_p, for which $\phi_p'\phi_p > 0$ and $\phi(p_0) = 0$,

(iii) p has a continuous p.d.f. with $\text{pdf}(p_0) > 0$, then $\hat{\beta}$ has no integral moments.

□

This theorem applies readily to a wide variety of reduced-form estimators of the type $\hat{\Pi} = -\hat{C}(\text{adj}\,\hat{B})/\det(\hat{B})$ considered above. Its conclusion gives some

---

[24] Dhrymes (1973) showed that this ranking in terms of asymptotic efficiency does not hold for RRF estimators, such as 2SLS, which are not fully efficient

general analytical support for the presumption that reduced-form coefficients extracted from OLS, 2SLS, and 3SLS structural estimators have no moments in overidentified models. The particular two-equation case studied by McCarthy (1972) also follows directly. Sargan (1976b) extends McCarthy's analysis to the general case. His results confirm that for most overidentified models 2SLS and 3SLS reduced-form coefficient estimators possess no integral moments. Exceptions do occur when the model is of a recursive type in which det $\hat{B}$ = const. Another important exception occurs for just identified models. Here the structural coefficient estimators for 2SLS, 3SLS, and FIML are all equal to indirect least squares and the corresponding reduced-form coefficients are equal to the OLS estimates, which have moments of all orders under normality.

We can also consider estimation of the reduced form by FIML. If θ is the vector of unconstrained elements in the structural coefficient matrices (B, C) of (3.1) then we may write $\Pi = \Pi(\theta)$ in (3.2) and θ_{FIML} is obtained by minimizing

$$\ln\left[\det\{T^{-1}(Y - Z\Pi(\theta))'(Y - Z\Pi(\theta))\}\right]. \tag{3.67}$$

From the criterion function (3.67) it is clear that low probabilistic weight will be attached to events in θ space which imply large values of Π since the latter will normally imply large values for the criterion (3.67). This will not be the case as the columns of Z become highly collinear or more generally when the complete data matrix $T^{-1}W'W$ is close to singularity. Thus, we might expect the FIML reduced-form $\Pi_{\text{FIML}} = \Pi(\theta_{\text{FIML}})$ to possess finite moments provided T is large in relation to the number of variables, $n + K$, in the system. In fact, Sargan (1976b) proves that Π_{FIML} has finite moments of integral order up to $T - n - K$.

The fact that many reduced-form estimators possess no integral moments has led to the suggestion of improved estimators which combine URF and RRF estimators in such a way that the tail behavior of the combined estimator is improved. A fundamental contribution in this area is due to Maasoumi (1978).[25] Maasoumi develops a new reduced-form estimator which combines the corresponding restricted 3SLS and the unrestricted OLS estimators. The new estimator incorporates the outcome of an asymptotic χ^2 test of the model's overidentifying restrictions and thereby opens up a middle road of methodology that lies between completely unrestricted and fully restricted estimation. Specifically, Maasoumi proposes the following estimator:

$$\Pi^* = \lambda\Pi_{3\text{SLS}} + (1 - \lambda)\Pi_{\text{OLS}}, \tag{3.68}$$

[25]More recent work by Maasoumi (1981) dealing with generic reduced forms that allow for reduced-form estimation in the light of intrinsically uncertain structural information is also pertinent to this discussion Nagar pseudo-moment expansions for 3SLS reduced-form coefficients have also been developed in Maasoumi (1977)

where

$$\lambda = \begin{cases} 1 & \text{if } \phi \leq C_\alpha, \\ \left(\dfrac{C_\alpha}{\phi}\right)^{1/2} \quad \text{or} \quad \left(\dfrac{C_\alpha}{\phi}\right) & \text{if } \phi > C_\alpha. \end{cases} \tag{3.69}$$

The weight coefficient λ depends on the outcome of a test of the overidentifying restrictions based on the statistic [see, for example, Malinvaud (1980, p. 378)]:

$$\phi = \text{tr}\{W^{-1}(\Pi_{OLS} - \Pi_{3SLS})'Z'Z(\Pi_{OLS} - \Pi_{3SLS})\}, \tag{3.70}$$

where $W = T^{-1}Y'(I - P_Z)Y$ is the usual consistent estimator of the reduced-form error covariance matrix and C_α in (3.69) is the (asymptotic) critical value of the test corresponding to a chosen significance level α.

The combined estimator Π^* can be additionally motivated by appealing to the principles of Stein–James estimation,[26] mixed regression,[27] and minimum expected loss[28] (MELO) methodologies, all of which lead to estimators which can be expressed as combinations of restricted and unrestricted estimators. To highlight the similarity of (3.68) with Stein-like procedures [and in particular the positive rule estimator proposed by Sclove (1968, 1971)] we may write Π^* as

$$\Pi^* = \Pi_{3SLS} + I_{(C_\alpha,\infty)}\left(1 - \frac{C_\alpha}{\phi}\right)(\Pi_{OLS} - \Pi_{3SLS}), \tag{3.71}$$

where $I_{(\ )}$ is an indicator function equal to unity when ϕ is in the indicated range and equal to zero otherwise. This estimator differs from the traditional Stein-like variety in that it takes the unrestricted estimate Π_{OLS} as the point of attraction rather than simply the origin.

The finite sample and asymptotic properties of the combined estimator Π^* are investigated in Maasoumi (1978).[29] It is shown that Π^* has finite integral moments to the order $T - n - K$ (as for the FIML reduced-form discussed earlier in this section) and that the limiting distribution of $\sqrt{T}(\Pi^* - \Pi)$ is close to that of $\sqrt{T}(\Pi_{3SLS} - \Pi)$ for conventional choices of the significance level C_α. Thus, Π^* has close to asymptotic equivalence with Π_{3SLS} and has apparently superior small sample properties in terms of outlier elimination. Practical implementation of the method is as straightforward as 3SLS. What remains problematic is the selection

[26]See, for example, James and Stein (1961), Zellner and Vandaele (1975), and Chapter 10 in this Handbook by Judge and Bock

[27]See Goldberger (1973)

[28]See Zellner (1978) and Zellner and Park (1979)

[29]The finite sample properties of Stein-like improved estimators in the context of the linear regression model have been studied by Ullah (1974, 1980)

of the critical level, C_α. The statistic ϕ in (3.70) has a limiting χ^2_N distribution, where N is the total number of overidentifying restrictions. Even in moderately sized models, N may be quite large and strict application of the test based on (3.70) at conventional significance levels usually leads to a rejection of the restrictions. Thus, frequent occurrence of $\Pi^* = \Pi_{OLS}$ in practical situations might be expected and this might raise the very genuine objection to the combined estimator that it will frequently result in the extreme alternative of unrestricted reduced-form estimation by OLS. This criticism should be tempered by the knowledge that the critical value, C_α, will often be a very poor (asymptotically based) indicator of the correct finite sample critical value for a test with a chosen size of α. Monte Carlo results by Basmann (1960), Byron (1974), Maddala (1974), Basmann, Richardson and Rohr (1974), Maasoumi (1977), Laitinen (1978), Meisner (1979), Hausman and McFadden (1981), and Rhodes and Westbrook (1982) all indicate that many conventional asymptotic tests of restrictions lead to an unduly high rate of rejection (that is often severe) in small sample situations. This evidence suggests that conventional asymptotic tests are often not sufficiently reliable to justify the extreme alternative of completely unrestricted reduced-form estimation. It would therefore seem wise in the light of this evidence to set the size of the test at a level much lower than usual so that the implied (asymptotic) critical value, C_α, is larger and the probability of a test rejection reduced. The problem of the most appropriate selection of C_α for a given model, data set, and limited knowledge about the exact distribution of ϕ clearly warrants substantially more attention than it has received. Mechanical correctors to the asymptotic critical region (C_α, ∞) can be based on Edgeworth expansions along the lines of Section 2.3 and this is an area of extensive current research in mathematical statistics. However, little is known at present concerning the adequacy of such corrections.

In addition to the above work on reduced forms, attention has also been given in the literature to the partially restricted reduced-form (PRRF) suggested by Amemiya (1966) and Kakwani and Court (1972). The PRRF coefficients can be obtained equation by equation from relationships such as

$$\begin{bmatrix} \pi_{11} \\ \pi_{21} \end{bmatrix} = \begin{bmatrix} \Pi_{12} \\ \Pi_{22} \end{bmatrix} \beta + \begin{bmatrix} \gamma \\ 0 \end{bmatrix} = \begin{bmatrix} \pi_{12} & I \\ \pi_{22} & 0 \end{bmatrix} \begin{bmatrix} \beta \\ \gamma \end{bmatrix} \tag{3.72}$$

[deduced from (3.6) above] which relate the reduced-form coefficients of one (relevant) equation to those of other equations in terms of the identifying restrictions. The PRRF estimator of the coefficients in the first reduced-form equation [given by the left-hand side of (3.72)] is then

$$\begin{bmatrix} \hat\pi_{11} \\ \hat\pi_{21} \end{bmatrix} = \begin{bmatrix} (Z'Z)^{-1}Z'Y & I \\ & 0 \end{bmatrix} \begin{bmatrix} \beta_{2SLS} \\ \gamma_{2SLS} \end{bmatrix}, \tag{3.73}$$

where $(Z'Z)^{-1}Z'Y_2$ is the OLS estimator of the coefficients in the reduced-form equations for the variables Y_2 appearing in the structural equation (3.3), and β_{2SLS} and γ_{2SLS} are the usual structural 2SLS estimators. The small sample properties of the estimator (3.73) and associated forecasts have been studied by Knight (1977), Nagar and Sahay (1978), Swamy and Mehta (1980), and Sahay (1981). Knight proved that this reduced-form estimator has finite moments of all orders. Nagar and Sahay found expressions for the exact bias and mean squared error of forecasts based on (3.73) in the two endogenous variable case; and Sahay (1981) has extended part of this work to the case of an equation with three endogenous variables. Their conclusions suggest that the mean squared error of forecasts based on (3.73) will be smaller than that based on unrestricted reduced-form estimation by OLS unless $\|\beta\|$ is large. These authors work with the standardized model studied in Section 3.3. Their conclusion accords with the intuitive ideas discussed in the paragraph following (3.32) that when β has large elements 2SLS estimators may display considerable dispersion in view of the correlation between the included endogenous variables and the structural error [which in the standardized model is measured by $-\beta(1 + \beta'\beta)^{-1/2}$ as in (3.32)].

3.10. *Improved estimation of structural coefficients*

The idea of developing alternative estimators which may improve on certain aspects of the small behavior behavior of the more conventional estimators has been applied to structural as well as reduced-form coefficient estimation. Here, many of the ideas are based on simple corrections to the usual formulae that are suggested by an analysis of the higher order terms [frequently terms up to $O(T^{-1})$] in series expansions of the distributions and (pseudo-) moments of the conventional estimators.[30] While these corrections are for the most part mechanical, there is an accumulating body of evidence which supports the view that their use will result in actual improvements in estimation as measured by the probability of concentration about true parameter values. In this respect, Morimune (1981) has provided a useful review of the performance characteristics of the main improved estimators.

One of the first suggestions appears in Nagar (1959) who provided (pseudo-) moment expansions for the first two moments of the k-class estimator and deduced values of k which removed the bias to $O(T^{-1})$ and minimized the determinant of the mean squared error moment matrix to $O(T^{-2})$. Zellner (1978) and Zellner and Park (1979) also developed an optimal member of the k-class

---

[30]A similar analysis of higher order terms (in Edgeworth expansions) of the distributions of conventional test statistics can be performed Much work has already been done on this topic in mathematical statistics leading to some general results on the higher order efficiency of tests based on maximum likelihood estimators. See, for example, Pfanzagl and Wefelmeyer (1978, 1979)

family in terms of the minimum expected loss (MELO) criterion whereby the posterior expectation of a weighted quadratic loss function is minimized with respect to the structural coefficients. Both Nagar and Zellner–Park reported applications of their improved estimators in the context of small macroeconometric models. Zellner and Park found in their application that the (asymptotic) standard errors of the MELO estimates were consistently smaller and often much smaller than their 2SLS counterparts.

Alternative estimators constructed by taking linear combinations of 2SLS with OLS and 2SLS with LIML were proposed by Sawa (1973a, 1973b) and Morimune (1978),[31] respectively. The weights in these combined estimators were selected so as to remove the bias (or pseudo-bias when this is appropriate) in the estimator up to terms of $O(\sigma^2)$, where σ^2 is a scalar multiple of the covariance matrix of the errors in the model. That is, the improvements were based on the use of small-σ asymptotic expansions (see footnote 5 in Section 2). Sawa (1973b) numerically computed the first two exact moments of the combined 2SLS–OLS estimator but no clear conclusion concerning its superiority over 2SLS emerged from these computations. Morimune (1978) examined the (asymptotic) mean squared error[32] of the 2SLS–LIML combination and demonstrated its superiority over LIML according to this criterion. In the context of systems estimation related work has been done by Maasoumi (1980) on a ridge-like modification to the 3SLS estimator.

Fuller (1977) introduced modifications to the LIML and fixed k-class estimators which ensure that the new estimators possess finite moments. The modifications add weight to the denominators in the matrix ratios that define the unmodified estimators. Their generic form, in the notation of Section 3.2 above, is as follows:

$$\beta_M = \left[A_{22}(I) - l\{A_{22}(I) - A_{22}(P_Z)\} \right]^{-1} \left[a_{21}(I) - l\{a_{21}(I) - a_{21}(P_Z)\} \right],$$

$$(3.74)$$

where

$$l = a\lambda + b, \qquad a, b = \text{const.,} \tag{3.75}$$

and λ is the smallest latent root of the equation:

$$\det\left[A(I) - \lambda\{A(I) - A(P_Z)\} \right] = 0 \tag{3.76}$$

as in (3.19) above. The estimator β_M in (3.74) specializes: to LIML for $a = 1$,

[31] See also Morimune and Kunitomo (1980).

[32] That is, the mean squared error of the asymptotic expansion of the distribution up to a certain order.

$b = 0$; to the fixed k-class for $a = 0$, $b =$ const.; to 2SLS for $a = 0$, $b = 1$; and to OLS for $a = 0$, $b = 0$.

Fuller's proposal is to set $a = 1$ and $b = -\alpha/(T - K)$ for some fixed real number $\alpha > 0$. For these values of a and b in (3.75), β_M is a direct modification of β_{LIML} [compare (3.20) above]. Fuller shows that when estimators of the class (3.74) are restricted to have the same bias to $O(T^{-1})$ the modified LIML estimator dominates the class according to mean squared error through to terms of $O(T^{-2})$. This result can be regarded in the light of the second-order efficiency of the maximum likelihood estimator, as has been mentioned by Rothenberg (1980). Asymptotic expansions of the densities of estimators in the class (3.74) have been explicitly derived and compared by Kunitomo (1981) who gives the same conclusion concerning the second-order efficiency of LIML. Additional small sample comparisons of estimators in the class (3.74) have been made by Morimune (1981) who suggests a variant of Fuller's modified LIML that displays superior concentration when the degree of equation overidentification is large. The numerical computations in this paper report the empirical distribution functions of the various estimators from Monte Carlo experiments with extensive replications. They indicate that the modifications to LIML thin out the tails of the distribution, as we would expect. They also confirm results recorded above (in Section 3.6) concerning the good locational properties of the LIML estimator and evaluate the performance of some new asymptotic approximations based on a large degree of equation overidentification.

3.11. Supplementary results on moments

In Sections 3.4–3.9 questions of existence of moments were dealt with in our discussion of the respective exact distributions. In most of these cases, direct formulae for the moments that exist can be extracted using term-by-term integration of the series expressions for the p.d.f.s. Direct results concerning the existence of moments and formulae for them have also appeared in the literature. The following result, which was first established in its complete form by Kinal (1980), encapsulates our knowledge concerning the existence of moments for k-class estimators.

Theorem 3.10.1 (Existence of moments for k-class estimators, $0 \le k \le 1$)

Integral moments of the estimator $\beta_{(k)}$ given by (3.17) for non-stochastic k in the structural equation (3.3) exist up to order M where

$$M = \begin{cases} T - K_1 - n & \text{for} \quad 0 \le k < 1, \\ K_2 - n & \text{for} \quad k = 1. \end{cases} \qquad \square$$

Earlier work was done by Mariano (1973) who covered the 2SLS ($k = 1$) case for even-order moments and by Hatanaka (1973) who gave sufficient conditions for existence. Sawa (1972) dealt with the two endogenous variable case, established the above result, and further demonstrated that $\beta_{(k)}$ has no integral moments when $k > 1$. Sawa also gave exact formulae for the first two moments when $0 \le k \le 1$ and developed asymptotic expansions for them in terms of the reciprocal of the concentration parameter, namely $1/\mu^2$. Similar formulae were derived by Takeuchi (1970) for OLS, 2SLS, and IV estimators in the two endogenous variable case.[33] Ullah and Nagar (1974) gave analytic formulae for the mean of the 2SLS estimator and their results were used by Sahay (1979) in finding an expression for the mean of the 2SLS structural equation residuals. Extending this work to the general single equation case (with n and K_2 arbitrary), Hillier and Srivastava (1981) and Kinal (1982) have derived exact formulae for the bias and mean squared error of the OLS and 2SLS estimator of a single endogenous variable coefficient. This generalizes the work of Sawa (1972). Unfortunately, the presence of zonal-type polynomials in the final formulae prevents their use for numerical computations in the general single equation case, at least with present-day tabulations and algorithmic machinery (see the discussion of this point in Section 3.5 above).

Before leaving this topic it may be worth mentioning that moments are useful to the extent that they shed light on the distribution itself. In particular, they provide summary information about the location, dispersion, and shape of a distribution. However, as many of the cases that are analyzed in Sections 3.5 and 3.6 attest, an important feature of many exact distributions in econometrics is their asymmetry. Obviously, moment analyses of higher order than the second are necessary to inform on such aspects of the shape of a distribution. In some cases, of course, such higher order moments may not exist. When they do, the formulae will often be as complicated as the series expressions for the p.d.f.s themselves. Considerations of research strategy therefore indicate that it may well be wise to direct most of our attention to the distributions, their numerical evaluation, and their approximation rather than that of the moments.[34]

Finally, we remark that direct results concerning the moments of estimated coefficients of the exogenous variables in a structural equation can be deduced from the relevant formulae given in Section 3.3 above and the results for the coefficients of the included endogenous variables. Thus, in the case of the IV

---

[33]As reported by Maasoumi and Phillips (1982a) there appear to be errors in his expression arising out of his formulae (2-7) and (2-8) which appear to confuse even- and odd-order moments

[34]The issues raised in this section have an obvious bearing on Monte Carlo experimentation, where it is customary to work with summary measures defined in terms of low order moments Caution in the use of such methods has been advised by several authors, for example Basmann (1961) and Maasoumi and Phillips (1982a) Problems of accurately estimating high order moments by Monte Carlo replications (and the demands this may place on the experimental design) are apposite here but seem not to have been discussed in the literature in this field.

estimator we have

$$\gamma_{IV} = T^{-1} Z_1' y_1 - T^{-1} Z_1' Y_2 \beta_{IV} \tag{3.77}$$

(assuming standardizing transformations are performed). Then $Z_1' X = Z_1'[y_1 \vdots Y_2]$ is statistically independent of β_{IV} (under error normality of course) and moments of γ_{IV} in (3.77) are defined to the same order as those of β_{IV} (see Section 3.5). Formulae for these moments can also be deduced from those that apply for β_{IV} and $Z_1' X$. Similar considerations enable us to treat the general k-class estimator of the exogenous variable coefficients.

3.12. Misspecification

Earlier results in this section have all been obtained on the presumption that the model has been correctly specified. When this is not so, the sampling distributions undergo modifications contingent upon the nature and extent of misspecification and earlier conclusions about estimator and test statistic performance in small samples no longer necessarily apply. Fisher (1961, 1966, 1967) carried out an asymptotic analysis of estimator performance in the presence of specification error consisting of incorrectly excluded variables in structural equations such as (3.3). An exact small sample theory can also be developed for this problem using the approach of Sections 3.5 and 3.6. We illustrate by considering OLS and 2SLS estimation of the (incorrectly specified) structural equation (3.3) when the true equation includes additional exogenous variables and is of the form

$$y_1 = Y_2 \beta + Z_1 \gamma_1 + Z_4 \gamma_4 + u_{\cdot} \tag{3.78}$$

We write the reduced form as

$$\left[y_1 \vdots Y_2 \right] = X = \left[Z_1 \vdots Z_4 \vdots Z_5 \right] \begin{bmatrix} \pi_{11} & \Pi_{12} \\ \pi_{41} & \Pi_{42} \\ \pi_{51} & \Pi_{52} \end{bmatrix} + \left[v_1 \vdots V_2 \right]$$

$$= \left[Z_1 \vdots Z_4 \vdots Z_5 \right] \begin{bmatrix} \Pi_1 \\ \cdots \\ \Pi_4 \\ \cdots \\ \Pi_5 \end{bmatrix} + V, \tag{3.79}$$

where the coefficients satisfy

$$\pi_{11} - \Pi_{12} \beta = \gamma_1, \qquad \pi_{41} - \Pi_{42} \beta = \gamma_4, \qquad \pi_{51} - \Pi_{52} \beta = 0. \tag{3.80}$$

We define $Z_2 = [Z_4 \vdots Z_5]$ and then

$$M = E\big[T^{-1/2}X'Z_2\big] = T^{-1/2}\Pi'Z'Z_2 = T^{1/2}\big[\Pi'_4 \vdots \Pi'_5\big]. \tag{3.81}$$

The non-centrality parameter matrix is

$$MM' = T\big[\Pi'_4\Pi_4 + \Pi'_5\Pi_5\big] = TD, \quad \text{say.} \tag{3.82}$$

We may note that this reduces to

$$\begin{bmatrix} \beta' \\ I \end{bmatrix}(\Pi'_{42}\Pi_{42} + \Pi'_{52}\Pi_{52})\big[\beta \vdots I\big] = \begin{bmatrix} \beta' \\ I \end{bmatrix}\Pi'_{22}\Pi_{22}\big[\beta \vdots I\big] \tag{3.83}$$

when $\gamma_4 = 0$ and the eq. (3.3) is correctly specified.

As in Section 3.5, OLS and 2SLS estimators of β in (3.78) depend on the matrix

$$A = T^{-1}X'FF'X, \tag{3.84}$$

with $FF' = Z_2Z'_2$ in the case of 2SLS and $FF' = I - T^{-1}Z_1Z'_1$ for OLS, where F is a $T \times f$ matrix of rank f and $F'F = TI_f$. Formulae for the exact densities of β_{OLS} and β_{2SLS} in the general case are then obtained by arguments which closely follow those of Section 3.5, as shown by Maasoumi and Phillips (1982b). For the two endogenous variable case we obtain [see Phillips (1982e) for derivations]:

$$\mathrm{pdf}(r) = \frac{\mathrm{etr}\left(\dfrac{-TD}{2}\right)}{B\left(\dfrac{1}{2},\dfrac{f}{2}\right)(1+r^2)^{(f+1)/2}} \sum_{k=0}^{\infty} \frac{\left\{\det\left(\dfrac{TD}{2}\right)\right\}^k \left(\dfrac{f+1}{2}\right)_k}{\left(\dfrac{f}{2}\right)_{2k} k!}$$

$$\times \sum_{h=0}^{\infty} \frac{\left(\dfrac{T}{2}\right)^h \left(\dfrac{f+1}{2}+k\right)_h (r^2 d_{11} + 2rd_{12} + d_{22})^h}{\left(\dfrac{f}{2}+2k\right)_h h!(1+r^2)^{k+h}}$$

$$\times {}_1F_1\left(\frac{f-1}{2}+k, \frac{f}{2}+h+2k; \frac{Tdu}{2}\right). \tag{3.85}$$

This expression gives the exact density under misspecification of β_{OLS} when $f = T - K_1$ and of β_{2SLS} when $f = K_2$. The density reduces to (3.45) when the structural equation is correctly specified ($\gamma_4 = 0$) as can be shown by rearrangement of the series.

Formula (3.85) was derived by Rhodes and Westbrook (1981)[35] and formed the basis of the computational work reported in their paper. These numerical computations provide valuable evidence concerning the practical consequences of misspecification. Two principal results emerge from their study: misspecification can substantially increase the concentration of the distribution of both OLS and 2SLS; and in some cases it may also reduce the bias (as well as the dispersion) of both estimators. These results led Rhodes and Westbrook to conclude that, when a structural equation is misspecified by incorrectly excluded variables, OLS may indeed be a superior technique of estimation to 2SLS.

The same general conclusion was reached by Hale, Mariano and Ramage (1980) who examined exact and approximate asymptotic expressions for the bias and mean squared error (MSE) of k-class estimators (for k non-stochastic in the interval $0 \le k \le 1$). Their results, which also refer to the two endogenous variable case, show that OLS is relatively insensitive to specification error and that when errors of specification are a more serious problem than simultaneity, OLS is preferable to 2SLS. Moreover, the entire k-class is dominated in terms of MSE under misspecification by either OLS or 2SLS.

Similar analysis of the effect of misspecification upon the LIML estimator in the two endogenous variable case has been performed by Mariano and Ramage (1978). Some extensions of this work, involving asymptotic expansions and moment approximations to the general single equation case, are contained in Mariano and Ramage (1979). Exact formulae for the p.d.f.s of OLS and 2SLS estimators in the general single equation case under misspecification are given by Maasoumi and Phillips (1982b).[36-39]

---

[35]Their stated result in theorem 2.1 contains a small error in that $|T_n/4|^k$ in their formula (2 11) should be replaced by $|T_n/2|^k$

[36]In addition, Knight (1981) has shown how, in the two endogenous variable case, expressions for the exact moments of k-class estimators under misspecification can be extracted from the corresponding expressions that apply in correctly specified situations.

[37]Related work on the effect of multicollinearity on the shape of the distributions of OLS and 2SLS estimators has been done by Mariano, McDonald and Tishler (1979)

[38]Hale (1979) has also studied the effects of misspecification on the two-stage Aitken estimator (2SAE) and OLS estimator in a two-equation seemingly unrelated regression model Hale's main conclusion is that the distribution of 2SAE appears to be more affected by misspecification than that of OLS

[39]Analysis of the effects of distributional shape are also possible. Knight (1981) has, in particular, found expressions for the first two exact moments of the k-class estimator in the two endogenous variable case when the reduced-form errors follow a non-normal distribution of the Edgeworth type Phillips (1980b) indicated generalizations of existing results for asymptotic expansions of coefficient estimators and test statistics under non-normal errors of this type Explicit formulae for such asymptotic expansions have in fact been derived by Satchell (1981) for the distribution of the serial correlation coefficient

4. A new approach to small sample theory

4.1. Intuitive ideas

This section outlines the elements of a new approach to small sample theory that is developed in Phillips (1982c). The idea that underlies the method in this article is very simple. It is motivated by the observation that, in spite of the complex analytic forms of many of the exact p.d.f.s presently known for econometric statistics (such as those in Section 3), when we do turn around and obtain numerical tabulations or graphical plots of the densities we typically end up with well-behaved, continuous functions that tend to zero at the limits of their domain of definition. The form of these p.d.f.s strongly suggests that we should be able to get excellent approximations to them in the class of much simpler functions and certainly without the use of multiple infinite series. We need to deal with approximating functions (or approximants as they are often called) that are capable of capturing the stylized form of a density: in particular, we want the approximant to be able to go straight for long periods in a direction almost parallel to the horizontal axis and yet still be able to bend, quite sharply if necessary, to trace out the body of the distribution wherever it is located. One class of functions that seems particularly promising in this respect, as well as being simple in form, are rational functions. Even low degree rational functions can go straight for long periods and then bend quite sharply. In this, of course, they are very different from low degree polynomials whose graphs typically display a distinct roly-poly character.

The possibility of finding rational functions which provide good global approximations to a general class of p.d.f.s is considered in Phillips (1982c). The technique developed there is based on the idea of working from local Taylor series approximations at certain points of the distribution towards a global approximation which performs well in the whole domain over which the distribution is defined and yet retains the good performance of the Taylor series approximations in the immediate locality of the points of expansion. This is, in part, achieved by the use of multiple-point Padé approximants. These Padé approximants[40] are rational functions constructed so as to preserve the local Taylor series behavior of the true p.d.f. (or d.f.) to as high an order as possible. The points selected for local expansion will often be simply the origin (in the central body of the distribution) and the tails. These local expansions can, in fact, be obtained from information about the characteristic function of the distribution so that direct knowledge of the local behavior of the true p.d.f. is not necessary for the successful application

---

[40]Pade approximants have a long tradition in mathematics and have recently been successfully applied to a large number of problems in applied mathematics and mathematical physics References to this literature may be found in Phillips (1982c).

of the technique. Local information may also be based on estimates obtained from the empirical d.f. arising in Monte Carlo simulations. Supplementary information about the distribution such as (i) its analytic form in leading cases (compare Section 3.4), (ii) knowledge of its moments where these exist, (iii) leading terms in its Edgeworth expansion (see Section 2.3), or even (iv) the crude asymptotic distribution, may all be utilized in the construction of the approximant. The final step in the method is to modify the Padé approximant so that it does display appropriate global behavior. This may involve the removal of unwanted zeros and poles which occur in the bridging region between the points of local expansion and possible modifications to ensure non-negativity in the approximant.

4.2. Rational approximation

Phillips (1982c) considers a general class of continuous marginal densities and defines the following family of potential rational function approximants:

$$R_{m,n}(r) = s(r)\frac{P_m(r)}{Q_n(r)} = s(r)\frac{a_0 + a_1 r + \ldots + a_m r^m}{b_0 + b_1 r + \ldots + b_n r^n}, \tag{4.1}$$

where m and n are even integers with $m \leq n$ and $s(r)$ is a real continuous function satisfying $s(r) > 0$ and $s(r) \rightarrow 0$ as $r \rightarrow \pm \infty$.

The coefficient function $s(r)$ in (4.1) is a vehicle by which additional information about the true density can be readily embodied in the approximant. This can be soft quantitative information, for example of the type that $\mathrm{pdf}(r) > 0$ and $\mathrm{pdf}(r) \rightarrow 0$ as $r \rightarrow \pm \infty$ [already explicit in $s(r)$]; or hard quantitative information, for example of the type (i) that $\mathrm{pdf}(r)$ has moments up to a certain order or (ii) that $\mathrm{pdf}(r)$ takes an especially simple form in an important and relevant leading case or (iii) that $\mathrm{pdf}(r)$ has a known Edgeworth series expansion up to a certain order (suitably modified to ensure that it is everywhere positive and still tends to zero at infinity).

Practical considerations frequently suggest a specialization of (4.1) to the family of rational fractions in which numerator and denominator polynomials have the same degrees (i.e. $m = n$).[41] In addition, a normalization condition is imposed on the coefficients of the polynomials in (4.1) to eliminate the redundancy that results from the multiplication of $P_n(r)$ and $Q_n(r)$ by an arbitrary constant. In density function approximation this can be simply achieved by setting $b_0 = 1$, which also ensures that the rational approximant is well behaved as

---

[41] This is discussed at some length in Phillips (1982c, 1982d)

r passes through the origin. In distribution function approximation it is convenient to set $a_n = b_n = 1$ and then, after an appropriate choice of $s(r)$ as a primitive distribution, the resulting approximant $R_{nn}(r)$ embodies desirable tail behavior as $r \to \pm \infty$.

A theory of goodness of approximation to continuous p.d.f.s based on rational approximants of the type (4.1) is developed in Phillips (1982c, 1982d). This theory uses the uniform norm

$$\| \text{pdf}(r) - R_{nn}(r) \| = \sup_{r \in (-\infty, \infty)} |\text{pdf}(r) - R_{nn}(r)| \qquad (4.2)$$

to measure the error in the approximation. Under this error norm it is shown that best uniform approximants within the family (4.1) exist and are unique for a general class of continuous p.d.f.s. Setting $m = n$ and defining $\gamma' = \{a_0, \ldots, a_n; b_1, \ldots, b_n\}$ in (4.1) means that there exists a vector γ^* and a corresponding rational fraction $R'_{nn}(r)$ for which

$$\| \text{pdf}(r) - R'_{nn}(r) \| = \inf_{\gamma} \| \text{pdf}(r) - R_{nn}(r) \| \qquad (4.3)$$

given some continuous density $\text{pdf}(r)$; and, moreover, the rational fraction $R'_{nn}(r)$ with the property (4.3) is unique. As $n \to \infty$ $R'_{nn}(r)$ converges uniformly to $\text{pdf}(r)$. Hence, an arbitrarily good approximation is possible within this family of rational functions.

Practical implementation of rational approximation requires the degree of $R_{nn}(r)$ to be prescribed, the coefficient function $s(r)$ to be selected, and the parameters of the polynomials to be specified. The problem is one of constructive functional approximation to a given distribution within the family of approximants (4.1) Operational guidelines for this constructive process are laid out in Phillips (1982c) and the final solution in any particular case will rely intimately on the information that is available about the true distribution. Typically, we will want the approximant to embody as much analytic and reliable experimental information about the distribution as possible. This will directly affect the choice of $s(r)$ and the prescribed degree of $R_{nn}(r)$. Leading case analyses such as those in Section 3.4 will often lead to a suitable choice of $s(r)$. Knowledge of the local behavior of the distribution in the body and in the tails can be used to determine the polynomial coefficients in $R_{nn}(r)$ which will then magnify or attenuate as appropriate the leading case distribution. Local information about the distribution may take the form of Taylor expansions at certain points or estimates of the function values obtained from Monte Carlo simulations. In cases where numerical or Monte Carlo integration is possible, a selected set of points within the main body and in the tails of the distribution can be used for these evaluations, which can then assist in determining the parameters of $R_{nn}(r)$. This has the advantage

of keeping the number of numerical integrations within economic limits and at the same time marrying the information from these integrations with other useful knowledge about the distribution. Constructive functional approximants of this type will have Bayesian applications in the problem of reducing multidimensional posterior distributions to manageable and readily interpreted marginal posteriors. They also provide a convenient avenue for combining Monte Carlo experimental evidence and analytic knowledge in a simple useable form.

The procedure outlined above was successfully applied in the context of approximating the exact density (3.45) of the 2SLS estimator by the rational family $R_{nm}(r)$ in (4.8) with $n = m = 4$ and $s(r)$ set equal to the leading density given in (3.38). The results obtained were very encouraging even for extremely small values of the concentration parameter μ^2, when other approximations such as those based on Edgeworth and saddlepoint methods produced very poor results. In particular, the modified two-point Padé approximant (using information at only the origin and infinity) yielded two decimal place accuracy to the exact distribution over the whole real axis and an unmodified seven-point Padé approximant gave three decimal place accuracy with a maximum error of 0.0008. Further development of these approximants seems likely on the basis of these results to produce a method that is flexible and general enough to be widely used and sufficiently accurate to be relied upon in empirical work.

4.3.   Curve fitting or constructive functional approximation?

The above discussion and reported application present a favorable picture of the strengths and potential of this new approach. An important contributory factor in this optimistic view is the flexible mathematical apparatus that underlies constructive functional approximation in the class defined by (4.1). As much analytic knowledge as is available about a distribution can be embodied in $R_{mn}(r)$ through the dual vehicles of the coefficient function $s(r)$ and the rational coefficients $(a_0, \ldots, a_m, b_1, \ldots, b_n)$. Thus, Edgeworth expansions and saddlepoint approximations are just subcases of (4.1). For if these expansions are known to yield good approximants in certain problems they themselves may be used to construct $s(r)$. Simple modifications to the Edgeworth expansion will ensure that $s(r)$ is everywhere positive, continuous, and still tends to zero as $|r| \to \infty$. Additional information about the distribution can then be incorporated in the rational coefficients and in adjustments to $s(r)$ that ensure the same tail behavior as the true distribution, where this is known by separate analytic investigation. Other choices of $s(r)$ that stem directly from analytic knowledge of the true distribution are also possible, as the example cited demonstrates. Moreover, experimental data about the distribution can be utilized in the choice of the rational coefficients by least squares or generalized least squares fitting to the

empirical distribution in place of (or in addition to) exact solutions. Thus, constructive approximants belonging to the family (4.1) can successfully embrace a wide range of different types of analytic and experimental information in a functional form that is useful for both descriptive and inferential purposes.

In a more limited mode of thinking, rational approximation *per se* is an exercise in interpolation or curve fitting. Of course, rational approximants do provide an excellent medium for such exercises, as the extensive examples in Hastings (1955) amply demonstrate, and they are already a standard technique in computer function algorithms because of their accuracy and economy [see Hart (1968)]. But such an interpretation in the present context would be myopic, ignoring as it does the extent to which the family (4.1) can build on the strengths of alternative, less flexible approximation methods and simultaneously blend analytic and experimental information from many diverse sources. It is this constructive mode of analysis that underlies the new approach and distinguishes it from more mechanical methods of asymptotic analysis and exercises in curve fitting.

5. Concluding remarks

This review began with some remarks taken from the first edition of R. A. Fisher's (1925) influential manual for practising statisticians. Fisher's keen awareness of the limitations of asymptotic theory, his emphasis on statistical tools which are appropriate in the analysis of small samples of data, and his own research on the exact sampling distributions of variance ratios and correlation coefficients contributed in significant ways to the growth of what is now an extensive literature in mathematical statistics on small sample distribution theory. The challenge of developing such a theory in models that are of interest to econometricians has produced the corpus of knowledge that forms the subject-matter of this review. Questions of the relevance of this research and its operational payoff in terms of empirical practice are as much a topic of debate in econometrics as they were (and still are to a lesser degree) in mathematical statistics.

In contrast to small sample theory, the power of asymptotic theory lies unmistakedly in the generality with which its conclusions hold, extending over a wide domain of models and assumptions that now allow for very general forms of dependent random processes, non-linear functional forms, and model misspecifications. However, the generality of this theory and the apparent robustness of many of its conclusions should not necessarily be presumed to be strengths. For the process by which asymptotic machinery works inevitably washes out sensitivities that are present and important in finite samples. Thus, generality and robustness in asymptotic theory are achieved at the price of insensitivity with

respect to such ingredients as the distributional characteristics of a model's random elements and the values of many of its parameters. These ingredients do influence finite sample behaviour, often in vital ways (as the numerical work reported in Sections 3.5 and 3.6 substantiates). But their effects fade out in very large samples and are totally lost in an asymptotic theory.

There is a second major weakness in the operation of traditional asymptotic machinery. In econometrics, models are most frequently interpreted as approximate data-generating mechanisms that are useful to the extent that they assist in the explanation of observed data, in making predictions, and in other relevant empirical goals. In such exercises, the fact that an investigator can marshall only a finite sample of data is itself a critical factor. For, in the absence of experimentation, the design of a model for empirical use is always tailored by the data that is available and the extent to which the latter may limit the goals of the investigation. As the size and coverage of the sample increase, there is a natural tendency to model an always more complex phenomenon in increasing degrees of sophistication. This process of sophistication arises from the knowledge that a model may be an adequate descriptive and inferential tool over a certain span of data but may be far less adequate as the data evolves further and as the horizons of the investigation widen. When a model grows in complexity with increases in the availability of data, traditional asymptotic methods of statistical analysis inevitably become self-defeating. To extend Fisher's own metaphor, by the time the cannon has rumbled into place, and been loaded and sighted to fire, the sparrow will have flown away. [42]

The central issue is, and will remain, how relevant asymptotic results are in the context of a given model, data set, and plausible hypotheses concerning the stochastic environment. Refinements of asymptotic theory such as those in Section 2.3 can shed light on this issue but are far from conclusive. The ultimate criterion of evaluation will inevitably be the relevant sampling distributions themselves. The recent advancements we have reviewed in the mathematical task of representing the analytic form of these distributions already play an important role in the evaluation of asymptotic theory as the results of Section 3 demonstrate. Moreover, these exact mathematical results and the construction of reliable functional approximants to them, such as those given in Section 4, suggest an alternative basis for estimator choice and a means by which the accuracy of inferential procedures may be improved. It is the challenge of future researchers to extend these results to more sophisticated models and to transmit the machinery in an operational form to the practitioner by the development of appropriate computer software.

---

[42]Some further reflections on the problems inherent in asymptotic theory are given in Phillips (1982b)

References

Amemiya, T (1966) "On the Use of Principal Components of Independent Variables in Two-Stage Least-Squares Estimation", *International Economic Review*, 7, 283–303

Anderson, T. W (1976) "Estimation of Linear Functional Relationships: Approximate Distribution and Connections with Simultaneous Equations in Econometrics", *Journal of the Royal Statistical Society*, B38, 1–36

Anderson, T W (1977) "Asymptotic Expansions of the Distributions of Estimates in Simultaneous Equations for Alternative Parameter Sequences", *Econometrica*, 45, 509–518.

Anderson, T W (1982) "Some Recent Developments on the Distributions of Single-Equation Estimators", in. W Hildebrand (ed.), *Advances in Econometrics* Amsterdam North-Holland Publishing Co. (forthcoming)

Anderson, T W., N Kunitomo and T. Sawa (1980) "Evaluation of the Distribution Function of the Limited Information Maximum Likelihood Estimator", Technical Report No 319, The Economic Series, Stanford University.

Anderson, T W., K. Morimune and T. Sawa (1978) "The Numerical Values of Some Key Parameters in Econometric Models", Stanford University, IMSSS TR no. 270.

Anderson, T W. and H Rubin (1949), "Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations", *Annals of Mathematical Statistics*, 20, 46–63

Anderson, T W and H. Rubin (1950), "The Asymptotic Properties of Estimates of the Parameters of a Single Equation in a Complete System of Stochastic Equations", *Annals of Mathematical Statistics*, 21, 570–582

Anderson, T W. and T Sawa (1973) "Distributions of Estimates of Coefficients of a Single Equation in a Simultaneous System and Their Asymptotic Expansions", *Econometrica*, 41, 683–714

Anderson, T W and T Sawa (1975) "Distribution of a Maximum Likelihood Estimate of a Slope Coefficient The LIML Estimate for Known Covariance Matrix", Technical Report no. 174, IMSSS, Stanford University.

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

Anderson, T W and T. Sawa (1979) "Evaluation of the Distribution Function of the Two-Stage Least Squares Estimate", *Econometrica*, 47, 163–182

Bartlett, M S. (1948) "A Note on the Statistical Estimation of Demand and Supply Relations from Time Series", *Econometrica*, 16, 323–329

Barndorff-Nielson, O. and D R. Cox (1979) "Edgeworth and Saddle-point Approximations with Statistical Applications", *Journal of the Royal Statistical Society*, Ser. B, 41, 279–312

Basmann, R L. (1960) "On Finite Sample Distributions of Generalized Classical Linear Identifiability Test Statistics", *Journal of the American Statistical Association*, 55, 650–659

Basmann, R L (1961) "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) "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 (1965) "On the Application of the Identifiability Test Statistic in Predictive Testing o Explanatory Economic Models", *Econometric Annual of the Indian Economic Journal*, 13, 387–423

Basmann, R. L. (1974) "Exact Finite Sample Distribution for Some Econometric Estimators and Test Statistics. A Survey and Appraisal", M. D Intriligator and D A. Kendricks (eds.), in. *Frontiers o Quantitative Economics*, vol 2. Amsterdam: North-Holland Publishing Co., ch 4, pp. 209–271

Basmann, R L and R H Richardson (1969) "The Finite Sample Distribution of a Structura Variance Estimator", Research Papers in Theoretical and Applied Economics, 24, University o Kansas, mimeographed

Basmann, R L and D H. Richardson (1973), "The Exact Finite Sample Distribution of ; Non-Consistent Structural Variance Estimator", *Econometrica*, 41, 41–58.

Basmann, R L., D H Richardson and R J Rohr (1974) "An Experimental Study of Structural Estimators and Test Statistics Associated with Dynamical Econometric Models", *Econometrica*, 42, 717–730

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

Bhattacharya, R N and R R Rao (1976) *Normal Approximation and Asymptotic Expansions* New York: John Wiley & Sons

Bleistein, N and R A Handelsman (1976) *Asymptotic Expansions of Integrals* New York Holt, Rinehart and Winston

Byron, R P (1974) "Testing Structural Specification Using the Unrestricted Reduced Form", *Econometrica*, 42, 869–883.

Chikuse, Y (1981) "Invariant Polynomials with Matrix Arguments and Their Applications", in R P Gupta (ed), *Multivariate Statistical Analysis* Amsterdam: North-Holland Publishing Co

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

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

Cramér, H (1972) "On the History of Certain Expansions Used in Mathematical Statistics", *Biometrika*, 59, 204–207

Daniels, H E (1954) "Saddlepoint Approximations in Statistics", *Annals of Mathematical Statistics*, 25, 631–650

Daniels, H E (1956) "The Approximate Distribution of Serial Correlation Coefficients", *Biometrika*, 43, 169–185

Daniels, H E (1980) "Exact Saddlepoint Approximations", *Biometrika*, 67, 59–63

Davies, R B (1973) "Numerical Inversion of a Characteristic Function", *Biometrika*, 60, 415–417

Davis, A W (1980a) "Invariant Polynomials with Two Matrix Arguments Extending the Zonal Polynomials", in P R Krishnaiah (ed), *Multivariate Analysis*. Amsterdam North-Holland Publishing Co

Davis, A W (1980b) "Invariant Polynomials with Two Matrix Arguments Extending the Zonal Polynomials Applications to Multivariate Distribution Theory", *Annals of the Institute of Statistical Mathematics*, forthcoming

De Bruijn, N G. (1958) *Asymptotic Methods in Analysis* Amsterdam. North-Holland Publishing Co

Dhrymes, P. J (1969) "Alternative Asymptotic Tests of Significance and Related Aspects of 2SLS and 3SLS Estimated Parameters", *Review of Economic Studies*, 36, 213–226

Drymes, P J (1973) "Restricted and Unrestricted Reduced Froms Asymptotic Distribution and Relative Efficiency", *Econometrica*, 41, 119–134

Durbin, J (1980a) "Approximations for Densities of Sufficient Estimators", *Biometrika*, 67, 311–333

Durbin, J (1980b) "The Approximate Distribution of Partial Serial Correlation Coefficients Calculated from Residuals from Regression on Fourier Series ", *Biometrika*, 67, 335–349

Durbin, J and G S Watson (1971) "Testing for Serial Correlation in Least Squares Regression, III", *Biometrika*, 58, 1–19.

Ebbeler, D H and J B. McDonald (1973) "An Analysis of the Properties of the Exact Finite Sample Distribution of a Nonconsistent GCL Structural Variance Estimator", *Econometrica*, 41, 59–65

Erdeyh, A (1953) *Higher Transcendental Functions*, vol 1. New York McGraw-Hill.

Feller, W (1971) *An Introduction to Probability Theory and Its Applications*, vol II Wiley: New York

Fieller, E C (1932) "The Distribution of the Index in a Normal Bivariate Population", *Biometrika*, 24, 428–440

Fisher, F M (1961) "On the Cost of Approximate Specification in Simultaneous Equation Estimation". *Econometrica*, 29, 139–170

Fisher, F M (1966) "The Relative Sensitivity to Specification Error of Different *k*-Class Estimators", *Journal of the American Statistical Association*, 61, 345–356

Fisher, F M (1967) "Approximate Specification and the Choice of a *k*-Class Estimator", *Journal of the American Statistical Association*, 62, 1265–1276

Fisher, R A (1921) "On the Probable Error of a Coefficient of Correlation Deduced From a Small Sample", *Metron*, 1, 1–32

Fisher, R A (1922) "The Goodness of Fit of Regression Formulae and the Distribution of Regression Coefficients", *Journal of the Royal Statistical Society*, 85, 597–612

Fisher, R A (1924) "The Distribution of the Partial Correlation Coefficient", *Metron*, 3, 329–332

Fisher, R A (1925) *Statistical Methods for Research Workers* Edinburgh· Oliver and Boyd,

Fisher, R A (1928a) "On a Distribution Yielding the Error Functions of Several Well Known Statistics", in *Proceedings of the International Congress of Mathematics* Toronto, pp 805–813

Fisher, R A (1928b) "The General Sampling Distribution of the Multiple Correlation Coefficient", *Proceedings of the Royal Statistical Society*, 121, 654–673

Fisher, R A (1935) "The Mathematical Distributions Used in the Common Tests of Significance", *Econometrica*, 3, 353–365.

Fuller, W A. (1977) "Some Properties of a Modification of the Limited Information Estimator", *Econometrica*, 45, 939–953

Gil-Pelaez, J. (1951) "Note on the Inversion Theorem". *Biometrika*. 38, 481–482

Goldberger, A (1973) "Efficient Estimation in Overidentified Models· An Interpretive Analysis", in· A S Goldberger and O D Duncan (eds.), *Structural Equation Models in the Social Sciences* Seminar Press

Haavelmo, T (1947) "Methods of Measuring the Marginal Propensity to Consume ", *Journal of the American Statistical Association*, 42, 105–122

Hale, C. (1979) "Misspecification in Seemingly Unrelated Regression Equations", Kent University, mimeo

Hale, C, R S Mariano and J G Ramage (1980), "Finite Sample Analysis of Misspecification in Simultaneous Equation Models", *Journal of the American Statistical Association*, 75, 418–427

Hart, J F. (1968) *Computer Approximations* New York· John Wiley & Sons

Hastings, C (1955) *Approximations for Digital Computers* Princeton: Princeton University Press

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

Hausman, J A and D McFadden (1981) "Specification Tests for the Multinomial Logit Model", Discussion Paper no 292, Department of Economics, Massachusetts Institute of Technology.

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

Hillier, G H. and V R. Srivastava (1981) "The Exact Bias and Mean Square Error of the k-Class Estimators for the Coefficient of an Endogenous Variable in a General Structural Equation", Monash University, mimeo

Holly, A and P. C B Phillips (1979) "A Saddlepoint Approximation to the Distribution of the k-Class Estimator of a Coefficient in a Simultaneous System", *Econometrica*, 47, 1527–1547

Hood, Wm C and T C. Koopmans (1953) "The Estimation of Simultaneous Linear Economic Relationships", in: Wm. C. Hood and T C Koopmans (eds), *Studies in Econometric Method*, Cowles Commission. New York. John Wiley & Sons.

Hurwicz, L (1950) "Least Squares Bias in Time Series", in: T C Koopmans (ed.), *Statistical Inference in Dynamic Economic Models* New York. John Wiley & Sons.

Imhof, J. P (1961) "Computing the Distribution of Quadratic Forms in Normal Variables", *Biometrika*, 48, 419–426

Intriligator, M D and D A Kendrick (eds.) (1974) *Frontiers of Quantitative Economics*, vol 11 Amsterdam· North-Holland Publishing Co

James, A T (1961) "Zonal Polynomials of the Real Positive Definite Symmetric Matrices", *Annals of Mathematics*, 74, 456–469

James, A T (1964) "Distribution of Matrix Variates and Latent Roots Derived from Normal Samples", *Annals of Mathematical Statistics*, 35, 475

James, A T (1968) "Circulation of Zonal Polynomial Coefficients by Use of the Laplace–Beltrami operator", *Annals of Mathematical Statistics*, 39, 1711–1718

James, W and C Stein (1961) "Estimation with Quadratic Loss", in *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, vol 1 Berkeley. University of California Press, pp 361–379

Johnson, N L and S Kotz (1970) *Continuous Univariate Distributions–2* Boston: Houghton Mifflin

Johnson, N L and S Kotz (1972) *Distributions in Statistics Continuous Multivariate Distributions* New York· John Wiley & Sons

Kabe, D G (1963) "A Note on the Exact Distributions of the GCL Estimators in Two-Leading Overidentified Cases", *Journal of the American Statistical Association*, 58, 535–537

Kabe, D G (1964) "On the Exact Distributions of the GCL Estimators in a Leading Three-Equation Case", *Journal of the American Statistical Association*, 58, 535–537

Kadane, J (1971) "Comparison of *k*-Class Estimators When the Disturbances are Small", *Econometrica*, 39, 723–737

Kakwani, N C and R. H Court (1972) "Reduced-Form Coefficient Estimation and Forecasting from a Simultaneous Equation Model", *Australian Journal of Statistics*, 14, 143–160

Kataoka, Y (1974) "The Exact Finite Sample Distributions of Joint Least Squares Estimators for Seemingly Unrelated Regression Equations", *Economic Studies Quarterly*, 25, 36–44

Kelejian, H H (1974) "Random Parameters in a Simultaneous Equation Framework: Identification and Estimation", *Econometrica*, 42, 517–527

Kinal, T W (1980) "The Existence of Moments of *k*-Class Estimators", *Econometrica*, 48, 241–249

Kinal, T W (1982) "On the Comparison of Ordinary and Two Stage Least Squares Estimators", SUNY, Albany, mimeo

Knight, J L (1977) "On the Existence of Moments of the Partially Restricted Reduced-Form Estimators from a Simultaneous-Equation Model", *Journal of Econometrics*, 5, 315–321

Knight, J. L (1981) "Non-Normality of Disturbances and the *k*-Class Structural Estimator", School of Economics, The University of New South Wales, mimeo

Kunitomo, N (1981) "On a Third Order Optimum Property of the LIML Estimator When the Sample Size is Large", Discussion Paper no 502, Department of Economics, Northwestern University

Laitinen, K (1978) "Why is Demand Homogeneity Rejected so Often?", *Economic Letters*, 1, 187–191

Lebedev, N N (1972) *Special Functions and Their Application*. New York Dover

Maasoumi, E (1977) "A Study of Improved Methods of Estimating Reduced Form Coefficients Based Upon 3SLS", unpublished Ph.D Thesis, London School of Economics

Maasoumi, E (1978) "A Modified Stein-Like Estimator for the Reduced Form Coefficients of Simultaneous Equations", *Econometrica*, 46, 695–703

Maasoumi, E (1980) "A Ridge-like Method for Simultaneous Estimation of Simultaneous Equations", *Journal of Econometrics*, 12, 161–176

Maasoumi, E (1981) "Uncertain Structural Models and Generic Reduced Form Estimation", Iowa University, mimeo

Maasoumi, E. and P Phillips (1982a) "On the Behaviour of Instrumental Variable Estimators", *Journal of Econometrics* (forthcoming)

Maasoumi, E and P. Phillips (1982b) "Misspecification in the General Single Equation Case", Yale University, mimeo

Maddala, G S. (1974) "Some Small Sample Evidence on Tests of Significance in Simultaneous Equations Models", *Econometrica*, 42, 841–851

Maekawa, K (1980) "An Asymptotic Expansion of the Distribution of the Test Statistics for Linear Restrictions in Zellner's SUR Model", *The Hiroshima Economic Review*, 4, 81–97

Malinvaud, E (1980) *Statistical Methods of Econometrics* Amsterdam· North-Holland Publishing Co

Mariano, R S (1973) "Approximations to the Distribution Functions of the Ordinary Least-Squares and Two-stage Least Squares Estimates in the Case of Two Included Endogenous Variables", *Econometrica*, 41, 67–77.

Mariano, R S (1975) "Some Large-Concentration-Parameter Asymptotics for the *k*-Class Estimators", *Journal of Econometrics* 3, 171–177

Mariano, R S (1977) "Finite-Sample Properties of Instrumental Variable Estimators of Structural Coefficients", *Econometrica*, 45, 487–496

Mariano, R S (1982) "Analytical Small-Sample Distribution Theory in Econometrics The Simultaneous-Equations Case", *International Economic Review* (forthcoming)

Mariano, R S and J McDonald (1979) "A Note on the Distribution Functions of LIML and 2SLS Coefficient Estimators in the Exactly Identified Case", *Journal of the American Statistical Association*, 74, 847–848

Mariano, R S, J McDonald and A Tishler (1979) "On the Effects of Multicollinearity upon the Properties of Structural Coefficient Estimators", mimeo

Mariano, R S and J G Ramage (1978) "Ordinary Least Squares versus other Single Equation Estimators A Return Bout under Misspecification in Simultaneous Systems", University of Pennsylvania Department of Economics Discussion Paper no 400

Mariano, R S and J G Ramage (1979), "Large Sample Asymptotic Expansions for General Linear Simultaneous Systems under Misspecification", University of Pennsylvania, mimeo

Mariano, R S and T Sawa (1972) "The 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.

McCarthy, M D (1972) "A Note on the Forecasting Properties of 2SLS Restricted Reduced Forms", *International Economic Review*, 13, 757–761

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

McDonald, J B. (1974) "An Approximation of the Distribution Function of the LIML Identifiability Test Statistic Using the Method of Moments", *Journal of Statistical Computation and Simulation*, 3, 53–66

McLaren, M L (1976) "Coefficients of the Zonal Polynomials", *Applied Statistics*, 25, 82–87

Meisner, J F (1979) "The Sad Fate of the Asymptotic Slutsky Symmetry Test for Large Systems", *Economic Letters*, 2, 231–233

Mikhail, W M (1969) "A Study of the Finite Sample Properties of Some Economic Estimators", unpublished Ph.D Thesis, London School of Economics

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

Morimune, K (1978) "Improving the Limited Information Maximum Likelihood Estimator When the Disturbances are Small", *Journal of the American Statistical Association*, 73, 867–871.

Morimune, K (1981) "Approximate Distributions of the k-Class Estimators when the Degree of Overidentifiability is Large Compared with the Sample Size", Discussion Paper no 159, Institute of Economic Research, Kyoto University.

Morimune, K and N Kunitomo (1980) "Improving the Maximum Likelihood Estimate in Linear Functional Relationships for Alternative Parameter Sequences", *Journal of the American Statistical Association*, 75, 230–237

Muirhead, R J (1978) "Latent Roots and Matrix Variates: A Review of Some Asymptotic Results", *The Annals of Statistics*, 6, 5–33

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

Nagar, A L. and S N Sahay (1978) "The Bias and Mean Squared Error of Forecasts from Partially Restricted Reduced Form", *Journal of Econometrics*, 7, 227–243.

Nagel, P. J A (1981) "Programs for the Evaluation of Zonal Polynomials", *American Statistician*, 35, 53

Pan Jie-Jian (1968) "Distribution of the Noncircular Serial Correlation Coefficients", *American Mathematical Society and Institute of Mathematical Statistics, Selected Translations in Probability and Statistics*, 7, 281–291

Pfanzagl, J and W Wefelmeyer (1978) "A Third-Order Optimum Property of the Maximum Likelihood Estimator", *Journal of Multivariate Analysis*, 8, 1–29

Pfanzagl, J and W Wefelmeyer (1979) "Addendum to a Third-Order Optimum Property of the Maximum Likelihood Estimator", *Journal of Multivariate Analysis*, 9, 179–182

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

Phillips, P C B (1977b) "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 (1977c) "An Approximation to the Finite Sample Distribution of Zellner's Seemingly Unrelated Regression Estimator", *Journal of Econometrics*, 6, 147–164

Phillips, P C B (1978) "Edgeworth and Saddlepoint Approximations in the First-Order Noncircular Autoregression", *Biometrika*, 65, 91–98

Phillips, P C B (1980a) "The Exact Finite Sample Density of Instrumental Variable Estimators in an Equation with n + 1 Endogenous Variables", *Econometrica*, 48, no 4, 861–878

Phillips, P C B (1980b) "Finite Sample Theory and the Distributions of Alternative Estimators of the Marginal Propensity to Consume", *Review of Economic Studies*, 47, no 1, 183–224

Phillips, P C B (1982a) "Marginal Densities of Instrumental Variable Estimators in the General Single Equation Case", *Advances in Econometrics* (forthcoming)

Phillips, P C B (1982b) "Comments on the Unification of Asymptotic Theory for Nonlinear Econometric Models", *Econometric Reviews* (forthcoming)

Phillips. P C B. (1982c) "A New Approach to Small Sample Theory", Cowles Foundation. Yale University, mimeo

Phillips. P C B (1982d) "Best Uniform and Modified Padé Approximation of Probability Densities in Econometrics", in W Hildenbrand (ed.), *Advances in Econometrics* Cambridge: Cambridge University Press

Phillips, P C B (1982e) "Small Sample Distribution Theory in Econometric Models of Simultaneous Equations", Cowles Foundation Discussion Paper no 617, Yale University

Phillips. P C B. and M R Wickens (1978) *Exercises in Econometrics*, vols I and II Ballinger & Philip Allan

Rainville. E D (1963) *Special Functions*. New York. Macmillan

Rao, C R (1973) *Linear Statistical Inference and its Applications* New York John Wiley & Sons

Revankar, N S and P Mallela (1972) "The Power Function of an F-Test in the Context of a Structural Equation", *Econometrica*, 40, 913–916

Rhodes, G F (1981) "Exact Density Functions and Approximate Critical Regions for Likelihood Ratio Identifiability Test Statistics", *Econometrica*, 49, 1035–1056

Rhodes, G. F and M D Westbrook (1982) "Simultaneous Equations Estimators, Identifiability Test Statistics, and Structural Forms", *Advances in Econometrics* (forthcoming)

Rhodes, G F and M D Westbrook (1981) "A Study of Estimator Densities and Performance Under Misspecification", *Journal of Econometrics* (forthcoming)

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

Richardson, D. H (1969) "On the Distribution of the Identifiability Test Statistic". in J P Quirk and A M Zarley (eds). *Papers in Quantitative Economics* Kansas: Lawrence University Press

Richardson, D H and R J. Rohr (1971) "The Distribution of a Structural t-Statistic for the Case of Two Included Endogenous Variables", *Journal of the American Statistical Association*, 66, 375–382

Richardson, D H and R J. Rohr (1981) "Experimental Evidence on Approximations to Two-Stage Least-Squares Distributions", Brown University, mimeo

Richardson, D. H and R. J Rohr (1982), "An Experimental Study of Two Stage Least Squares Distributions in Equations Containing Three Endogenous Variables", *Advances in Econometrics* (forthcoming)

Richardson, D H and D Wu (1969) "A Note on the Comparison of Ordinary and Two-Stage Least Squares Estimators", Research Papers in Theoretical and Applied Economics (University of Kansas), Paper no 25

Richardson, D H and D. Wu (1970) "Least Squares and Grouping Method Estimators in the Errors in Variables Model", *Journal of the American Statistical Association*, 65, 724–748

Rothenberg, T J (1977) "Edgeworth Expansions for Multivariate Test Statistics", University of California, Berkeley Discussion Paper, unpublished.

Rothenberg, T. J (1978) "Second-Order Efficiency of Estimators and Tests in Simultaneous Equations Models", University of California. Berkeley, mimeo

Rothenberg, T J (1980) "Comparing Alternative Asymptotically Equivalent Tests", presented to World Congress of Econometric Society, Aix-en-Provence 1980

Sahay, S N (1979) "The Exact Mean of the Two-Stage Least Squares Residuals in a Structural Equation Having Three Endogenous Variables", Discussion Paper no 69/79, Department of Economics. University of Bristol

Sahay, S N (1981) "The Bias of the Forecast from Partially Restricted Reduced Form When There are Three Endogenous Variables in the Structural Equation", Discussion Paper no 81, University of Bristol

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 (1975) "Gram–Charlier Approximations Applied to t-Ratios of k-Class Estimators", *Econometrica*, 43, 326–346

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

Sargan, J D (1976b) "Existence of the Moments of Estimated Reduced Form Coefficients", London School of Economics Discussion Paper no A6.

Sargan, J D (1978) "On the Existence of the Moments of 3SLS Estimators", *Econometrica*, 46, 1329–1350

Sargan, J D (1980) "Some Approximation to the Distribution of Econometric Criteria which are Asymptotically Distributed as Chi-squared", *Econometrica*, 48, 1107–1138.

Sargan, J D and S E Satchell (1981) "The Validity of Edgeworth Expansions for Autoregressive Models", in preparation.

Satchell, S E. (1981) "Edgeworth Approximations in Linear Dynamic Models", unpublished Ph.D dissertation, London School of Economics

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

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

Sclove, S. (1968) "Improved Estimators for Coefficients in Linear Regression", *Journal of the American Statistical Association*, 63, 596–606

Sclove, S (1971) "Improved Estimation of Parameters in Multivariate Regression", *Sankya*, Ser. A, 61–66

Slater, L J (1965) "Confluent Hypergeometric Functions", in M Abramowitz and I A Stegun (eds.), *Handbook of Mathematical Functions* New York: Dover

Srinivasan, T N (1970) "Approximations to Finite Sample Moments of Estimators Whose Exact Sampling Distributions are Unknown", *Econometrica*, 38, 533–541

Swamy, P A. V B and J. S Mehta (1980) "On the Existence of Moments of Partially Restricted Reduced Form Coefficients", *Journal of Econometrics*, 14, 183–194.

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

Ullah, A (1974) "On the Sampling Distribution of Improved Estimators for Coefficients in Linear Regression", *Journal of Econometrics*, 2, 143–150

Ullah, A (1980) "The Exact, Large-Sample and Small-Disturbance Conditions of Dominance of Biased Estimators in Linear Models", *Economic Letters*, 6, 339–344

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

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

Wegge, L L (1971) "The Finite Sampling Distribution of Least Squares Estimators with Stochastic Regressors", *Econometrica*, 38, 241–251

Whittaker, E T. and G. N Watson (1927) *Modern Analysis*. Cambridge.

Widder, D. V (1941) *The Laplace Transform* Princeton: Princeton University Press

Widder, D V. (1961) *Advanced Calculus*. Prentice-Hall

Zellner, A. (1978) "Estimation of Functions of Population Means and Regression Coefficients Including Structural Coefficients: A Minimum Expected Loss (MELO) Approach", *Journal of Econometrics*, 8, 127–158

Zellner, A and W Vandaele (1975) "Bayes-Stein Estimators for *k*-Means, Regression and Simultaneous Equation Models", in. S E Feinberg and A. Zellner (eds.), *Studies in Bayesian Econometrics and Statistics in Honor of Leonard J Savage* Amsterdam. North-Holland Publishing Co

Zellner, A. and S B Park (1979) "Minimum Expected Loss (MELO) Estimators for Functions of Parameters and Structural Coefficients of Econometric Models", *Journal of the American Statistical Association*, 74, 183–185