

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS

AT YALE UNIVERSITY

Box 2125, Yale Station
New Haven, Connecticut 06520

COWLES FOUNDATION DISCUSSION PAPER NO. 617

Note: Cowles Foundation Discussion Papers are preliminary materials circulated to stimulate discussion and critical comment. Requests for single copies of a Paper will be filled by the Cowles Foundation within the limits of the supply. References in publications to Discussion Papers (other than mere acknowledgment by a writer that he has access to such unpublished material) should be cleared with the author to protect the tentative character of these papers.

SMALL SAMPLE DISTRIBUTION THEORY IN ECONOMETRIC
MODELS OF SIMULTANEOUS EQUATIONS

Peter C. B. Phillips

February 8, 1982

CONTENTS

	page
1. INTRODUCTION.....	1
2. FORMULATION OF THE DISTRIBUTION PROBLEM AND METHODS OF SOLUTION.....	7
2.1. Primitive Exact Relations and Useful Inversion Formulae.....	7
2.2. Approach via Sample Moments of the Data.....	10
2.3. The Edgeworth Expansion.....	13
2.4. Nagar Moment Approximations.....	22
2.5. Steepest Descents, Saddlepoints and Durbin's Method	25
2.6. Moment Formulae for Ratios.....	34
2.7. The Wishart Distribution and Related Issues.....	36
3. EXACT THEORY IN THE SIMULTANEOUS EQUATIONS MODEL.....	43
3.1. The Model and Notation.....	43
3.2. Generic Statistical Forms of Common Single Equation Estimators.....	45
3.3. The Standardizing Transformations.....	49
3.4. The Analysis of Leading Cases.....	62
3.5. The Exact Distribution of the IV Estimator in the General Single Equation Case.....	70
3.6. The Case of Two Endogenous Variables.....	79
3.7. Structural Variance Estimators.....	88
3.8. Test Statistics.....	92
3.9. Systems Estimators and Reduced Form Coefficients.....	103
3.10. Improved Estimation of Structural Coefficients.....	116
3.11. Supplementary Results on Moments.....	119
3.12. Misspecification.....	122
4. A NEW APPROACH TO SMALL SAMPLE THEORY.....	127
4.1. Intuitive Ideas.....	127
4.2. Rational Approximation.....	129
4.3. An Example of Multiple Point Padé Approximation.....	132
4.4. Curve Fitting or Constructive Functional Approximation?..	137

	page
5. CONCLUDING REMARKS.....	139
FOOTNOTES.....	142
REFERENCES.....	150

SMALL SAMPLE DISTRIBUTION THEORY IN ECONOMETRIC
MODELS OF SIMULTANEOUS EQUATIONS

by

P. C. B. PHILLIPS

Cowles Foundation
Yale University

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! 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 for 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 1960's 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.

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 impression³ 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 1960's and 1970's 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 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 by Hausmann (1982) in this volume of the Handbook. However, the technical machinery which is used in this chapter and which is developed in paragraphs 2.1 through 2.7 has application outside the realm of the SEM. For example, the refinements of asymptotic theory that are provided by Edgeworth expansions (see paragraph 2.3) have important applications in the general statistical

analysis of the second order efficiency of estimators and test statistics. These applications are reviewed in detail by Rothenberg (1982) in Volume 2 of the Handbook. 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 1950's and 1960's with an attempt to improve certain features of the design and efficiency of the experiments, together with the means by which the results of the experiments are characterized. These methods are described by Hendry (1982). 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 describes the genesis, and essential structure of the Edgeworth, Nagar and saddlepoint approximations, all of which have lately 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 (with some fully illustrated derivations) 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 and reduced form coefficient estimators. Section 4 outlines the essential features of a new approach to small sample theory that seems promising for future research. The concluding remarks in Section 5 include reflections on the limitations of traditional asymptotic methods in econometric modeling.

2. FORMULATION OF THE DISTRIBUTION PROBLEM AND METHODS OF SOLUTION

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 ($t=1, \dots, T$), viz. $\{y_1, \dots, y_T\}$, conditional on certain fixed exogenous variables $\{x_1, \dots, x_T\}$ and possibly on certain initial values $\{y_{-k}, \dots, y_0\}$. This distribution can be completely represented by its distribution function (d.f.), $df(y|x, y_{-1}; \theta)$ or its probability density function (p.d.f.), $pdf(y|x, y_{-}; \theta)$, both of which depend on an unknown vector of parameters θ and where we have set $y' = (y'_1, \dots, y'_T)$, $x' = (x'_1, \dots, x'_T)$ and $y'_{-} = (y'_{-k}, \dots, y'_0)$. 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 $pdf(y|x;\theta)$ where the meaning is clear from the context. Estimation of θ or a sub-vector 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

$$(2.1) \quad df(r) = P(\theta_T \leq r) = \int_{y \in \Theta(r)} \text{pdf}(y) dy, \quad \Theta(r) = \{y : \theta_T(y, x) \leq r\}.$$

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

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

$$(2.2) \quad \text{cf}(s) = E(e^{is\theta_T}) = \int e^{is\theta_T(y, x)} \text{pdf}(y) dy$$

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

$$(2.3) \quad \text{pdf}(r) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-isr} \text{cf}(s) ds$$

and this inversion formula is valid provided $\text{cf}(s)$ 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):

$$(2.4) \quad df(r) - df(0) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{1 - e^{-isr}}{is} \text{cf}(s) ds$$

and

$$(2.5) \quad df(r) = \frac{1}{2} + \frac{1}{2\pi} \int_0^{\infty} \frac{e^{isr} \text{cf}(-s) - e^{-isr} \text{cf}(s)}{is} ds.$$

The first of these formulae is valid whenever the integrand on the right 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

$$(2.6) \quad df(b) - df(a) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{e^{-isa} - e^{-isb}}{is} cf(s) ds .$$

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

In the special case where $\theta_T = A_T/B_T$ is a ratio of random variables, some useful inversion formulae have been developed which work directly from the joint characteristic function, $cf(s_1, s_2)$, of the component variates (A_T, B_T) . In particular, if $B_T \geq 0$ and has a finite mean, we have the following formula for the density of θ_T :

$$(2.7) \quad pdf(r) = \frac{1}{2\pi i} \int_{-\infty}^{\infty} \left[\frac{\partial cf(s_1, s_2)}{\partial s_2} \right]_{s_2 = -rs_1} ds_1$$

given by Cramér (1946, p. 317) and Geary (1944). Gurland (1948) generalized (2.7) by considering the multidimensional case of a vector of ratios. He also relaxed the requirement that B_T have a finite mean and gave the following inversion formulae for the density and distribution function:

$$(2.8) \quad pdf(r) = \frac{1}{2\pi i} \int_{-\infty}^{\infty} \left[\frac{\partial cf(s_1, s_2)}{\partial s_2} \right]_{s_2 = -rs_1} ds_1$$

$$(2.9) \quad df(r) = \frac{1}{2} - \frac{1}{2\pi i} \int_{-\infty}^{\infty} \frac{cf(s, -sr)}{s} ds$$

where \oint denotes the Cauchy principal value (see, for example, Olmsted (1961, p. 145)) $\lim_{\epsilon \rightarrow 0, R \rightarrow \infty} \left\{ \int_{\epsilon}^R + \int_{-R}^{-\epsilon} \right\}$, and r is a point of continuity of $df(r)$ in the case of (2.9). Gurland also provides the appropriate inversion formulae for the case in which the denominator of θ_T may be positive or negative (provided $P(B_T = 0) = 0$ still applies).

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, viz. $\text{pdf}(m)$, 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

$$(2.10) \quad \text{pdf}(m) = \int_{a \in A} \text{pdf}(y) \left| \frac{\partial y}{\partial (m,a)} \right| da$$

has already been reduced. In (2.10) a is a vector of auxiliary variates defined over the space A 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

$$(2.11) \quad \text{pdf}(\tau) = \int_{b \in \mathcal{B}} \text{pdf}(m) \left| \frac{\partial m}{\partial(\tau, b)} \right| db$$

where \mathcal{B} is the space of definition of b . The simplicity of the representation (2.11) often belies the major analytic difficulties that are involved in the practical execution of this step.¹ These difficulties center on the selection of a suitable set of auxiliary variates b for which the integration in (2.11) 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 (2.11) is most frequently expressed in terms of infinite 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 Erdélyi (1953). At least in the simpler cases, these series representations can be used for numerical computations of the densities.

An alternative to searching for an exact mathematical solution to the problem of integration in (2.11) is to take the density $\text{pdf}(m)$ 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. The

genesis of these methods will be described in the ensuing paragraphs 2.3-2.5. Both can be considered in terms of the following versions of the primitive inversion formulae (2.2) and (2.3), which involve $\text{pdf}(m)$, viz.

$$(2.12) \quad \text{cf}(s) = \int e^{is\theta_T^*(m)} \text{pdf}(m) dm$$

or, in the form of a Laplace transform (see, for example, Widder (1941)),

$$(2.13) \quad \text{lt}(w) = \int e^{w\theta_T^*(m)} \text{pdf}(m) dm$$

for some complex variable w and where the integral converges in some strip of the imaginary axis defined by $-c_1 \leq \text{Re}(w) \leq c_2$ where $c_1, c_2 \geq 0$. The analogue of (2.3) for the density of θ_T is now

$$(2.14) \quad \text{pdf}(r) = \frac{1}{2\pi i} \int_{-i\infty}^{i\infty} e^{-wr} \text{lt}(w) dw .$$

2.3. The Edgeworth Expansion

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 $\sqrt{T}m \xrightarrow{D} N(0, V)$ as $T \rightarrow \infty$, where D denotes "tends in distribution." Then, if $f(m)$ 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 $\sqrt{T}\{f(m) - f(0)\} \xrightarrow{D} N(0, W)$ where $W = (\partial f(0)/\partial m') V \partial f(0)/\partial m$. 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 $T \rightarrow \infty$,

$m \rightarrow 0$ 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.

In what follows we will detail in skeletal form the constructive process by which the Edgeworth approximation can be derived in the general case. This process was first developed in full by Chambers (1967) but has its origins in the work of Wallace and Edgeworth.² It was first utilized in econometrics by Sargan (1976a). We start our outline by supposing for the sake of convenience that m is already appropriately centered about its mean value and that $\theta_T(m)$ is centered about its value at $m = 0$ so that $\theta_T(0) = 0$. ($\theta_T(m)$ may then represent the error in an estimator rather than the estimator itself. Note that we are also dropping the $*$ in the represen-

tation $\theta_T^*(m)$ used in paragraph 2.2.) We assume:³ that m has moments of a high enough order for the operations that follow to be valid; that $\sqrt{T}m$ has moments of the same order of magnitude in T as would a simple standardized mean of i.i.d. variates (viz. non-zero moments of Tm are $O(T)$ as $T \rightarrow \infty$); that $\theta_T(m)$ has continuous derivatives of a high enough order for the expansions below to be valid; and that the length of the vector $\partial\theta_T(0)/\partial m$ is bounded above zero as $T \rightarrow \infty$. We employ the following notation:

$$\bar{m} = \sqrt{T}m \xrightarrow{D} N(0, V)$$

$$\theta_T(m) = \sqrt{T}e(m) = \sqrt{T}e(\bar{m}/\sqrt{T}), \quad e(0) = 0$$

$$cf(s; \theta_T) = \int e^{is\theta_T(m)} pdf(m) dm = \text{characteristic function of } \theta_T$$

$$cf(z; \bar{m}) = \int e^{iz'\bar{m}} pdf(\bar{m}) d\bar{m} = \text{characteristic function of } \bar{m}$$

$$cgf(z; \bar{m}) = \text{Log}\{cf(z; \bar{m})\} = \text{cumulant generating function of } \bar{m}$$

$$e_r = \frac{\partial e(0)}{\partial m_r}, \quad e_{rs} = \frac{\partial^2 e(0)}{\partial m_r \partial m_s}, \quad cgf_{ij} = \frac{\partial^2 cgf(0)}{\partial z_i \partial z_j}, \quad \text{etc.}$$

and the notational summation device that a repeated subscript is summed over itself:

$$e_{rr} = \sum_r e_{rr}, \quad e_{rsrr} = \sum_r \sum_s e_{rsrr}, \quad \text{etc.}$$

Our constructive process now begins with the Taylor representation

$$\begin{aligned} \theta_T(m) &= \sqrt{T}e_T(m) = \sqrt{T} \left\{ e_{rr} m_r + \frac{1}{2} e_{rs} m_r m_s + \frac{1}{6} e_{rst} m_r m_s m_t + O_p(T^{-2}) \right\} \\ (2.15) \quad &= e_{rr} \bar{m}_r + \frac{1}{2\sqrt{T}} e_{rs} \bar{m}_r \bar{m}_s + \frac{1}{6T} e_{rst} \bar{m}_r \bar{m}_s \bar{m}_t + O_p(T^{-3/2}). \end{aligned}$$

Using this polynomial representation of θ_T in the characteristic function of θ_T we have by formal operations⁴

$$\begin{aligned}
(2.16) \quad cf(s : \theta_T) &= \int e^{ise_r \bar{m}_r + is \left[\frac{1}{2\sqrt{T}} e_{rs} \bar{m}_r \bar{m}_s + \dots \right]} pdf(\bar{m}) d\bar{m} \\
&= \int e^{ise_r \bar{m}_r} \left[1 + \frac{is}{2\sqrt{T}} e_{rs} \bar{m}_r \bar{m}_s + O(T^{-1}) \right] pdf(\bar{m}) d\bar{m}
\end{aligned}$$

and noting that

$$cf_{ab}(z : \bar{m}) = \int (i\bar{m}_a)(i\bar{m}_b) e^{iz' \bar{m}} pdf(\bar{m}) d\bar{m}$$

we deduce that

$$\begin{aligned}
(2.17) \quad cf(s : \theta_T) &= cf(se^0 : \bar{m}) + \frac{ise_{rs}}{2\sqrt{T}} \left(\frac{1}{i} \right)^2 cf_{rs}(se^0 : \bar{m}) + O(T^{-1}) \\
&= cf(se^0 : \bar{m}) - \frac{is}{2\sqrt{T}} e_{rs} cf_{rs}(se^0 : \bar{m}) + O(T^{-1})
\end{aligned}$$

where $e^0 = (e_r)$ is the vector of first derivatives of $e(\bar{m})$ evaluated at the origin. In (2.17) we have written the c.f. of θ_T as a series involving the c.f. of \bar{m} and its subsequent derivatives (with coefficients depending on the derivatives of $e(\bar{m})$). The next step in the process is to expand the c.f. of \bar{m} in a series about the c.f. of the limiting distribution of \bar{m} . To achieve this we first expand the c.g.f. of \bar{m} in a Taylor series at the origin as follows

$$(2.18) \quad cgf(z : \bar{m}) = \frac{1}{2} cgf_{ij} z_i z_j + \frac{1}{6} cgf_{ijk} z_i z_j z_k + O(z^4)$$

where we note that the first non-zero term is quadratic since \bar{m} is centered about its mean value. Now, by definition,

$$(2.19) \quad cf(z : \bar{m}) = \exp\{cgf(z : \bar{m})\}$$

so that setting $z = se^0$ in (2.18) and (2.19) we obtain⁵

$$\begin{aligned}
 \text{cf}(se^0 : \bar{m}) &= \exp\left\{\frac{s^2}{2}\text{cgf}_{ij}e_i e_j + \frac{s^3}{6}\text{cgf}_{ijk}e_i e_j e_k + O(s^4)\right\} \\
 (2.20) \quad &= \exp\left\{\frac{s^2}{2}\text{cgf}_{ij}e_i e_j\right\} \left[1 + \frac{s^3}{6}\text{cgf}_{ijk}e_i e_j e_k + O(s^4)\right]
 \end{aligned}$$

and similarly find that

$$(2.21) \quad \text{cf}_{ab}(se^0 : \bar{m}) = \exp\left\{\frac{s^2}{2}\text{cgf}_{ij}e_i e_j\right\} \left[\text{cgf}_{ab} + (\text{cgf}_{aj}e_j)(\text{cgf}_{bk}e_k)s^2 + O(T^{-1/2})\right]$$

Using (2.20) and (2.21) in (2.17) we find the following series expansion of the c.f. of θ_T :

$$\begin{aligned}
 (2.22) \quad \text{cf}(s : \theta_T) &= \exp\left\{\frac{s^2}{2}\text{cgf}_{ij}e_i e_j\right\} \left[1 - \frac{is}{2\sqrt{T}}\text{cgf}_{rs}e_r e_s + \frac{s^3}{6}\left\{\text{cgf}_{ijk}e_i e_j e_k \right. \right. \\
 &\quad \left. \left. - \frac{3i}{\sqrt{T}}e_{rs}(\text{cgf}_{rj}e_j)(\text{cgf}_{sk}e_k)\right\}\right] + O(T^{-1}) .
 \end{aligned}$$

We let $\omega^2 = -\text{cgf}_{ij}e_i e_j = (\partial e(0)/\partial m') \vee (\partial e(0)/\partial m)$ and note that this is the variance of the limiting distribution of $\theta_T(m) = \sqrt{T}e(m)$ and that $\exp\left\{\frac{s^2}{2}\omega^2\right\}$ is its limiting c.f. Thus, (2.22) is an asymptotic expansion of the c.f. of θ_T about the c.f. of its limiting distribution and the approximation obtained by truncating terms of $O(T^{-1})$ is just a polynomial (in s) multiple of the asymptotic c.f. By inverting (2.22) term by term we deduce the corresponding asymptotic expansion of the p.d.f. of θ_T . To do so we use the relation

$$(-1)^n \phi^{(n)}(z) = \frac{1}{2\pi} \int_{-\infty}^{\infty} (it)^n e^{-t^2/2} e^{-itz} dt$$

or, more appropriately here,

$$(2.23) \quad \frac{1}{\omega^{n+1}} H_n \left(\frac{x}{\omega} \right) \phi \left(\frac{x}{\omega} \right) = (-1)^n \frac{1}{\omega^{n+1}} \phi^{(n)} \left(\frac{x}{\omega} \right) = \frac{1}{2\pi} \int_{-\infty}^{\infty} (is)^n e^{-s^2 \omega^2 / 2} e^{-isx} ds$$

where $\phi(z) = (2\pi)^{-1/2} \exp(-z^2/2)$ is the standard normal density and $H_n(z)$ are the Hermite polynomials (see, for example, Cramér (1946, p. 133)), the first four of which are:

$$(2.24) \quad H_0(z) = 1, \quad H_1(z) = z, \quad H_2(z) = z^2 - 1, \quad H_3(z) = z^3 - 3z.$$

From (2.23) and (2.24) we deduce by inversion of (2.22) that the p.d.f. of θ_T has the following asymptotic expansion as $T \rightarrow \infty$

$$(2.25) \quad \text{pdf}(r) = \frac{1}{\omega} \phi \left(\frac{r}{\omega} \right) \left\{ 1 + \frac{1}{\sqrt{T}} \left[a_1 \left(\frac{r}{\omega} \right) + a_3 \left(\frac{r}{\omega} \right)^3 \right] \right\} + O(T^{-1})$$

where

$$(2.26) \quad a_1 = -\frac{1}{2\omega} \text{cgf}_{rs} e_{rs} - \frac{1}{2\omega^3} [3\{e_{rs} (\text{cgf}_{rj} e_j) (\text{cgf}_{sk} e_k) + iT^{1/2} \text{cgf}_{ijk} e_i e_j e_k\}]$$

$$(2.27) \quad a_3 = \frac{1}{6\omega^3} [3\{e_{rs} (\text{cgf}_{rj} e_j) (\text{cgf}_{sk} e_k) + iT^{1/2} \text{cgf}_{ijk} e_i e_j e_k\}]$$

$$(2.28) \quad \omega^2 = -\text{cgf}_{ij} e_i e_j.$$

Equation (2.25) yields the Edgeworth approximation to $O(T^{-1/2})$ of $\text{pdf}(r)$, which we see comprises a simple polynomial (in (x/ω)) multiple of the limiting normal distribution. The coefficients in the polynomial are given by (2.26), (2.27) and can be deduced from the cumulants of \bar{m} (which are simply proportional to the derivatives of the c.g.f. of \bar{m}) and the derivatives of the function $\theta_T(m) = \sqrt{T}e(m)$. Terms of higher order in $T^{-1/2}$ can be obtained in the same way and explicit

formulae are detailed in Sargan (1976a) and Phillips (1977a). As the above constructive process should make clear the steps leading to (2.25) or its generalized form containing higher order terms are completely mechanical. In particular, the algebra involved in these steps and in the computation of the final formulae (2.26)-(2.28) can in principle be performed by computer symbolic manipulation. This was envisaged by Chambers (1967) in his original article and pilot studies that involve the practical implementation of this process have been performed in a simple setting by Phillips (1980b) and Sargan and Tse (1980).

It is important to note that three different levels of approximation are involved in the constructive process leading to the Edgeworth approximation. These are:

- (i) The approximation of the statistic θ_T by a low degree polynomial in a vector of more basic statistics, \bar{m} , whose moments do exist, at least to a high enough order to ensure the validity of the expansion: see equation (2.15) above;
- (ii) The approximate representation of the c.f. of θ_T in terms of the c.f. of \bar{m} and its successive derivatives: see equations (2.16) and (2.17) above;
- (iii) The approximate representation of the c.f. of \bar{m} in terms of the c.f. of the limiting distribution of \bar{m} : see equation (2.20) above.

Of these approximations, (i) is the most heroic. The adequacy of a polynomial representation of θ_T in terms of \bar{m} will depend on many factors. The most important of these is the width of the interval in θ_T space (or more accurately $e_T(m)$ space) over which a good approximation to the distribution of θ_T is required. For large (small) sample sizes this

interval will generally be smaller (larger) in view of the fact that $m \rightarrow 0$ and $e_T(m) \rightarrow 0$ in probability as $T \rightarrow \infty$ (so that, asymptotically at least, we are inevitably working in the immediate locality of the origin). Thus, when T is small, it may not be possible to adequately capture the random behavior of θ_T by a polynomial approximation in anything else but a very small neighborhood of the origin. When this happens, the Edgeworth expansion will produce a poor approximation to the distribution of θ_T , particularly in the tails where the shape of the true density necessarily depends on the behavior of θ_T in outlying regions of its space of definition. In this connection it is worthy of mention that, whereas θ_T itself may possess no finite integral order moments, all moments of (2.25) are finite (just as they are for the normal distribution). In such cases the (finite) moments of the approximating polynomial representation of θ_T become pseudo-moments of θ_T . In more common parlance of the econometrics literature, they are called moments of an approximating distribution or moments of expansions of distributions (see, for example, Anderson (1977)). Moreover, there is nothing in the definition of the Edgeworth expansion to ensure that it yields a proper distribution with a non-negative p.d.f. that integrates to unity. Thus, the approximation implied by the asymptotic normal distribution and $O(T^{-1/2})$ correction term of (2.25) may become negative as the argument, r , of the density becomes large. Additionally, when terms of $O(T^{-1})$ are included in the correction, the resulting approximation integrates to a constant whose value depends on the numerical values of the coefficients. (Note that the $O(T^{-1/2})$ correction terms involve odd powers of r and integrals over intervals symmetric about the origin have the same value as those based solely on the asymptotic normal.) These features

of the Edgeworth expansion arise from the mechanical process by which it is constructed and point to weaknesses that are thereby inherent in the approximations to which it gives rise. These weaknesses necessarily carry over to the various statistical applications of Edgeworth expansions, such as the analysis of higher order asymptotic efficiency (see, for example, Pfanzagl (1973, 1980), Akahira and Takeuchi (1979) and, for a review of some aspects of this work, Rothenberg (1982) in this Handbook).

Recent literature in mathematical statistics and econometrics has addressed the validity of series expansions of distributions (such as (2.25)) as proper asymptotic series in the mathematical sense. This demands a rigorous analysis of the order of magnitude of the remainder terms and a verification that the error after truncation of the series at a finite number of terms has the same order of magnitude as the first neglected term. Rigorous work in the statistical literature has recently been surveyed by Ibragimov and Linnik (1971), Petrov (1975) and Bhattacharya and Rao (1976). The theory developed in this literature is not sufficiently general to extend to the generic class of problem above for which we have by formal manipulations (that is, without paying attention to the rigorous

justification of the error order of magnitude) extracted the series expansion (2.25). Validity theorems which justify these expansions as valid asymptotic series in such generic situations have recently been proved in the econometric literature by Sargan (1975), Phillips (1977b), and Sargan and Satchell (1981). A general discussion of the issues and the associated literature in more detail may be found in Sections 3 and 4 of Phillips (1980b).

Expansions similar to (2.25) can be extracted for the distribution of similarly smooth functions of more basic statistics such as m which are asymptotically distributed as chi-squared (χ^2). Since the limiting distribution is χ^2 , these expansions involve a leading χ^2 term (corresponding to the leading normal term in (2.25)) and correction terms based on a combination of successive derivatives of the leading term arranged in a series of increasing powers of $T^{-1/2}$. Such expansions have been used for some time in mathematical statistics (see, for example, Fujikoshi (1973, 1977)). In econometrics, related work has been done by Rothenberg (1977) for test statistics in the linear model (see also Rothenberg (1982) in Volume 2 of this Handbook); and a general validity theorem has been proved by Sargan (1980) that applies to (asymptotic χ^2) test statistics in most econometric models which contain only non-random exogenous variables and normally distributed errors.

2.4. Nagar Moment Approximations

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. 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 can be illustrated as follows.⁶

We write $\mu = E(m)$ and suppose that $m = \mu + T^{-1/2} \eta_T$ where $\eta_T = O_p(1)$. We also assume that the function $\theta_T(m)$ is sufficiently smooth to admit the Taylor series expansion

$$\theta_T(m) = \sum_{r=0}^{k-1} \frac{1}{r!} \left[\left\{ (m-\mu)^r \frac{\partial}{\partial m} \right\}^r \theta_T(m) \right]_{m=\mu} + R_{Tk}(m)$$

where $R_{Tk} = O_p(T^{-k/2})$. If we now write

$$\begin{aligned} (2.29) \quad \theta_{Tk} &= \sum_{r=0}^{k-1} \frac{1}{r!} \left[\left\{ (m-\mu)^r \frac{\partial}{\partial m} \right\}^r \theta_T(m) \right]_{m=\mu} \\ &= \sum_{r=0}^{k-1} \frac{1}{r! T^{r/2}} \left[\left\{ \eta_T^r \frac{\partial}{\partial m} \right\}^r \theta_T(m) \right]_{m=\mu} \end{aligned}$$

we may expect that

$$(2.30) \quad E(|\theta_T(m)|^j) = E(|\theta_{Tk}(m)|^j) + O(T^{-k/2})$$

at least to a certain order j less than k and provided $E(|\theta_T(m)|^j)$ is indeed finite. If (2.30) is valid⁷ then $E(\theta_{Tk}(m)^j)$ is the Nagar approximation to the exact moment $E(\theta_T(m)^j)$ up to $O(T^{-(k-1)/2})$.

To illustrate this process we take the case considered in paragraph 2.3 in which $\mu = E(m) = 0$ and $\eta_T = \bar{m}$. From (2.15) we have

$$(2.31) \quad \theta_T(\bar{m}) = e_{r\bar{r}} + \frac{1}{2\sqrt{T}} e_{rs} \bar{m}_r \bar{m}_s + O_p(T^{-1}) .$$

Assuming validity of the moment approximation we deduce that

$$E(\theta_T(\bar{m})) = \frac{1}{2\sqrt{T}} e_{rs} E(\bar{m}_r \bar{m}_s) + O(T^{-1}) .$$

Now, from the c.f. of \bar{m} , we have

$$E(\bar{m}_r \bar{m}_s) = -cf_{rs}(0 : \bar{m})$$

so that

$$(2.32) \quad E(\theta_T(\bar{m})) = -\frac{1}{2\sqrt{T}} e_{rs} cf_{rs}(0 : \bar{m}) + O(T^{-1}) .$$

In fact, we can also deduce (2.32) from the Edgeworth expansion (2.25). We notice that the first moment of the approximating distribution to $O(T^{-1/2})$ is

$$\begin{aligned} & \int_{-\infty}^{\infty} \frac{1}{\omega} \phi\left(\frac{r}{\omega}\right) \left\{ r + \frac{\omega}{\sqrt{T}} \left[a_1 \left(\frac{r}{\omega}\right)^2 + a_3 \left(\frac{r}{\omega}\right)^4 \right] \right\} dr \\ &= \int_{-\infty}^{\infty} \omega \phi(x) \left\{ x + \frac{1}{\sqrt{T}} [a_1 x^2 + a_3 x^4] \right\} dx \\ &= \frac{\omega}{\sqrt{T}} (a_1 + 3a_3) = -\frac{1}{2\sqrt{T}} cgf_{rs} e_{rs} = -\frac{1}{2\sqrt{T}} e_{rs} cf_{rs}(0 : \bar{m}) \end{aligned}$$

as given above in (2.32). The last equality follows by differentiating (2.19) and evaluating the result at the origin.

The approximating statistic θ_{Tk} in (2.29) can also be used to construct a distributional approximation in the following way. We may replace the mean and variance of the asymptotic distribution of θ_T by the mean and variance of θ_{Tk} in (2.29), retaining the normality of the

limiting distribution. This provides a rudimentary correction to the location and variance of the usual asymptotic distribution and is sometimes called the Nagar approximation to the distribution of θ_T . Since normality is retained, this approximation makes no adjustment for possible skewness in the small sample distribution of θ_T , unlike the Edgeworth approximation (2.25). Both approximations are computed and certain aspects of their adequacy are investigated in simple time series models by Phillips (1977a) and Tse (1981).

A worked example of the computation of the Nagar approximation to the bias of the 2SLS estimator in a simple SEM is detailed in Phillips and Wickens (1978, exercise and solution 6.20). Explicit formulae for the Nagar approximation to the bias of the 2SLS estimator taken up to terms of $O(T^{-2})$ are derived by Mikhail (1972).

2.5. Steepest Descents, Saddlepoints and Durbin's Method

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.⁸

The conventional approach to the SP method has its starting point in inversion formulae for the probability density like that of (2.14) given earlier. This integral 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 in (2.14) 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 altitude he has to climb. This particular physical analogy is developed at some length by De Bruijn (1958).

In practice the algebra leading to this new expansion is along the following lines. (Later we will outline an elegant alternative approach that is available in certain cases and is due to a recent significant contribution by Durbin (1980a).) We first suppose that the relevant inversion formula, such as (2.14), admits an alternative representation of the form:

$$(2.33) \quad \text{pdf}(r) = \frac{1}{2\pi i} \int_{-i\infty}^{i\infty} B(w) \exp\{T\Psi(w)\} dw$$

involving the parameter T which we allow to become indefinitely large in order to develop the asymptotics (this may be the sample size or a function of it which also $\rightarrow \infty$ as $T \rightarrow \infty$). We assume that the functions $B(w)$ and $\Psi(w)$ that appear in (2.33) are analytic in some strip of the imaginary axis containing in its interior the point c . In general,

we will then be allowed to deform the path of integration in (2.33) so that the integral can be written as

$$(2.34) \quad \text{pdf}(r) = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} B(w) \exp\{T\Psi(w)\} dw$$

where the path of integration now passes through c .

To obtain an approximate evaluation of the integral (2.34) we select a path of integration in which $c = w^0$ is a suitable SP at which

$$(2.35) \quad \Psi'(w^0) = 0.$$

Solutions of (2.35) are called SP's because, if we write $x = x + iy$ and $\Psi(w) = a(x,y) + ib(x,y)$, then (x^0, y^0) is an SP of the function $a(x,y)$ and hence of the function

$$(2.36) \quad \left| \exp\left\{\frac{\mu}{2}\Psi(w)\right\} \right| = \exp\left\{\frac{\mu}{2}\text{Re}(\Psi(w))\right\} = \exp\left\{\frac{\mu}{2}a(x,y)\right\}.$$

To see this we note that, since $\Psi(w)$ is analytic, $a(\)$ and $b(\)$ satisfy the Cauchy-Riemann conditions (see, for example, Miller (1960))

$$(2.37) \quad \frac{\partial a}{\partial x} = \frac{\partial b}{\partial y}, \quad \frac{\partial a}{\partial y} = -\frac{\partial b}{\partial x}$$

so that

$$\frac{\partial^2 a}{\partial x^2} = \frac{\partial^2 b}{\partial x \partial y}, \quad \frac{\partial^2 a}{\partial y^2} = -\frac{\partial^2 b}{\partial y \partial x}$$

and hence, from the continuity of the second derivatives of b ,

$$(2.38) \quad \frac{\partial^2 a}{\partial x^2} + \frac{\partial^2 a}{\partial y^2} = 0$$

for all values of x and y . Since the first derivatives of $a(\)$ are zero at (x^0, y^0) , and since from (2.38)

$$\left(\frac{\partial^2 a}{\partial x^2} \right) \left(\frac{\partial^2 a}{\partial y^2} \right) - \left(\frac{\partial^2 a}{\partial x \partial y} \right)^2 < 0$$

it now follows that (x^0, y^0) is an SP of $a(x, y)$.

It also follows from (2.37) that the gradient vectors $(\partial a / \partial x, \partial a / \partial y)$ and $(\partial b / \partial x, \partial b / \partial y)$ of $a(\)$ and $b(\)$ are orthogonal. This means that a normal vector to the surface $a(x, y) = \text{const.}$ is tangential to a member of the family of surfaces $\{b(x, y) = \text{const.}\}$. We can deduce that the curve of steepest descent along which $a(x, y)$, and hence (2.36), decreases most rapidly on either side of the SP w^0 is such that

$$\text{Im}(\Psi(w)) = \text{Im}(\Psi(w^0))$$

where $\text{Im}(\)$ denotes the imaginary part. That is,

$$(2.39) \quad b(x, y) = b(x^0, y^0).$$

Frequently, we find that the SP w^0 is real, so that $\Psi(w^0)$ is real and (2.39) becomes

$$(2.40) \quad b(x, y) = 0.$$

Moreover, if we expand $\Psi(w)$ in a region of w^0 we have, in view of (2.35),

$$(2.41) \quad \Psi(w) = \Psi(w^0) + \frac{1}{2!} \Psi''(w^0) (w - w^0)^2 + O(\|w - w^0\|^3).$$

In the vicinity of $w^0 = x^0$, the behavior of $\Psi(w)$ is determined by the first two terms of (2.41). If $\Psi''(w^0) > 0$, then (2.40) holds and $\text{Re}(\Psi(w))$ decreases most rapidly when $x = x^0$ and $y \neq 0$. Thus, in the vicinity of w^0 , the curve of steepest descent becomes the straight line

$$(2.42) \quad w = x^0 + iy$$

which crosses the real axis orthogonally at $w^0 = x^0$. We note also that, on the path defined by (2.42) and in the vicinity of x^0 , (2.36) is dominated by

$$(2.43) \quad \exp\{T\Psi(w^0)\} \exp\left\{-\frac{T}{2}\Psi''(w^0)y^2\right\}.$$

As T becomes large it is clear that most of the contribution of this factor (i.e., of $\exp\{T\Psi(w)\}$) to the value of the integrand arises in the immediate vicinity of the SP w^0 . The geometry of the landscape is crudely illustrated in Figure 1.

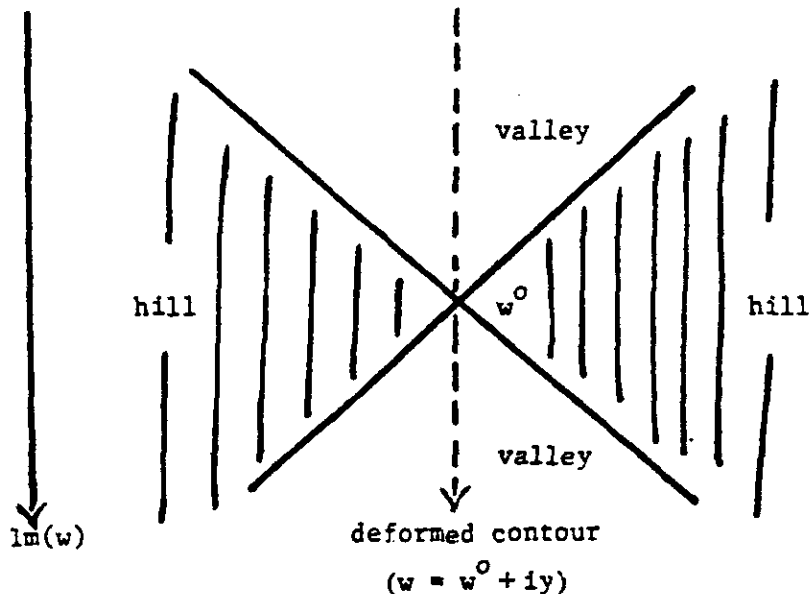


FIGURE 1. Relief diagram for $|\exp\{T\Psi(w)\}| = \exp\{T\text{Re}\{\Psi(w)\}\}$ and deformed path of integration.

We now set $c = w^0$ in (2.34) and change the variable of integration from w to y in $w = w^0 + iy$ giving

$$(2.44) \quad \text{pdf}(r) = \frac{1}{2\pi} \int_{-\infty}^{\infty} B(w^0 + iy) \exp\{T\Psi(w^0 + iy)\} dy .$$

Since $B(w)$ and $\Psi(w)$ are analytic on the path of integration we replace them in (2.44) by the first few terms in their Taylor expansions. Noting that, on this path, $\Psi(w) \sim \Psi(w^0) - \frac{1}{2}\Psi''(w^0)y^2$ in a neighborhood of the SP where the maximum contribution to the integral occurs, we deduce by formal integration the following expansion (where we note that the SP w^0 will depend on the argument r of the p.d.f. of θ_T):

$$(2.45) \quad \text{pdf}(r) = \frac{B(w^0) \exp\{T\Psi(w^0)\}}{\sqrt{2\pi}\{T\Psi''(w^0)\}^{1/2}} [1 + O(T^{-1})] , \quad w^0 = w^0(r) .$$

The first factor in (2.45) is called the SP approximation and the justification of the expansion as a valid asymptotic series is provided by Watson's Lemma (see, for example, Daniels (1954), p. 633). In fact, (2.45) gives the dominant term in an asymptotic series in powers of T^{-1} . The Edgeworth expansion in (2.25) for example, is by comparison a series in powers of $T^{-1/2}$ and thus (2.45) should in general provide a much better asymptotic approximation in its dominant term than the Edgeworth approximation. Moreover, the dominant term of (2.45) is generally positive and renormalization of the area under this curve so that it is unity will produce a proper p.d.f. This implies that the SP approximation will often provide a superior approximation, particularly in the tails of the distribution where the Edgeworth series can produce negative probabilities. More detailed arguments for, and convincing examples of, the favorable nature

of SP approximations are given in Daniels (1954). A successful application of the method in econometrics to k class estimators is illustrated in Holly and Phillips (1979).

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 $\text{pdf}(y;\theta)$ and the limiting mean information matrix $\lim_{T \rightarrow \infty} E\{-T^{-1} \partial^2 \ln \text{pdf}(y;\theta) / \partial \theta \partial \theta'\}$ 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. To see the operational role of the latter step we can recall the general form of the Edgeworth expansion given in (2.25) and note that the series expansion is in powers of $T^{-1/2}$ except when $r = 0$. At this value of r (or alternatively for the corresponding parametric value of θ for which the appropriately centered argument of the density is zero) we have an asymptotic series in powers of T^{-1} , as indeed we have for the SP approximation (recall (2.45)).

The algebra underlying the procedure is as follows. Suppose θ_T is a sufficient statistic for θ so that

$$(2.46) \quad \text{pdf}(y;\theta) = \text{pdf}(r;\theta)h(y)$$

where as usual $\text{pdf}(r; \theta)$ is the p.d.f. of θ_T and $h(y)$ is some function of the data, independent of θ . If we wish to approximate $\text{pdf}(r) = \text{pdf}(r; \theta)$ at the parameter point θ_0 we may deduce from (2.46) the form

$$(2.47) \quad \text{pdf}(r; \theta_0) = \frac{\text{pdf}(y; \theta_0)}{\text{pdf}(y; \theta)} \text{pdf}(r; \theta) .$$

Now (2.47) may be approximated by developing an Edgeworth asymptotic expansion of $\text{pdf}(r; \theta)$ on the right side of (2.47) for a suitable value of θ . As indicated in the last paragraph we may in particular select a value of θ (which will often be $\theta = r$) for which the error on the dominant term of the resulting approximation to $\text{pdf}(r; \theta_0)$ is of $O(T^{-1})$. This process can be illustrated in the general case by using the results of paragraph 2.3. In developing the Edgeworth expansion (2.25) it was assumed that θ_T was already appropriately centered and standardized by \sqrt{T} about θ . Otherwise (2.25) is the expansion of the density of $\sqrt{T}(\theta_T - \theta)$ and the corresponding approximation for θ_T is (setting $s = \sqrt{T}(r - \theta)$):

$$(2.48) \quad \text{pdf}(r; \theta) = \text{pdf}(s) \left| \frac{ds}{dr} \right| = \frac{\sqrt{T}}{\omega} \left(\frac{\sqrt{T}(r - \theta)}{\omega} \right) \left\{ 1 + \frac{1}{\sqrt{T}} \left[a_1 \left(\frac{\sqrt{T}(r - \theta)}{\omega} \right) + a_3 \left(\frac{\sqrt{T}(r - \theta)}{\omega} \right)^3 \right] + O(T^{-1}) \right\} .$$

We now set $\theta = r$ in (2.48) and substituting in (2.47) we get

$$(2.49) \quad \text{pdf}(r; \theta_0) = \frac{\text{pdf}(y; \theta_0)}{\text{pdf}(y; r)} \left(\frac{T}{2\pi\omega^2} \right)^{1/2} \left[1 + O(T^{-1}) \right] , \quad \omega^2 = \omega^2(r) .$$

When higher order terms are included in (2.25) we get an asymptotic series expansion in powers of T^{-1} .

To illustrate the use of (2.49) let us take the case of i.i.d.

gamma variates $\{y_t : t = 1, \dots, T\}$ with p.d.f.

$$(2.50) \quad \text{pdf}(y; \theta) = \prod_{t=1}^T \frac{\lambda^\alpha}{\Gamma(\alpha)} y_t^{\alpha-1} e^{-\lambda y_t} = \prod_{t=1}^T \frac{\alpha^\alpha \theta^{-\alpha}}{\Gamma(\alpha)} y_t^{\alpha-1} e^{-\alpha y_t / \theta}$$

and $E(y_t) = \alpha/\lambda = \theta$. We suppose α is known, so that the sample mean $\theta_T = \bar{y}$ is a sufficient statistic for λ and, hence, θ (see, for example, Cox and Hinkley (1974, p. 27)). We note that $\sqrt{T}(\bar{y} - \theta)$ has a limiting normal distribution with p.d.f. $\omega^{-1} \phi(r/\omega)$ in which $\omega^2 = \alpha/\lambda^2$. Using (2.49) we deduce the following approximation to the finite sample distribution of \bar{y}

$$(2.51) \quad \text{pdf}(r; \theta_0) = \frac{\prod_{t=1}^T \Gamma(\alpha)^{-1} \alpha_0^\alpha \theta_0^{-\alpha} y_t^{\alpha-1} e^{-\alpha y_t / \theta_0}}{\prod_{t=1}^T \Gamma(\alpha)^{-1} \alpha_0^\alpha r^{-\alpha} y_t^{\alpha-1} e^{-\alpha y_t / r}} \left(\frac{T\alpha}{2\pi r^2} \right)^{1/2} [1 + O(T^{-1})]$$

$$(2.52) \quad = \left(\frac{T\alpha}{2\pi} \right)^{1/2} \theta_0^{-T\alpha} r^{T\alpha} r^{T\alpha-1} e^{-T\alpha r / \theta_0} [1 + O(T^{-1})]$$

where we have used the fact that $r = \bar{y}$, $\theta_0 = \alpha/\lambda_0$ and $\omega^2 = \alpha/\lambda^2 = r^2/\alpha$ when we set $\theta = \alpha/\lambda = r$ (that is, $\lambda = \alpha/r$). In the original (α, λ) parametric form we have

$$(2.53) \quad \text{pdf}(r) = \left(\frac{T\alpha}{2\pi} \right)^{1/2} \left(\frac{\lambda}{\alpha} \right)^{T\alpha} r^{T\alpha} r^{T\alpha-1} e^{-T\lambda r} [1 + O(T^{-1})].$$

This result (2.53) was first obtained by Daniels (1954) using the traditional contour integration approach outlined in the first part of this paragraph.

It is a remarkable fact that (2.52) and (2.53) differ from the exact density at $\theta_T = \bar{y}$, viz.

$$(2.54) \quad \text{pdf}(r) = \frac{(\tau\lambda)^{\tau\alpha}}{\Gamma(\tau\alpha)} r^{\tau\alpha-1} e^{-\tau\lambda r}$$

only by the proportional constant. (In fact, if we replace $\Gamma(\tau\alpha)$ by Stirling's approximation to it, i.e., $\Gamma(\tau\alpha) \sim (2\pi)^{1/2} (\tau\alpha)^{\tau\alpha-1/2} e^{-\tau\alpha}$, we find directly:

$$\text{pdf}(r) \sim \left(\frac{\tau\alpha}{2\pi}\right)^{1/2} \left(\frac{\lambda}{\alpha}\right)^{\tau\alpha} e^{\tau\alpha} r^{\tau\alpha-1} e^{-\tau\lambda r}$$

that is the approximation obtained in (2.53) above.) Thus, by renormalizing the SP approximation (2.53) so that the area under the curve is unity we will in this case reproduce the true density of the mean $\theta_T = \bar{y}$ exactly. This remarkable feature of the renormalized SP approximation powerfully illustrates the considerable advantage this method of approximation has over the Edgeworth technique. Daniels (1980) has recently shown that this property of exact reproduction of the true density of the sample mean holds if and only if the underlying distribution of the (i.i.d.) observations is normal, inverse normal or gamma. This characterizes the class of densities for which the property applies in the case of the sample mean. However, the property also holds for certain ratios of sums (see Daniels (1954)) and it has more generally been found that the SP approximation provides a substantially improved approximation over the Edgeworth in a wide variety of cases where the methods have been used in practice.

2.6. Moment Formulae for Ratios

Primitive moment formulae for random variates and ratios of random variates can be obtained in terms of derivatives and integrals of derivatives of the appropriate c.f.'s of the relevant variates. For example, if $cf(s_1, \dots, s_k)$ is the joint c.f. of ξ_1, \dots, ξ_k and if the r^{th}

joint moment $E(\xi_1^{r_1}, \dots, \xi_k^{r_k})$, $r = \sum_{j=1}^k r_j$, of the distribution exists we have the basic expression:

$$(2.55) \quad E(\xi_1^{r_1}, \dots, \xi_k^{r_k}) = (-1)^r \frac{\partial^r \text{cf}(0, \dots, 0)}{\partial s_1^{r_1} \partial s_2^{r_2} \dots \partial s_k^{r_k}}$$

(see, for example, Wilks (1962), p. 119).

Frequently, it is useful to have expressions similar to (2.55) but which apply directly to moments of ratios of variates. Under the assumption that the joint moment generating function (m.g.f.) of the variates exists, viz.

$$\text{mgf}(s_1, s_2) = E(e^{s_1 \xi_1 + s_2 \xi_2})$$

the following two expressions can readily be deduced:

$$(2.56) \quad E \left[\frac{\xi_1}{\xi_2} \right]^r = \int_{-\infty}^0 \dots \int_{-\infty}^0 \frac{\partial^r \text{mgf}(s_1, s_2)}{\partial s_1^r} \Big|_{s_1=0} ds_{21} \dots ds_{2r}, \quad s_2 = \sum_{j=1}^r s_{2j}$$

$$(2.57) \quad E \left[\frac{\xi_1}{\xi_2} \right]^r = \frac{1}{\Gamma(r)} \int_{-\infty}^0 (-s_2)^{r-1} \left[\frac{\partial^r \text{mgf}(s_1, s_2)}{\partial s_1^r} \right]_{s_1=0} ds_1 .$$

The first of these is established by noting that

$$(2.58) \quad \left[\frac{\partial^r \text{mgf}(s_1, s_2)}{\partial s_1^r} \right]_{s_1=0} = \int x_1^r e^{s_2 x_2} \text{pdf}(x_1, x_2) dx_1 dx_2$$

and, since $\int_{-\infty}^0 e^{s_2 x_2} ds_{2j} = 1/x_2$, successive integrations of (2.58) over

$-\infty < s_{2j} < 0$ ($j = 1, \dots, r$) lead directly to (2.56). The second expression (2.57) is established by writing ξ_2^{-r} as a gamma integral

$$\left(\frac{\xi_1}{\xi_2}\right)^r = \frac{\xi_1^r}{\Gamma(r)} \int_0^\infty t^{r-1} e^{-\xi_2 t} dt$$

(see, for example, Lebedev (1965), p. 13) and then

$$E\left(\frac{\xi_1}{\xi_2}\right)^r = \frac{1}{\Gamma(r)} \int_{-\infty}^0 (-t)^{r-1} E(\xi_1^r e^{t\xi_2}) dt$$

which leads to (2.57).

A particularly simple result holds for moments of ratios of certain quadratic forms in normal variates. If y is $N(0, I)$, M is symmetric idempotent and A is a symmetric matrix then $E(R^r) = E(A^r)/E(B^r)$ where

$$R = \frac{y' M A M y}{y' M y} = \frac{A}{B}.$$

In this case, R is distributed independently of the denominator form, B . This result is especially useful in computing moments of the Durbin-Watson statistic. It is proved by Hannan (1970, p. 343) and is due originally to Pitman (1937). A closely related result was established using the general expression (2.56) above by Williams (1941).

2.7. The Wishart Distribution and Related Issues

If $X = [x_1, \dots, x_T]$ 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

$$(2.59) \quad \text{pdf}(A) = \frac{1}{(2)^{nT/2} \Gamma_n\left(\frac{T}{2}\right) (\det \Omega)^{T/2}} \text{etr}\left(-\frac{1}{2}\Omega^{-1}A\right) (\det A)^{(T-n-1)/2} .$$

Since A is symmetric $n \times n$, this density has $N = \frac{1}{2}n(n+1)$ 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 (2.59), although in transforming the distribution we must recognize the correct number of independent arguments. Thus, in transforming $A \rightarrow GAG' = B$ where G is a non-singular $n \times n$ matrix, the jacobian of the transformation is given by $dB = (\det G)^{n+1}dA$. This result is most easily established by writing down differentials $dB = GdAG'$ or $d(\text{vec } B) = (G \otimes G)d(\text{vec } A)$ and then rewriting this equation in terms of non-redundant elements of the differentials. In particular, we have $db = L_n(G \otimes G)D_n da$ where L_n and D_n are the elimination and duplication matrices of dimension $N \times n^2$ and $n^2 \times N$ respectively defined by $L_n \text{vec } B = b$, $D_n a = \text{vec } A$. We now deduce the jacobian as $\det\{L_n(G \otimes G)D_n\} = (\det G)^{n+1}$. This and other useful results involving matrix transformations are established by Magnus and Neudecker (1980).

In (2.59) above $\Gamma_n(z)$ is the multivariate gamma function defined by the integral

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

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

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

(see James (1964)). In (2.59) we also use the abbreviated operator representation $\text{etr}(\cdot) = \exp\{\text{tr}(\cdot)\}$.

The parameters of the Wishart distribution (2.59) are: (i) the order of the symmetric matrix A , viz. n ; (ii) the degrees of freedom T , of the component variates x_t in the summation $A = XX' = \sum_{t=1}^T x_t x_t'$; (iii) the covariance matrix, Ω , of the normally distributed columns x_t in X . A common notation for the Wishart distribution (2.59) is then $W_n(T, \Omega)$ --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 (2.59) reduces to $(2)^{-T/2} \Gamma\left(\frac{T}{2}\right)^{-1} a^{T/2-1} e^{-\frac{1}{2}a}$, 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 (noncentral) Wishart with noncentrality matrix $\bar{M} = MM'$ where $M = [m_1, \dots, m_T]$. This is frequently denoted $W_n(T, \Omega, \bar{M})$ although M is sometimes used in place of \bar{M} (as in Rao (1973), for example). The latter is a more appropriate parameter in the matrix case as a convenient generalization of the noncentrality parameter that is used in the case of the noncentral χ^2 distribution--a special case of $W_n(T, \Omega, \bar{M})$ in which $n = 1$, $\Omega = 1$ and $\bar{M} = \sum_{t=1}^T m_t^2$.

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

$$(2.60) \quad \text{pdf}(A) = \frac{\text{etr} \left\{ -\frac{1}{2} \Omega^{-1} \bar{M} \right\}}{2^{nT/2} \Gamma_n \left(\frac{T}{2} \right) (\det \Omega)^{T/2}} {}_0F_1 \left\{ \frac{T}{2}; \frac{1}{4} \Omega^{-1} \bar{M} \Omega^{-1} A \right\} \text{etr} \left\{ -\frac{1}{2} \Omega^{-1} A \right\} (\det A)^{(T-n-1)/2} .$$

In (2.60) 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, Erdélyi (1953)) to matrix argument functions by using multidimensional Laplace transforms and inverse transforms. Constantine (1963) discovered that hypergeometric functions ${}_pF_q$ of matrix argument have a general series representation in terms of zonal polynomials as follows:

$$(2.61) \quad {}_pF_q(a_1, \dots, a_p; b_1, \dots, b_q; S) = \sum_{j=0}^{\infty} \sum_J \frac{(a_1)_J \dots (a_p)_J}{(b_1)_J \dots (b_q)_J} \frac{C_J(S)}{j!} .$$

In (2.61) 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 $\{j_1, \dots, j_r\}$ such that $\sum_{i=1}^r j_i = j$. For example $\{2,1\}$ and $\{1,1,1\}$ are partitions of 3 and are conventionally written (21) and (1^3) . The coefficients $(a)_J$ and $(b)_J$ in (2.61) are multivariate hypergeometric coefficients defined by

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

and where

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

The factor $C_j(S)$ in (2.61) 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

$$s_m = \text{sum of the } m^{\text{th}} \text{ powers of the latent roots of } S .$$

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	$\frac{1}{3}(s_1^3 - 3s_1s_2 + 2s_3)$
	21	$\frac{3}{5}(s_1^3 + s_1s_2 - 2s_3)$
	3	$\frac{1}{15}(s_1^3 + 6s_1s_2 + 8s_3)$

(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 (2.61). However, certain special cases of (2.61) are already recognizable in terms of simpler functions: when $n = 1$ we have the classical hypergeometric functions

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

(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 $p = 1$, $q = 0$ 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 noncentral Wishart density (2.60) generalizes the classical Bessel function. (The reader may recall that the noncentral 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 $n = 1$, $\Omega = 1$, $\bar{M} = \lambda$ and $A = a$ is a scalar we have

$$\begin{aligned}
 \text{pdf}(a) &= \frac{\exp\left\{-\frac{1}{2}(a+\lambda)\right\}}{2^{T/2}\Gamma(T/2)} a^{T/2-1} \sum_{j=0}^{\infty} \frac{\left(\frac{1}{4}\lambda a\right)^j}{(T/2)_j j!} \\
 (2.62) \quad &= \frac{\exp\left\{-\frac{1}{2}(a+\lambda)\right\}}{2^{T/2}} \sum_{j=0}^{\infty} \frac{\lambda^j a^{T/2+j-1}}{\Gamma(T/2+j)j!2^{2j}}.
 \end{aligned}$$

This is the usual form of the p.d.f. of a noncentral χ^2 variate.

The mean and variance of the noncentral Wishart matrix A with p.d.f. (2.60) have been derived by Magnus and Neudecker (1979) in the following convenient form (in which $\mathcal{D}(\xi)$ denotes the covariance matrix of the random vector ξ):

$$E(A) = T\Omega + \bar{M}$$

$$\mathcal{D}(\text{vec } A) = (I + K_n)[T(\Omega \otimes \Omega) + \Omega \otimes \bar{M} + \bar{M} \otimes \Omega]$$

where K_n is the $n^2 \times n^2$ (commutation) matrix which transforms $\text{vec}(A)$ into $\text{vec}(A')$. More explicitly

$$K_n = \sum_{i=1}^n \sum_{j=1}^n (E_{ij} \otimes E'_{ij})$$

where E_{ij} is the $n \times n$ matrix with 1 in its $(i,j)^{\text{th}}$ position and zeroes elsewhere.

Other useful properties of the central and noncentral Wishart distributions are discussed in the books by Anderson (1958), Johnson and Kotz (1972) and Rao (1973).

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

$$(3.1) \quad YB + ZC = U$$

and its reduced form as

$$(3.2) \quad Y = Z\bar{\gamma} + V$$

where $Y' = [y_1, \dots, y_T]$ is a $G \times T$ matrix of T observations of G endogenous variables, $Z' = [z_1, \dots, z_T]$ is a $K \times T$ matrix of T observations of K non-random exogenous variables and $U' = [u_1, \dots, u_T]$ 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 Hsiao (1982) and Hausman (1982) in this Handbook), it is also assumed that the u_t ($t = 1, \dots, T$) 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

($t = 1, \dots, T$) to be i.i.d. $N(0, \Omega)$.

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:

$$(3.3) \quad y_1 = Y_2\beta + Z_1\gamma + u,$$

or

$$(3.4) \quad y_1 = W_1\delta + u, \quad W_1 = [Y_2 : Z_1], \quad \delta' = (\beta', \gamma')$$

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

$$(3.5) \quad [y_1 : Y_2] = [Z_1 : Z_2] \begin{bmatrix} \pi_{11} & \pi_{12} \\ \pi_{21} & \pi_{22} \end{bmatrix} + [v_1 : v_2]$$

or

$$X = Z\Pi^S + V^S, \quad X = [y_1 : Y_2], \quad Z = [Z_1 : Z_2]$$

where Z_2 is a $T \times K_2$ matrix of exogenous variables excluded from (3.3). The system (3.5) represents $n+1$ columns of the complete reduced form (containing $G \geq n+1$ 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 $K_2 \geq n$ 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

$$(3.6) \quad \pi_{11} - \pi_{12}\beta = \gamma, \quad \pi_{21} - \pi_{22}\beta = 0.$$

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 paragraph 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 noncentrality 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

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

and writing

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

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

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

$$(3.10) \quad \beta_{IV} = \{Y_2' [P_H - P_H Z_1 (Z_1' P_H Z_1)^{-1} Z_1' P_H] Y_2\}^{-1} \{Y_2' [P_H - P_H Z_1 (Z_1' P_H Z_1)^{-1} Z_1' P_H] y_1\}.$$

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

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

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

We define the moment matrix

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

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

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

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

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

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

Setting $\beta'_\Delta = (1, -\beta')$ we may also note that

$$(3.17) \quad \beta_{OLS} \text{ minimizes } \beta'_\Delta A(I) \beta_\Delta = \beta'_\Delta X' Q_{Z_1} X \beta_\Delta$$

and

$$(3.18) \quad \beta_{2SLS} \text{ minimizes } \beta'_\Delta A(P_Z) \beta_\Delta = \beta'_\Delta X' (P_Z - P_{Z_1}) X \beta_\Delta .$$

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

$$(3.19) \quad \begin{aligned} \beta_{(k)} &= \{Y_2' [k(P_Z - P_{Z_1}) + (1-k)Q_{Z_1}] Y_2\}^{-1} \{Y_2' [k(P_Z - P_{Z_1}) + (1-k)Q_{Z_1}] y_1\} \\ &= [kA_{22}(P_Z) + (1-k)A_{22}(I)]^{-1} [ka_{21}(P_Z) + (1-k)a_{21}(I)] \end{aligned}$$

(see, for example, Mariano (1977)) and

$$(3.20) \quad \beta_{(k)} \text{ minimizes } \beta'_\Delta [kA(P_Z) + (1-k)A(I)] \beta_\Delta = \beta'_\Delta X' [kA(P_Z) + (1-k)A(I)] X \beta_\Delta .$$

We also note that for the LIML estimator, β_{LIML} , of β

$$(3.21) \quad \beta_{LIML} \text{ minimizes } \frac{\beta'_\Delta A(I) \beta_\Delta}{\beta'_\Delta [A(I) - A(P_Z)] \beta_\Delta} = \frac{\beta'_\Delta X' Q_{Z_1} X \beta_\Delta}{\beta'_\Delta X' Q_Z X \beta_\Delta}$$

and that β_{LIML} satisfies the system

$$(3.22) \quad \{A(I) - \lambda[A(I) - A(P_Z)]\} \beta_\Delta = 0$$

where λ is the minimum of the variance ratio in (3.21). Thus, β_{LIML} is given by the generic form

$$(3.23) \quad \beta_{LIML} = [\lambda A_{22}(P_Z) + (1-\lambda)A_{22}(I)]^{-1} [\lambda a_{21}(P_Z) + (1-\lambda)a_{21}(I)]$$

that is, the k class estimator (3.19) 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 (2.11) 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 last paragraph 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.¹ They will now be

discussed in more detail.

3.3. The Standardizing Transformations

Transformation into canonical form

We introduce a non-singular lower triangular matrix L for which

$$(3.24) \quad \Omega = L'L .$$

We will partition L conformably with $[y_1; Y_2]$, as in (3.5), leading to the block triangular form

$$(3.25) \quad L = \begin{bmatrix} l_{11} & 0 \\ l_{21} & L_{22} \end{bmatrix} .$$

The inverse of L is simply obtained as

$$(3.26) \quad L^{-1} = \begin{bmatrix} 1/l_{11} & 0 \\ -L_{22}^{-1}l_{21}/l_{11} & L_{22}^{-1} \end{bmatrix}$$

and will be used in what follows. We will occasionally use the explicit form of L given below in (3.27) which can be deduced from (3.24) and the following partitioned form of Ω :

$$(3.27) \quad \Omega = \begin{bmatrix} \omega_{11} & \omega'_{21} \\ \omega_{21} & \Omega_{22} \end{bmatrix} ;$$

that is

$$(3.28) \quad L = \left[\begin{array}{c|c} (\omega_{11} - \omega_{21}' \Omega_{22}^{-1} \omega_{21})^{1/2} & 0 \\ \hline \Omega_{22}^{-1/2} \omega_{21} & \Omega_{22}^{1/2} \end{array} \right].$$

Under the transformation L^{-1} (3.5) becomes

$$(3.29) \quad [y_1 : Y_2] L^{-1} = Z \Pi^S L^{-1} + V^S L^{-1}$$

and the rows of $V^S L^{-1}$, viz. $v_t' L^{-1}$ have covariance matrix I_{n+1} .

The corresponding transformation of the structural equation (3.3) is

$$(3.30) \quad [y_1 : Y_2] L^{-1} L \begin{bmatrix} 1 \\ -\beta \end{bmatrix} = Z_1 \gamma + u$$

or

$$(3.31) \quad [y_1^* : Y_2^*] \begin{bmatrix} \ell_{11} \\ \ell_{21} - L_{22} \beta \end{bmatrix} = Z_1 \gamma + u$$

and renormalization is achieved by division with ℓ_{11} giving

$$(3.32) \quad [y_1^* : Y_2^*] \begin{bmatrix} 1 \\ (\ell_{21} - L_{22} \beta) / \ell_{11} \end{bmatrix} = Z_1 \gamma / \ell_{11} + u / \ell_{11}.$$

We rewrite (3.32) as

$$[y_1^* : Y_2^*] \begin{bmatrix} 1 \\ -\beta^* \end{bmatrix} = Z_1 \gamma^* + u^*$$

or

$$(3.33) \quad y_1^* = Y_2^* \beta^* + Z_1^* \gamma^* + u^*$$

where

$$(3.34) \quad y_1^* = (y_1 - Y_2 L_{22}^{-1} l_{21}) / l_{11} = (\omega_{11} - \omega_{21}' \Omega_{22}^{-1} \omega_{21})^{-1/2} (y_1 - Y_2 \Omega_{22}^{-1} \omega_{21}) ,$$

$$(3.35) \quad Y_2^* = Y_2 L_{22}^{-1} = Y_2 \Omega_{22}^{-1/2} ,$$

and

$$(3.36) \quad \begin{aligned} \beta^* &= (L_{22} \beta - l_{21}) / l_{11} = (\omega_{11} - \omega_{21}' \Omega_{22}^{-1} \omega_{21})^{-1/2} (\Omega_{22}^{1/2} \beta - \Omega_{22}^{-1/2} l_{21}) \\ &= (\omega_{11} - \omega_{21}' \Omega_{22}^{-1} \omega_{21})^{-1/2} \Omega_{22}^{1/2} (\beta - \Omega_{22}^{-1} l_{21}) , \end{aligned}$$

and

$$(3.37) \quad \gamma^* = \gamma / l_{11} = (\omega_{11} - \omega_{21}' \Omega_{22}^{-1} \omega_{21})^{-1/2} \gamma .$$

Equation (3.33) is called the canonical form of the original structural equation (3.3).

Orthonormalization of the exogenous variables

We introduce a non-singular upper triangular matrix J for which

$$T^{-1} Z' Z = J' J$$

(sometimes it is convenient to work with $Z' Z$ rather than $T^{-1} Z' Z$ and this will involve only a rescaling of the elements of J). Under the transformation J we can write (3.5) in the form

$$(3.38) \quad X = Z J^{-1} J \Pi^s + V^s = \bar{Z} \bar{\Pi} + V$$

where $\bar{Z} = Z J^{-1}$ and $\bar{\Pi} = J \Pi^s$. We drop for convenience the superscript s on $\bar{\Pi}$ and note that $\bar{Z}' \bar{Z} = T L_K$ where $K = K_1 + K_2$.

We now partition J conformably with $[Z_1 : Z_2]$ and write

$$(3.39) \quad J = \begin{bmatrix} K_1 & K_2 \\ J_{11} & J_{12} \\ 0 & J_{22} \end{bmatrix} \begin{matrix} K_1 \\ K_2 \end{matrix}.$$

The inverse of J will be used in what follows and is given by

$$(3.40) \quad J^{-1} = \begin{bmatrix} J_{11}^{-1} & | & -J_{11}^{-1}J_{12}J_{22}^{-1} \\ \hline 0 & | & J_{22}^{-1} \end{bmatrix}.$$

We can also partition the matrix $Z'Z$ as

$$(3.41) \quad Z'Z = \begin{bmatrix} Z_1'Z_1 & Z_1'Z_2 \\ Z_2'Z_1 & Z_2'Z_2 \end{bmatrix}.$$

It is sufficient in this case to work with a block triangular J in (3.39) and the following explicit form is useful:

$$(3.42) \quad J = \frac{1}{\sqrt{T}} \begin{bmatrix} (Z_1'Z_1)^{1/2} & | & (Z_1'Z_1)^{-1/2}Z_1'Z_2 \\ \hline 0 & | & [Z_2'Z_2 - Z_2'Z_1(Z_1'Z_1)^{-1}Z_1'Z_2]^{1/2} \end{bmatrix}.$$

Under the orthonormalizing transformation (3.42) the reduced form matrices in (3.38) take the form

$$(3.43) \quad \bar{Z} = ZJ^{-1} = [Z_1J_{11}^{-1} : -Z_1J_{11}^{-1}J_{12}J_{22}^{-1} + Z_2J_{22}^{-1}] = [\bar{Z}_1 : \bar{Z}_2]$$

and

$$(3.44) \quad \bar{\pi} = J\pi = \begin{bmatrix} \overset{1}{J_{11}\pi_{11} + J_{12}\pi_{21}} & \overset{n}{J_{11}\pi_{12} + J_{12}\pi_{22}} \\ \hline J_{22}\pi_{21} & J_{22}\pi_{22} \end{bmatrix} \begin{matrix} K_1 \\ K_2 \end{matrix}.$$

We now state the following result which summarizes the effect of the above standardizing transformations:

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*

$$(3.45) \quad T^{-1}Z'Z = I_K \quad \text{and} \quad \Omega = I_{n+1}.$$

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

$$(3.46) \quad y_1^* = Y_2^* \beta^* + \bar{Z}_1 \bar{\gamma}^* + u^*$$

and

$$(3.47) \quad [y_1^* : Y_2^*] = \bar{Z} \bar{\pi}^* + \bar{v}$$

where $T^{-1}\bar{Z}'\bar{Z} = I_K$ and the rows of $[y_1^* : Y_2^*]$ are uncorrelated with covariance matrix given by I_{n+1} . Explicit formulae for the new coefficients in (3.46) are

$$(3.48) \quad \beta^* = (L_{22}\beta - l_{21})/l_{11} = (\omega_{11} - \omega_{21}'\Omega_{22}^{-1}\omega_{21})^{-1/2} \Omega_{22}^{1/2} (\beta - \Omega_{22}^{-1}\omega_{21})$$

and

$$(3.49) \quad \bar{\gamma}^* = J_{11}\gamma^* = J_{11}\gamma/l_{11} = \left(\frac{Z_1'Z_1}{T} \right)^{1/2} (\omega_{11} - \omega_{21}'\Omega_{22}^{-1}\omega_{21})^{-1/2} \gamma.$$

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

PROOF. We use the transformations L and J as defined by (3.25), (3.28) and (3.39), (3.42). Writing

$$(3.50) \quad [y_1^* : y_2^*] = [y_1 : y_2]L^{-1}$$

and

$$(3.51) \quad \bar{Z} = ZJ^{-1}, \quad \bar{\Pi} = J\Pi$$

it follows from (3.33) and (3.43) that the structural equation (3.3) takes the form

$$y_1^* = Y_2^* \beta^* + Z_1 J_{11}^{-1} J_{11} \gamma^* + u^*$$

or

$$(3.52) \quad y_1^* = Y_2^* \beta^* + \bar{Z}_1 \bar{\gamma}^*$$

where β^* is given by (3.48) as established in the argument leading to (3.36) above, and where

$$\bar{\gamma}^* = J_{11} \gamma^* = J_{11} \gamma / \lambda_{11}$$

as in (3.49) and where $\bar{Z}_1 = Z_1 J_{11}^{-1}$.

Moreover, from (3.29) we have the reduced form

$$\begin{aligned} [y_1 : y_2]L^{-1} &= Z\Pi L^{-1} + VL^{-1} \\ &= ZJ^{-1}J\Pi L^{-1} + VL^{-1} \end{aligned}$$

or

$$(3.53) \quad [y_1^* : y_2^*] = \bar{Z} \bar{\Pi}^* + \bar{V}.$$

It is clear from (3.52) that these transformations L^{-1} and J preserve the number of excluded exogenous variables in the structural equation.

From (3.26) and (3.39) we now find

$$\begin{aligned} \bar{\Pi}^* &= J\Pi L^{-1} \\ (3.54) \quad &= \left[\begin{array}{c|c} (J_{11}\pi_{11} + J_{12}\pi_{21})/\ell_{11} & (J_{11}\pi_{12} + J_{12}\pi_{22})L_{22}^{-1} \\ - (J_{11}\pi_{12} + J_{12}\pi_{22})L_{22}^{-1}\ell_{21}/\ell_{11} & \\ \hline (J_{22}\pi_{21} - J_{22}\pi_{22}L_{22}^{-1}\ell_{21})/\ell_{11} & J_{22}\pi_{22}L_{22}^{-1} \end{array} \right]. \end{aligned}$$

Since J_{22} and L_{22} are non-singular it follows that the lower right hand submatrix of $\bar{\Pi}^*$, viz. $\bar{\Pi}_{22}^* = J_{22}\pi_{22}L_{22}^{-1}$, has the same rank as π_{22} , which by assumption is n . Thus, the rank condition for identifiability is preserved. \square

Expressions (3.48) and (3.49) in the statement of Theorem 3.3.1 show that the canonical form transformation affects both coefficient vectors β and γ in the structural equation (3.3). However, orthonormalization of the exogenous variables affects only the coefficients of the exogenous variables, viz. γ . We note also that in the transformed model (3.53) $\bar{Z}'\bar{Z} = T\Gamma_K$ as required.

Effect of the Standardizing Transformations on Econometric Estimators

It turns out that the commonly used econometric estimators of the standardized coefficients β^* and $\bar{\gamma}^*$ in (3.46) are related to the unstandardized coefficient estimators by the same relations which define the standardized coefficients viz. (3.48) and (3.49). We illustrate by considering the 2SLS estimators. If we set $H = Z$ in (3.12) we may use

the transformation matrices L and J as they have been derived in the previous analysis. We note that, by construction,

$J_{11}'^{-1}Z_1'Z_1J_{11}^{-1} = \Pi_{K_1}$ and $J'^{-1}Z'ZJ^{-1} = \Pi_K$. From (3.11) and (3.12) we deduce that

$$(3.55) \quad J_{11}\gamma_{2SLS} = (\bar{Z}_1'\bar{Z}_1)^{-1}\bar{Z}_1'(y_1 - Y_2\beta_{2SLS}) = T^{-1}\bar{Z}_1'(y_1 - Y_2\beta_{2SLS})$$

and

$$(3.56) \quad \beta_{2SLS} = \left\{ Y_2' \left[\frac{ZJ^{-1}J'^{-1}Z'}{T} - \frac{Z_1J_{11}^{-1}J_{11}'^{-1}Z_1'}{T} \right] Y_2 \right\}^{-1} \left\{ Y_2' \left[\frac{ZJ^{-1}J'^{-1}Z'}{T} - \frac{Z_1J_{11}^{-1}J_{11}'^{-1}Z_1'}{T} \right] y_1 \right\}$$

$$= \left[Y_2' \left(\frac{\bar{Z}\bar{Z}'}{T} - \frac{\bar{Z}_1\bar{Z}_1'}{T} \right) Y_2 \right]^{-1} \left[Y_2' \left(\frac{\bar{Z}\bar{Z}'}{T} - \frac{\bar{Z}_1\bar{Z}_1'}{T} \right) y_1 \right]$$

where $\bar{Z}'\bar{Z} = \Pi_K$ and $\bar{Z}_1'\bar{Z}_1 = \Pi_{K_1}$. We deduce from (3.55) and (3.56):

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 orthonormalized. 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 $\bar{\gamma}_{2SLS} = J_{11}\gamma_{2SLS}$.*

Note that Theorem 3.3.2 refers only to the orthonormalizing transformation and, in view of (3.42), we can write

$$(3.57) \quad \bar{\gamma}_{2SLS} = \left(\frac{Z_1'Z_1}{T} \right)^{1/2} \gamma_{2SLS}.$$

Thus, the γ coefficients and estimators in the transformed system depend on the sample second moments of the exogenous variables.

Note also that there is a simpler and more useful way of writing expression (3.56) for the 2SLS estimator of β . We deduce from (3.43) that

$$\bar{z} \bar{z}' = \bar{z}_1 \bar{z}_1' + \bar{z}_2 \bar{z}_2'$$

so that

$$(3.58) \quad \hat{\beta}_{2SLS} = [Y_2' \bar{z}_2 \bar{z}_2' Y_2]^{-1} [Y_2' \bar{z}_2 \bar{z}_2' y_1]$$

where

$$\begin{aligned} \bar{z}_2 &= Z_2 J_{22}^{-1} - Z_1 J_{11}^{-1} J_{12} J_{22}^{-1} \\ &= \sqrt{T} [Z_2 - Z_1 (Z_1' Z_1)^{-1} Z_1' Z_2] [Z_2' Z_2 - Z_2' Z_1 (Z_1' Z_1)^{-1} Z_1' Z_2]^{-1/2} \\ (3.59) \quad &= \sqrt{T} [(I - P_{Z_1}) Z_2 [Z_2' (I - P_{Z_1}) Z_2]^{-1/2}] \end{aligned}$$

and, as expected, $\bar{z}' \bar{z}_2 = T I_{K_2}$.

For the complete set of standardizing transformations we have:

THEOREM 3.3.3. *The 2SLS estimators of β^* and $\bar{\gamma}^*$ in the standardized model (3.46) are related to the corresponding estimators of β and γ in the unstandardized model (3.3) by the equations*

$$(3.60) \quad \beta_{2SLS}^* = (\ell_{21} - L_{22} \ell_{2SLS}) / \ell_{11} = (\omega_{11} - \omega_{21}' \Omega_{22}^{-1} \omega_{21})^{-1/2} \Omega_{22}^{1/2} (\beta_{2SLS} - \Omega_{22}^{-1} \omega_{21})$$

$$(3.61) \quad \bar{\gamma}_{2SLS}^* = J_{11} \gamma_{2SLS} / \ell_{11} = \left(\frac{Z_1' Z_1}{T(\omega_{11} - \omega_{21}' \Omega_{22}^{-1} \omega_{21})} \right)^{1/2} \gamma_{2SLS} .$$

PROOF. From (3.58) and under the assumption that all standardizing transformations have been carried out, we have

$$(3.62) \quad \beta_{2SLS}^* = (Y_2^* \bar{Z}_2 \bar{Z}_2' Y_2^*)^{-1} (Y_2^* \bar{Z}_2 \bar{Z}_2' y_1^*)$$

$$(3.63) \quad \bar{Y}_{2SLS}^* = (\bar{Z}_1 \bar{Z}_1')^{-1} \bar{Z}_1' (y_1^* - Y_2^* \beta_{2SLS}^*) .$$

From (3.34) and (3.35) we deduce that (3.62) has the alternative form

$$\begin{aligned} \beta_{2SLS}^* &= (L_{22}'^{-1} Y_2^* \bar{Z}_2 \bar{Z}_2' Y_2^* L_{22}^{-1})^{-1} \{L_{22}'^{-1} Y_2^* \bar{Z}_2 \bar{Z}_2' (y_1 - Y_2 L_{22}^{-1} \ell_{21}) / \ell_{11}\} \\ &= (L_{22} \beta_{2SLS} - \ell_{21}) / \ell_{11} \end{aligned}$$

as required. The last line follows because of the invariance of β_{2SLS} to the orthonormalization transformation as already established in Theorem 3.3.2.

Using (3.30) to (3.32) and (3.63) we have:

$$\begin{aligned} \bar{Y}_{2SLS}^* &= T^{-1} \bar{Z}_1' (y_1^* - Y_2^* \beta_{2SLS}^*) \\ &= T^{-1} \bar{Z}_1' (y_1 - Y_2 \beta_{2SLS}) / \ell_{11} \\ &= J_{11} Y_{2SLS} / \ell_{11} \end{aligned}$$

as required, the last line following directly from (3.55). \square

Results that correspond to these for 2SLS can be derived similarly for other estimators. In the case of the general IV estimator given in (3.11) and (3.12) we need only construct an appropriately partitioned block upper triangular matrix J that corresponds to the instrument set H and the remaining variables. To illustrate briefly let us write

$Z = [Z_1 : Z_3 : Z_4]$ where $H = [Z_1 : Z_3]$ and $Z_2 = [Z_3 : Z_4]$. A conformable partition of J is

$$(3.64) \quad J = \begin{bmatrix} J_{11} & J_{13} & J_{14} \\ 0 & J_{33} & J_{34} \\ 0 & 0 & J_{44} \end{bmatrix} = \begin{bmatrix} J_{HH} & J_{H4} \\ 0 & J_{44} \end{bmatrix}.$$

The submatrices of (3.64) can be constructed from the blocks of $Z'Z$ just as in (3.42). The conclusions of Theorems 3.3.1-3.3.3 then apply under this transformation and with the usual reduction (by L^{-1}) to canonical form. The formulae corresponding to (3.62) and (3.63) for the IV estimators are:

$$(3.65) \quad \beta_{IV}^* = [Y_2^* \bar{Z}_3 \bar{Z}_3' Y_2^*]^{-1} [Y_2^* \bar{Z}_3 \bar{Z}_3' y_1^*]$$

where

$$\bar{Z}_3 = \sqrt{T} (I - P_{Z_1}) Z_3 [Z_3' (I - P_{Z_1}) Z_3]^{-1/2}$$

and

$$(3.66) \quad \bar{y}_{IV}^* = (\bar{Z}_1' \bar{Z}_1)^{-1} \bar{Z}_1' (y_1^* - Y_2^* \beta_{IV}^*) = T^{-1} \bar{Z}_1' (y_1^* - Y_2^* \beta_{IV}^*).$$

Turning to the case of the LIML estimator, we note from (3.21) that β_{LIML} minimizes $\beta_{\Delta}' X' Q_{Z_1} X \beta_{\Delta} / \beta_{\Delta}' X' Q_Z X \beta_{\Delta}$. We may write

$$Z(Z'Z)^{-1}Z' = T^{-1} \bar{Z} \bar{Z}', \quad Z_1(Z_1'Z_1)^{-1}Z_1' = T^{-1} \bar{Z}_1 \bar{Z}_1'$$

so that β_{LIML} minimizes the ratio

$$(3.67) \quad \frac{\beta_{\Delta}' X' (I - T^{-1} \bar{Z}_1 \bar{Z}_1') X \beta_{\Delta}}{\beta_{\Delta}' X (I - T^{-1} \bar{Z}_1 \bar{Z}_1') X \beta_{\Delta}}$$

and is therefore invariant under the orthonormalizing transformation J .

Under the canonical transformation L^{-1} (3.67) becomes

$$(3.68) \quad \frac{\beta_{\Delta}' L' L'^{-1} X' Q_{\bar{Z}_1} X L^{-1} L \beta_{\Delta}}{\beta_{\Delta}' L' L'^{-1} X Q_{\bar{Z}_1} X L^{-1} L \beta_{\Delta}} = \frac{\beta_{\Delta}^{*'} X^{*'} Q_{\bar{Z}_1} X^{*} \beta_{\Delta}^{*}}{\beta_{\Delta}^{*'} X^{*'} Q_{\bar{Z}_1} X^{*} \beta_{\Delta}^{*}}$$

where

$$(3.69) \quad \beta_{\Delta}^{*} = L \beta_{\Delta} = \begin{bmatrix} \ell_{11} \\ \ell_{21} - L_{22} \beta \end{bmatrix}.$$

Renormalization of β_{Δ}^{*} gives us the transformed coefficient

$$(3.70) \quad \beta^{*} = (L_{22} \beta - \ell_{21}) / \ell_{11}$$

precisely as in (3.48). Minimization of (3.68) with respect to β^{*} then yields the LIML estimator of the transformed coefficient (3.70)

$$(3.71) \quad \beta_{LIML}^{*} = (L_{22} \beta_{LIML} - \ell_{21}) / \ell_{11}.$$

Properties of the transformed equation

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 (3.46) are worth examining. Let us first write (3.46) in individual observation form as

$$(3.72) \quad y_{1t}^* = y_{2t}^{*'}\beta^* + \bar{z}_{1t}'\bar{\gamma}^* + u_t^* .$$

Then, by simple manipulations we find that:

$$(3.73) \quad \text{cov}(y_{2t}^*, u_t^*) = (\ell_{21} - L_{22}\beta) / \ell_{11} = -\beta^*$$

$$(3.74) \quad \text{var}(u_t^*) = 1 + \left(\frac{\ell_{21} - L_{22}\beta}{\ell_{11}} \right)' \left(\frac{\ell_{21} - L_{22}\beta}{\ell_{11}} \right) = 1 + \beta^{*'}\beta^*$$

and

$$(3.75) \quad \begin{aligned} \text{corr}(y_{2t}^*, u_t^*) &= \mathcal{D}(y_{2t}^*)^{-1/2} \text{cov}(y_{2t}^*, u_t^*) \text{var}(u_t^*)^{-1/2} \\ &= -\beta^* / (1 + \beta^{*'}\beta^*)^{1/2} . \end{aligned}$$

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)² and the graphical analyses by Phillips (1980a, 1982) in the general case.

The vector of correlations corresponding to (3.75) in the unstandardized model is given by

$$\begin{aligned}
\text{corr}(y_{2t}, u_t) &= \mathcal{D}(y_{2t})^{-1/2} \text{cov}(y_{2t}, u_t) \text{var}(u_t)^{-1/2} \\
&= \frac{\Omega_{22}^{-1/2} (\omega_{21} - \Omega_{22}\beta)}{(\omega_{11} - 2\beta'\omega_{21} + \beta'\Omega_{22}\beta)^{1/2}} \\
(3.76) \quad &= \frac{\Omega_{22}^{1/2} (\Omega_{22}^{-1}\omega_{21} - \beta)}{(\omega_{11} - 2\beta'\omega_{21} + \beta'\Omega_{22}\beta)^{1/2}} = -\beta^*/(1 + \beta^*\beta^*)^{1/2}
\end{aligned}$$

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 (3.76), 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 (from (3.58) above)

$$(3.77) \quad \beta_{\text{ILS}} = [\bar{Z}'_2 Y_2]^{-1} [\bar{Z}'_2 y_1]$$

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 form³

$$(3.78) \quad \text{pdf}(r) = \frac{\exp\left\{-\frac{\mu^2}{2}(1+\beta^2)\right\}}{\pi(1+r^2)} {}_1F_1\left(1, \frac{1}{2}; \frac{\mu^2}{2} \frac{(1+\beta r)^2}{1+r^2}\right)$$

where $\mu^2 = \Pi \Pi'_{22} \Pi_{22}$ is the scalar concentration parameter.⁴ The density (3.78) is one of the interesting examples in which maximum likelihood estimators may have bimodal distributions for certain values of the parameters. An interesting numerical analysis of situations in which such bimodality occurs is given by Marsaglia (1965).⁵ In the general case of $n+1$ included endogenous variables the density (3.78) is replaced by a multivariate analogue in which the ${}_1F_1$ function has a matrix argument (see (3.102) below). The category of estimators that take the generic form of a matrix ratio of normal variates as in (3.77) 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 (3.77) occurs in random coefficient SEM's 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 (3.78) to the multivariate case in much the same way as the exact distribution theory of (3.77), which we will detail in paragraph 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 parameterized SEM.⁶ 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.⁷ 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 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.⁸ In order to keep the derivations as simple and brief as possible and to illustrate the main ideas, we will deal with the two endogenous variable case ($n = 1$) in this discussion. The results extend to the $n+1$ endogenous variable case with appropriate generalizations of the method of derivation given here.

The IV estimator

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 system⁹

$$(3.79) \quad y_1 = Z_1 \gamma + u, \quad y_2 = Z_1 \Pi_{12} + v_2.$$

The IV estimator of β is, from (3.65),

$$(3.80) \quad \beta_{IV} = (y_2' Z_3 Z_3' y_2)^{-1} (y_2' Z_3 Z_3' y_1)$$

where we have eliminated the asterisks and bars on the variables in (3.65) for convenience while retaining the assumption that standardizing transformations have been performed. Let Z_3 be $T \times K_3$ with $K_3 \geq 1$ so that the total number of instruments is $K_1 + K_3$. We note that the conditional distribution of $(y_2' Z_3 Z_3' y_1) / (y_2' Z_3 Z_3' y_2)^{1/2}$ given $Z_3' y_2$ is $N(0,1)$. This is independent of $Z_3' y_2$ and is, therefore, the unconditional distribution. We note also that $y_2' Z_3 Z_3' y_2$ has a $\chi_{K_3}^2$ distribution. Hence, we may write

$$(3.81) \quad \beta_{IV} = \frac{(y_2' Z_3 Z_3' y_1) / (y_2' Z_3 Z_3' y_2)^{1/2}}{(y_2' Z_3 Z_3' y_2)^{1/2}} \equiv \frac{N(0,1)}{(\chi_{K_3}^2)^{1/2}}$$

so that β_{IV} is proportional to a t variate with K_3 degrees of freedom. The p.d.f. of β_{IV} is therefore given by

$$(3.82) \quad \text{pdf}(r) = \left[B\left(\frac{1}{2}, \frac{K_3}{2}\right) \right]^{-1} (1+r^2)^{-(K_3+1)/2}$$

where $B(1/2, K_3/2)$ 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 $Q_{Z_1} = I - T^{-1} Z_1 Z_1' = C_1 C_1'$ 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 (3.82) 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 LIML estimator

From (3.21) β_{LIML} minimizes the ratio

$$(3.83) \quad \frac{\beta_{\Delta}' A(I) \beta_{\Delta}}{\beta_{\Delta}' [A(I) - A(P_Z)] \beta_{\Delta}} = 1 + \frac{\beta_{\Delta}' A(P_Z) \beta_{\Delta}}{\beta_{\Delta}' [A(I) - A(P_Z)] \beta_{\Delta}} = 1 + \frac{\beta_{\Delta}' W \beta_{\Delta}}{\beta_{\Delta}' S \beta_{\Delta}}, \text{ say.}$$

Under the same null hypothesis that $\beta = 0$ and $\Pi_{22} = 0$, we see from (3.79) that $W = X'(P_Z - P_{Z_1})X$ and $S = X'(I - P_Z)X$ have central Wishart distributions $\mathcal{W}(K_2, I_m)$ and $\mathcal{W}(T - K, I_m)$, respectively where $m = n+1$. Since $(P_Z - P_{Z_1})(I - P_Z) = 0$, W and S are independent with joint p.d.f. given by (see (2.59)):

$$(3.84) \quad \text{pdf}(W, S) = \frac{\text{etr}\left\{-\frac{1}{2}(W+S)\right\} (\det W)^{(K_2-m-1)/2} (\det S)^{(T-K-m-1)/2}}{2^{m(T-K+K_2)}/2 \Gamma_m\left(\frac{K_2}{2}\right) \Gamma_m\left(\frac{T-K}{2}\right)}.$$

We now transform $W \rightarrow S^{-1/2} W S^{-1/2} = F$ giving the joint density

$$(3.85) \quad \text{pdf}(F, S) = \frac{\text{etr}\left\{-\frac{1}{2}S(I+F)\right\} (\det F)^{(K_2-m-1)/2} (\det S)^{(T-K+K_2-m-1)/2}}{2^{m(T-K_1)}/2 \Gamma_m\left(\frac{K_2}{2}\right) \Gamma_m\left(\frac{T-K}{2}\right)}.$$

Using the following matrix variate gamma integral (Herz (1955))

$$(3.86) \quad \int_{S>0} \text{etr}(-SZ) (\det S)^{a-(m+1)/2} dS = \Gamma_m(a) (\det Z)^{-a}$$

for $\text{Re}(Z) > 0$ and $\text{Re}(a) > (m-1)/2$ we deduce from (3.85) the marginal matrix density of F

$$(3.87) \quad \text{pdf}(F) = \frac{\Gamma_m \left(\frac{T - K_1}{2} \right) (\det F)^{(K_2 - m - 1)/2} [\det(I+F)]^{-(T - K_1)/2}}{\Gamma_m \left(\frac{K_2}{2} \right) \Gamma_m \left(\frac{T - K}{2} \right)}$$

$$= \left[B_m \left(\frac{K_2}{2}, \frac{T - K}{2} \right) \right]^{-1} (\det F)^{(K_2 - m - 1)/2} [\det(I+F)]^{-(T - K_1)/2}$$

which is a multivariate Beta density of the second kind (see, for example, Tan (1969)).

Expression (3.87) now leads simply to the p.d.f. of the LIML estimator β_{LIML} . We note that if λ_m is the smallest latent root of $F = S^{-1/2}WS^{-1/2}$ it follows from (3.83) (or more directly from (3.22) given earlier) that β_{LIML} satisfies the system

$$(3.88) \quad (S^{-1/2}WS^{-1/2} - \lambda_m I)\beta_{\Delta} = (F - \lambda I)\beta_{\Delta} = 0$$

where λ_m is the smallest latent vector of the matrix F , β_{Δ} is its associated latent vector and $\beta_{\text{LIML}} = -\beta_{\Delta 2}/\beta_{\Delta 1}$ where $\beta_{\Delta 1}$ is the first component of β_{Δ} .

In order to extract the exact distribution of β_{LIML} we introduce the orthogonal transformation H by which F is diagonalized so that $H'FH = \Lambda = \text{diag}(\lambda_1, \dots, \lambda_n)$. This transformation is unique if we specify that $\lambda_1 > \lambda_2 > \dots > \lambda_m$ and that the elements in the first row of H are positive. (The latter eliminates the possibility of multiplying columns of H by -1 .) From (3.87) we can deduce the joint distribution of (Λ, H) . It is most convenient in the general multivariate case to work with the probability element $\text{pdf}(F)dF$. Under the transformation $F \rightarrow (\Lambda, H)$ we then have

$$\text{pdf}(\Lambda, H) d\Lambda(dH) = \text{pdf}(F) dF$$

$$(3.89) \quad = \text{pdf}(F) \prod_{i < j} (\lambda_i - \lambda_j) \left(\prod_{i=1}^m d\lambda_i \right) (dH)$$

$$(3.90) \quad = \left[B_n \left(\frac{K_2}{2}, \frac{T-K}{2} \right) \right]^{-1} \prod_{i=1}^m \lambda_i^{(K_2-m-1)/2} \\ \cdot \left[\prod_{i=1}^m (1 + \lambda_i) \right]^{-1} \prod_{i < j} (\lambda_i - \lambda_j) \left(\prod_{i=1}^m d\lambda_i \right) (dH) .$$

Line (3.89) follows from Constantine (1963, equation (43), p. 1280) and (dH) is the invariant measure on the orthogonal group (the group of orthogonal matrices) that is normalized so that the measure over the whole group (restricted so that $h_{ij} \geq 0$) is unity. The distribution of H is called the conditional Haar invariant distribution (see Anderson (1958, p. 322)) and we see from (3.90) that H is distributed independently of the latent roots that form the diagonal elements of Λ . To find the distribution of β_{LIML} we now need to concentrate on the final column of H.

It is most instructive to take the simple two endogenous variable case ($m = 2$, $n = 1$) as for the IV estimator above. First we make explicit the form of the orthogonal transformation H :

$$(3.91) \quad H = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix}, \quad \frac{3\pi}{2} \leq \theta \leq 2\pi \\ = \begin{bmatrix} -\sin \theta & \cos \theta \\ \cos \theta & \sin \theta \end{bmatrix}, \quad \frac{-\pi}{2} \leq \theta \leq 0 .$$

There is now only one free variable, θ , in this representation of H (note that we have restricted the domain of θ so that $h_{ij} \geq 0$ applies)

and the invariant measure (dH) is simply

$$(3.92) \quad (dH) = \frac{1}{\pi} d\theta, \quad \theta \in [-\pi/2, 0] \cup [3\pi/2, 2\pi].$$

In this case the transformation $F \rightarrow (\Lambda, H)$ becomes $F \rightarrow (\lambda_1, \lambda_2, \theta)$.

As in the multivariate case (λ_1, λ_2) and θ are statistically independent.

Moreover, from (3.90) and (3.92) we see that θ is uniformly distributed over the intervals $[-\pi/2, 0]$, $[3\pi/2, 2\pi]$. We deduce that

$$(3.93) \quad \text{LIML} = \frac{-\beta_{\Delta 2}}{\beta_{\Delta 1}} = \begin{cases} \frac{\cos \theta}{\sin \theta} = \cot \theta, & 3\pi/2 \leq \theta \leq 2\pi \\ \frac{-\sin \theta}{\cos \theta} = -\tan \theta, & -\pi/2 \leq \theta \leq 0 \end{cases}$$

The density function of β_{LIML} is therefore given by

$$(3.94) \quad \text{pdf}(r) = \frac{1}{\pi} \left| \frac{d}{dr} \cot^{-1}(r) \right|, \quad r \leq 0$$

$$= \frac{1}{\pi} \left| \frac{d}{dr} (-\tan^{-1}(r)) \right|, \quad r \geq 0$$

$$(3.95) \quad = [\pi(1+r^2)]^{-1}, \quad -\infty < r < \infty.$$

Thus, the exact sampling distribution of the LIML estimator is Cauchy in this leading case. In fact (3.95) 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 (3.95) that β_{LIML} has no finite moments of integral order, as was shown by Mariano and Sawa (1972) and Sargan (1970).

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

$$(3.96) \quad \beta_{IV} = (Y_2' Z_3 Z_3' Y_2)^{-1} (Y_2' Z_3 Z_3' y_1)$$

as shown in (3.65) and where the standardizing transformations are assumed to have been carried out. This is the case discussed in paragraph 3.3 where $H = [Z_1 : Z_3]$ is the 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 noncentral Wishart with a p.d.f. of the form:

$$(3.97) \quad \text{pdf}(A) = \frac{\text{etr} \left\{ -\frac{1}{2} M M' \right\}}{2^{\frac{1}{2} K_3 (n+1)} \Gamma_{n+1} \left(\frac{K_3}{2} \right)} {}_0F_1 \left(\frac{K_3}{2}; \frac{1}{4} M M' A \right) \text{etr} \left\{ -\frac{1}{2} A \right\} (\det A)^{\frac{1}{2} (K_3 - n - 2)}$$

(see (2.60) 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

$$M' = T^{-1/2} Z_3' Z_2 \Pi = T^{-1/2} [O S'] \Pi = T^{-1/2} S' [\pi_{21} \ \pi_{22}] = T^{-1/2} S' \Pi_{22} [\beta, I]$$

in view of the relations (3.6) given earlier. 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 (1) is assumed to be identified), we find that

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

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

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

(3.97) becomes

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

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

$$\begin{aligned}
\text{pdf}(w, r, A_{22}) &= \frac{\text{etr}\left\{-\frac{T}{2}(I + \beta\beta')\bar{\pi}'_{22}\bar{\pi}_{22}\right\}}{2^{\frac{1}{2}K_3(n+1)} \Gamma_{n+1}\left(\frac{K_3}{2}\right)} \\
&\cdot {}_0F_1\left\{\frac{K_3}{2}; \frac{T}{4}\left[w\bar{\pi}_{22}\beta\beta'\bar{\pi}'_{22} + \bar{\pi}_{22}(I + \beta r')A_{22}(I + r\beta')\bar{\pi}'_{22}\right]\right\} \\
&\cdot \exp\left\{-\frac{1}{2}(w+r'A_{22}r)\right\} \text{etr}\left\{-\frac{1}{2}A_{22}\right\} \left\{w \det A_{22}\right\}^{\frac{1}{2}(K_3-n-2)} \det A_{22} \\
&= \frac{\text{etr}\left\{-\frac{T}{2}(I + \beta\beta')\bar{\pi}'_{22}\bar{\pi}_{22}\right\}}{2^{\frac{1}{2}K_3(n+1)} \Gamma_{n+1}\left(\frac{K_3}{2}\right)} \\
&\cdot {}_0F_1\left\{\frac{K_3}{2}; \left\{\frac{T}{4} w\bar{\pi}_{22}\beta\beta'\bar{\pi}'_{22} + \bar{\pi}_{22}(I + \beta r')A_{22}(I + r\beta')\bar{\pi}'_{22}\right\}\right\} \\
&\cdot \exp\left\{-\frac{1}{2}w\right\} \text{etr}\left\{-\frac{1}{2}(I + rr')A_{22}\right\} w^{\frac{1}{2}(K_3-n-2)} (\det A_{22})^{\frac{1}{2}(K_3-n)}.
\end{aligned}$$

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

$$\begin{aligned}
(3.98) \quad \text{pdf}(w, r, B) &= \frac{\text{etr}\left\{-\frac{T}{2}(I + \beta\beta')\bar{\pi}'_{22}\bar{\pi}_{22}\right\}}{2^{\frac{1}{2}(L+n)(n+1)} \Gamma_{n+1}\left(\frac{L+n}{2}\right) [\det(I + rr')]^{(L+n+1)/2}} \\
&\cdot {}_0F_1\left[\frac{L+n}{2}; \frac{T}{4}\left[w\bar{\pi}_{22}\beta\beta'\bar{\pi}'_{22} + \bar{\pi}_{22}(I + \beta r')(I + rr')^{-1/2}\right.\right. \\
&\left.\left.\cdot B(I + rr')^{-1/2}(I + \beta r')\bar{\pi}'_{22}\right]\right] \exp\left\{-\frac{1}{2}w\right\} \text{etr}\left\{-\frac{1}{2}B\right\} w^{L/2-1} (\det B)^{L/2}.
\end{aligned}$$

As in the general scheme of development outlined in paragraph 2.1 we

need to integrate out the auxiliary variates (w, B) in (3.98) 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 important 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 (3.98) can be replaced by a power series in the argument

$$(3.99) \quad \frac{1}{4} \left\{ w \bar{\pi}_{22} \beta \beta' \bar{\pi}'_{22} + \bar{\pi}_{22} (I + B r') (I + r r')^{-1/2} B (I + r r')^{-1/2} (I + r s') \bar{\pi}'_{22} \right\}$$

which, when $n = 1$, is a scalar. Powers of this variable may now be expanded in binomial series and inspection of (3.98) shows that terms of this double series may then be integrated simply as gamma functions. Details of this argument will be given in the following paragraph 3.6.

When $n > 1$, (3.99) 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 (3.98) 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 (3.99) is thrown up into an exponent. The two elements of the binomial matrix sum (3.99) 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.)¹⁰ In short, the process leads to the following analytic form for the exact finite sample density of β_{IV} :

$$\begin{aligned}
(3.100) \quad \text{pdf}(r) &= \frac{\text{etr}\left\{-\frac{T}{2}(I + \beta\beta')\bar{\Pi}'_{22}\bar{\Pi}_{22}\right\} \Gamma_n\left(\frac{L+n+1}{2}\right)}{\pi^{n/2} [\det(I + rr')]^{(L+n+1)/2}} \\
&\cdot \sum_{j=0}^{\infty} \frac{\binom{L}{2}_j}{j! \Gamma_n\left(\frac{L+n}{2} + j\right)} \left[\left[\frac{T}{2} \beta' \bar{\Pi}'_{22} \left(\text{adj} \frac{\partial}{\partial W} \right) \bar{\Pi}_{22} \beta \right]^j (\det(I+W))^{(L-1)/2+j} \right. \\
&\cdot \left. {}_1F_1\left(\frac{L+n+1}{2}, \frac{L+n}{2} + j; \frac{T}{2}(I+W)\bar{\Pi}'_{22}(I + \beta r')(I + rr')^{-1}(I + r\beta')\bar{\Pi}_{22}\right) \right]_{W=0}.
\end{aligned}$$

In (3.100) $L = K_3 - n$ 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 $L = K_2 - n$, (3.100) 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 $L = T - K_1 - n$, (3.100) gives the p.d.f. of the OLS estimator of β .

The matrix W ($n \times n$) in (3.100) contains auxiliary variables that are useful in reducing the integral from which (3.100) 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 (3.100) becomes

$$\begin{aligned}
(3.101) \quad \text{pdf}(r) &= \frac{\exp\left\{-\frac{\mu^2}{2}(1 + \beta^2)\right\} \Gamma\left(\frac{L+2}{2}\right)}{\pi^{1/2} \Gamma\left(\frac{L+1}{2}\right) (1 + r^2)^{(L+2)/2}} \\
&\cdot \sum_{j=0}^{\infty} \frac{\binom{L}{2}_j}{j! \binom{L+1}{2}_j} \left(\frac{\mu^2}{2} \beta^2\right)^j {}_1F_1\left(\frac{L+2}{2}, \frac{L+1}{2} + j; \frac{\mu^2}{2} \frac{(1 + \beta r)^2}{1 + r^2}\right)
\end{aligned}$$

in which $\mu^2 = T\bar{\Pi}'_{22} = T\Pi'_{22}\Pi_{22}$ is the scalar concentration parameter (recall (3.78) and footnote 4). The density (3.101) 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 (3.100) the series corresponding to the suffix j terminates at the first term and we have

$$(3.102) \quad \text{pdf}(r) = \frac{\text{etr}\left\{-\frac{T}{2}(I + \beta\beta')\bar{\pi}'_{22}\bar{\pi}_{22}\right\} \Gamma_n\left(\frac{n+1}{2}\right)}{\pi^{n/2} \Gamma_n\left(\frac{n}{2}\right) [\det(I + rr')]^{(n+1)/2}} \\ \cdot {}_1F_1\left(\frac{n+1}{2}; \frac{n}{2}; \frac{T}{2}\bar{\pi}'_{22}(I + \beta r')(I + rr')^{-1}(I + r\beta')\bar{\pi}'_{22}\right).$$

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

While (3.100) 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 (c.f. 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 (3.100) is not by itself very revealing. Moreover series representations such as (3.100) and (3.102) 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 (2.61)). This is an area in which important developments are currently taking place (some discussion and references are given in Section 2 following (2.61)). Unfortunately, the availability of tabulations and algorithms for zonal-type polynomials¹¹ 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 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 parameter (that is, β in (3.101)) is greater than 5 and the concentration parameter (that is, ν^2 in (3.101)) 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 (3.101), 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 11).

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 paragraph 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 multi-dimensional 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 (see paragraph 2.5 above) 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 paragraph 3.6. In the following we report briefly the principal results which apply in the multi-endogenous variable cases investigated by Phillips (1980a, 1982a):

- (i) For comparable parameter values the marginal distributions of β_{IV} appear to concentrate more slowly as $T \rightarrow \infty$ when the number of endogenous variables $(n+1)$ in the equation increases.
- (ii) The marginal densities are particularly sensitive to the degree of correlation in the concentration parameter matrix $\bar{M} = T\bar{\Pi}'_{22}\bar{\Pi}_{22}$ in (3.100). Setting, for example,

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

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 \bar{M} approaches singularity as $|\rho| \rightarrow 1$ when the equation becomes unidentifiable (Π_{22} in (3.5) and hence $\bar{\Pi}_{22}$ 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 ν^2); but when β_1 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.

- (iii) 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.
- (iv) 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 (3.101) the general form of the joint density (3.100) 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 (3.101) 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 (3.98). For the special case in which $n = 1$ the joint density (3.98) becomes (setting $B = b$ and $u^2 = T\bar{\pi}_{22}^2$):

$$\begin{aligned}
 \text{pdf}(w, r, b) &= \frac{\exp\left\{-\frac{u^2}{2}(1+\beta^2)\right\} w^{L/2-1} \exp\left(-\frac{1}{2}w\right)}{2^{(L+1)} \Gamma_2\left(\frac{L+1}{2}\right) (1+r^2)^{(L+2)/2}} \\
 &\quad \cdot b^{L/2} \exp\left(-\frac{1}{2}b\right) \sum_{k=0}^{\infty} \frac{\left(\frac{1}{2}\right)^k \left[\frac{u^2}{2}\beta^2 w + \frac{1}{2}u^2 \frac{(1+\beta r)^2}{1+r^2} b\right]^k}{k! \left(\frac{L+1}{2}\right)_k} \\
 &= \frac{\exp\left\{-\frac{u^2}{2}(1+\beta^2)\right\} \exp\left(-\frac{1}{2}w\right) \exp\left(-\frac{1}{2}b\right)}{2^{L+1} \Gamma_2\left(\frac{L+1}{2}\right) (1+r^2)^{(L+2)/2}} w^{L/2-1} b^{L/2} \\
 &\quad \cdot \sum_{k=0}^{\infty} \sum_{j=0}^k \frac{\left(\frac{1}{2}\right)^k \binom{k}{j}}{k! \left(\frac{L+1}{2}\right)_k} \left(\frac{u^2}{2}\beta^2 w\right)^j \left(\frac{u^2}{2} \frac{(1+\beta r)^2}{1+r^2}\right)^{k-j} \\
 (3.103) \quad &= \frac{\exp\left\{-\frac{u^2}{2}(1+\beta^2)\right\} \exp\left(-\frac{1}{2}w\right) \exp\left(-\frac{1}{2}b\right)}{2^{L+1} \pi^{1/2} \Gamma\left(\frac{L}{2}\right) \Gamma\left(\frac{L+1}{2}\right) (1+r^2)^{(L+2)/2}} \\
 &\quad \cdot \sum_{\ell=0}^{\infty} \sum_{j=0}^{\infty} \frac{\left(\frac{1}{2}\right)^{\ell+j}}{\left(\frac{L+1}{2}\right)_{\ell+j} \ell! j!} \left(\frac{u^2}{2}\beta^2 w\right)^j \left(\frac{u^2}{2} \frac{(1+\beta r)^2}{1+r^2}\right)^{\ell} w^{L/2+j-1} b^{L/2+\ell} .
 \end{aligned}$$

The final equality in (3.103) follows by changing the index of summation from k to $\ell = k-j$ and by using the relation $\Gamma_2(z) = \pi^{1/2} \Gamma(z) \Gamma(z-1/2)$ (see paragraph 2.7). The elementary gamma integral

$$(3.104) \quad \int_0^{\infty} e^{-at} t^{z-1} dt = \Gamma(z)/a^z, \quad \text{Re}(a) > 0, \quad \text{Re}(z) > 0$$

now gives us the density of θ_T by integrating out w and b over the regions $0 \leq w < \infty$ and $0 \leq b < \infty$. We obtain

$$(3.105) \quad \text{pdf}(r) = \frac{\exp\left\{-\frac{v^2}{2}(1+\beta^2)\right\}}{2^{L+1} \pi^{1/2} \Gamma\left(\frac{L}{2}\right) \Gamma\left(\frac{L+1}{2}\right) (1+r^2)^{(L+2)/2}}$$

$$\cdot \sum_{\ell=0}^{\infty} \sum_{j=0}^{\infty} \frac{\left(\frac{1}{2}\right)^{\ell+j}}{\left(\frac{L+1}{2}\right)_{\ell+j} \ell! j!} \left(\frac{v^2 \beta^2}{2}\right)^j \left(\frac{v^2(1+\beta r)^2}{1+r^2}\right)^{\ell} \frac{\Gamma\left(\frac{L}{2}+j\right) \Gamma\left(\frac{L}{2}+\ell+1\right)}{\left(\frac{1}{2}\right)^{L/2+j} \left(\frac{1}{2}\right)^{L/2+\ell+1}}$$

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

$$\cdot \sum_{\ell=0}^{\infty} \sum_{j=0}^{\infty} \frac{\left(\frac{L}{2}\right)_j}{\left(\frac{L+1}{2}\right)_j j!} \left(\frac{v^2 \beta^2}{2}\right)^j \frac{\left(\frac{L+2}{2}\right)_{\ell}}{\ell! \left(\frac{L+1}{2}+j\right)_{\ell}} \left(\frac{v^2(1+\beta r)^2}{1+r^2}\right)^{\ell}$$

$$(3.106) \quad = \frac{\exp\left\{-\frac{v^2}{2}(1+\beta^2)\right\} \Gamma\left(\frac{L+2}{2}\right)}{\pi^{1/2} \Gamma\left(\frac{L+1}{2}\right) (1+r^2)^{(L+2)/2}} \sum_{j=0}^{\infty} \frac{\left(\frac{L}{2}\right)_j}{j! \left(\frac{L+1}{2}\right)_j} \left(\frac{v^2 \beta^2}{2}\right)^j$$

$$\cdot {}_1F_1\left[\frac{L+2}{2}, \frac{L+1}{2}+j; \frac{v^2(1+\beta r)^2}{1+r^2}\right]$$

as given in (3.101). The integration term by term that leads to (3.105) is justified because the series in (3.103) is absolutely convergent, which enables us to reverse the order of summation and integration (cf. Lebedev (1972, p. 239)). As indicated earlier, (3.106) 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 Appendix A of Phillips (1980b) where the inversion (of the form (2.7)) 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 (see paragraph 3.4) but now the joint density pdf(W,S) in (3.84) is noncentral. This joint density is a product of independent Wishart densities with different degrees of freedom (K_2 and $T-K$ respectively) and a noncentrality parameter matrix closely related to that which applies in the case of the IV estimator analyzed in paragraph 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 noncentrality 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).¹² The parameterization of the LIML density is different from that of the IV density given earlier. In particular, the relevant parameters of (3.106) 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 (3.19) 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 density for $0 \leq k \leq 1$ 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:

- (i) 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 (3.106) 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 $K_2 = 3$ 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 $\beta = 1$, $\mu^2 = 100$, $K_2 = 30$ the interquartile range (measured again in terms of asymptotic standard deviations) is 1.031 whereas at $\beta = 1$, $\mu^2 = 100$, $K_2 = 3$ the interquartile range is 1.321. The following table illustrates how these effects are magnified as β increases:

TABLE 1¹³

Median (MDN) and Interquartile Range (IQR)
of $\beta_{2SLS} - \beta$ in terms of asymptotic standard deviations

($\nu^2 = 100$)

$K_2 \backslash \beta$		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

- (ii) 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 ν^2 must be at least 100 to hold the maximum error on the asymptotic normal approximation to 0.05; but when $K_2 = 10$, ν^2 must be at least 3,000 to ensure the same maximum error on the asymptotic distribution.
- (iii) Since the exact distribution of β_{LIML} involves a triple infinite series, Anderson and Sawa (1977, 1978) tabulated the distribution of a closely related estimator known as LIMLK. This estimator represents what would be the LIML estimator if the covariance matrix of the reduced form errors were known. In terms of (3.83) and (3.88), 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 noncentral 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. The following table of selected computations from Anderson and Sawa (1978) illustrates these effects in relation to the corresponding results for 2SLS in Table 1.

TABLE 2^{13,14}

Median and Interquartile range of $B_{LIMLK} - \beta$
in terms of asymptotic standard deviations

($\mu^2 = 100$)

$K_2 \backslash \beta$		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

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 B_{2SLS} may be badly located and skewed unless β and K_2 are small or μ^2 is very large; (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 at length in paragraph 3.3. When we referred to the true coefficient β above, we therefore meant the true standardized coefficient (as given by β^* in expression (3.48) 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^* viz. $\text{corr}(y_t, u_t) = -\beta^*/(1+\beta^{*2})^{1/2}$ as given by (3.76)

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 1 and 2 illustrate.

3.7. Structural Variance Estimators

In paragraphs 3.3 to 3.6 our attention has focussed 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):

$$(3.107) \quad G_1(\beta) = \beta'_\Delta X Q_{Z_1} X \beta_\Delta = \beta'_\Delta A(I) \beta_\Delta,$$

$$(3.108) \quad G_2(\beta) = \beta'_\Delta X' Q_Z X \beta_\Delta = \beta'_\Delta [A(I) - A(P_Z)] \beta_\Delta,$$

$$(3.109)^{15} \quad Q(\beta) = G_1(\beta) - G_2(\beta) = \beta'_\Delta X' (P_Z - P_{Z_1}) X \beta_\Delta = \beta'_\Delta A(P_Z) \beta_\Delta.$$

Corresponding to (3.107) we have the most common structural error variance estimators, viz.

$$(3.110) \quad \begin{aligned} \sigma_{IV} &= T^{-1} G_1(\beta_{IV}) = T^{-1} (y_1 - Y_2 \beta_{IV})' Q_{Z_1} (y_1 - Y_2 \beta_{IV}) \\ &= T^{-1} (y_1 - Y_2 \beta_{IV} - Z_1 \gamma_{IV})' (y_1 - Y_2 \beta_{IV} - Z_1 \gamma_{IV}) \end{aligned}$$

and

$$\begin{aligned}
 (3.111) \quad \sigma_{LIML} &= T^{-1} G_1(\beta_{LIML}) = T^{-1} (y_1 - Y_2 \beta_{LIML})' Q_{Z_1} (y_1 - Y_2 \beta_{LIML}) \\
 &= T^{-1} (y_1 - Y_2 \beta_{LIML} - Z_1 \gamma_{LIML})' (y_1 - Y_2 \beta_{LIML} - Z_1 \gamma_{LIML})
 \end{aligned}$$

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 (3.110)-(3.111) to the required marginal densities. The analytic derivations relate very closely to those that apply in the distribution theory for the relevant coefficient estimator, which we have already discussed in paragraphs 3.4-3.6. 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 (3.107)-(3.109), and by McDonald (1972), who found the exact p.d.f.'s of the LIML variance estimators for the same trinity of quadratic forms (3.107)-(3.109). Some of their principal findings can otherwise be obtained by examination of the leading case considered earlier in paragraph 3.4.

We start our analysis by noting that the standardizing transformations of paragraph 3.3 (see (3.3) and (3.74)) imply the relationship

$$\sigma^2 = \text{var}(u_t) = \lambda_{11}^2 \text{var}(u_t^*) = (\omega_{11} - \omega_{21}' \Omega_{22}^{-1} \omega_{21}) \text{var}(u_t^*)$$

between the variances of the errors in the unstandardized and standardized (denoted by $*$) model, respectively. We also have, from (3.110)-(3.111) the corresponding relationships for the two estimators:

$$(3.112) \quad \sigma_{IV} = \epsilon_{11}^2 T^{-1} (y_1^* - Y_2^* \beta_{IV}^*)' Q_{Z_1} (y_1^* - Y_2^* \beta_{IV}^*) = \epsilon_{11}^2 \sigma_{IV}^*,$$

$$(3.113) \quad \sigma_{LIML} = \epsilon_{11}^2 T^{-1} (y_1^* - Y_2^* \beta_{LIML}^*)' Q_{Z_1} (y_1^* - Y_2^* \beta_{LIML}^*) = \epsilon_{11}^2 \sigma_{LIML}^*.$$

This correspondence can be used to deduce the exact densities of σ_{IV} and σ_{LIML} from their standardized forms in terms of σ_{IV}^* and σ_{LIML}^* . In what follows we will, as usual, work in standardized form and drop the $*$ for notational convenience.

We take the leading null hypothesis $\beta = 0$ and $\Pi_{22} = 0$ of paragraph 3.4 and the two endogenous variable case ($n = 1$) to illustrate the main derivations and results as simply as possible. Under the null hypothesis and conditional on β_{IV} , σ_{IV} is an idempotent quadratic form in independent normal variates with variance equal to $1 + \beta_{IV}^2$. Thus, conditional on β_{IV} , $T\sigma_{IV}/(1 + \beta_{IV}^2)$ is $\chi_{T-K_1}^2$, so that

$$(3.114) \quad \begin{aligned} \text{pdf}(\sigma_{IV} | \beta_{IV} = r) &= T(1+r^2)^{-1} \text{pdf}(\chi_{T-K_1}^2) \\ &= \frac{T(1+r^2)^{-1}}{2^{(T-K_1)/2} \Gamma\left(\frac{T-K_1}{2}\right)} \left(\frac{T\sigma_{IV}}{1+r^2}\right)^{(T-K_1)/2-1} \exp\left\{-\frac{1}{2} \frac{T\sigma_{IV}}{1+r^2}\right\}. \end{aligned}$$

It follows from (3.114) and the marginal density, (3.82), of β_{IV} that

$$(3.115) \quad \text{pdf}(\sigma_{IV}) = \frac{T(T\sigma_{IV})^{(T-K_1)/2-1}}{2^{(T-K_1)/2} \Gamma\left(\frac{T-K_1}{2}\right) B\left(\frac{1}{2}, \frac{K_3}{2}\right)} \int_{-\infty}^{\infty} \frac{\exp\left\{-\frac{1}{2} \frac{T\sigma_{IV}}{1+r^2}\right\} dr}{(1+r^2)^{(T-K_1+K_3+1)/2}}$$

$$= \frac{(T/2)^{(T-K_1)/2} \sigma_{IV}^{(T-K_1)/2-1} e^{-T\sigma_{IV}/2}}{\Gamma\left(\frac{T-K_1}{2}\right) B\left(\frac{1}{2}, \frac{K_3}{2}\right)} \int_{-\infty}^{\infty} \frac{\exp\left\{\frac{1}{2} \frac{T\sigma_{IV} r^2}{1+r^2}\right\} dr}{(1+r^2)^{(T-K_1+K_3+1)/2}}$$

$$(3.116) = \frac{(T/2)^{(T-K_1)/2} \sigma_{IV}^{(T-K_1)/2-1} e^{-T\sigma_{IV}/2}}{\Gamma\left(\frac{T-K_1}{2}\right) B\left(\frac{1}{2}, \frac{K_3}{2}\right)} \sum_{j=0}^{\infty} \frac{\left(\frac{T\sigma_{IV}}{2}\right)^j}{j!} \int_{-\infty}^{\infty} \frac{r^{2j} dr}{(1+r^2)^{(T-K_1+K_3+1)/2+j}}$$

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

$$(3.117)^{16} = \frac{TB\left(\frac{1}{2}, \frac{T-K_1+K_3}{2}\right) e^{-T\sigma_{IV}/2}}{2\Gamma\left(\frac{T-K_1}{2}\right) B\left(\frac{1}{2}, \frac{K_3}{2}\right)} \sum_{j=0}^{\infty} \frac{\left(\frac{1}{2}\right)_j \left(\frac{T\sigma_{IV}}{2}\right)^j}{\left(\frac{T-K_1+K_3+1}{2}\right)_j j!} \cdot$$

The term by term integration in (3.116) is justified by the absolute and uniform convergence of the series, which enables us to reverse the order of summation and integration. The density (3.117) was first given by Basmann (1974) for the general case of 2SLS.¹⁷

Expression (3.117) 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 (3.95) and this is just the special case

of (3.82) when $K_3 = 1$.

Analysis of (3.115) and (3.117) shows that moments, $E(\sigma_{IV}^h)$, 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 β_{2SLS} minimizes $Q(\beta)$ in (3.109) (see (3.18)) and since $Q(\beta)$ is proportional to a $\chi_{K_2}^2$ variate when β takes on the true value of the coefficient vector, the structural variance estimator $\hat{\sigma}_{2SLS} = Q(\beta_{2SLS})/(K_2 - n)$ possesses finite moments of all orders. However, $\hat{\sigma}_{2SLS}$ (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 $\hat{\sigma}_{2SLS}$ has a limiting $\chi_{K_2-1}^2$ distribution as the concentration parameter $\mu^2 \rightarrow \infty$. The errors involved in this asymptotic distribution of $\hat{\sigma}_{2SLS}$ 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 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 in earlier paragraphs of this section 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 along the lines of paragraph 2.3 and then 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, 1976,

1980) and will be considered more closely by Rothenberg (1982) in Volume 2 of this Handbook.

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.

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).¹⁸ 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 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

$$(3.118) \quad F = \frac{T-K}{K_2} \frac{y^{*'}(Q_{Z_1} - Q_Z)y^*}{y^{*'}Q_Z y^*}$$

and has an $F_{K_2, T-K}$ distribution under H_0 . When H_0 is false, the denominator of (3.118) is still proportional to a χ_{T-K}^2 variate while the numerator becomes noncentral $\chi_{K_2}^2$ with the noncentrality dependent on the vector of coefficient inaccuracy under the null, $\beta - \beta_0$, and a subset of the reduced form parameters. Thus (3.118) is noncentral $F_{K_2, T-K}$ 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 (3.118) is consistent and its power function was considered briefly by Revankar and Mallela (1972). Confidence regions follow from (3.119) in the usual way as the set of all β satisfying the inequality

$$(3.119) \quad \frac{(y_1 - Y_2\beta)'(Q_{Z_1} - Q_Z)(y_1 - Y_2\beta)}{(y_1 - Y_2\beta)'Q_Z(y_1 - Y_2\beta)} \leq \frac{K_2}{T-K} F_{K_2, T-K}(\alpha)$$

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

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 SFM's. 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 $\mu^2 \rightarrow \infty$. 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 paragraph 3.4. The Richardson-Rohr structural t-statistic is given by

$$(3.120) \quad t = (y_2' Z_3 Z_3' y_2)^{1/2} \beta_{IV} / s$$

where

$$(3.121) \quad s^2 = \frac{1}{K_3 - 1} Q(\beta_{IV}) = \frac{1}{K_3 - 1} e'_{\Delta IV} X' Z_3 Z_3' X e_{\Delta IV}$$

in the notation of (3.109).¹⁹ We note that (3.120) is based on a less common estimator of the structural variance, viz (3.121) (compare the more usual estimator σ_{IV} in (3.110)). Simple manipulations show that S^2 can be rewritten as

$$(3.122) \quad S^2 = \frac{1}{K_3 - 1} y_1' Z_3' M Z_3 y_1, \quad M = I - Z_3' y_2 y_2' Z_3 / y_2' Z_3 Z_3' y_2.$$

We introduce the orthogonal $K_3 \times K_3$ matrix C whose first column is $Z_3' y_2 / (y_2' Z_3 Z_3' y_2)^{1/2}$ and whose remaining columns are the latent vectors of M corresponding to the unit latent roots of this matrix. Conditional on $Z_3' y_2$, we see that $z = CZ_3' y_1$ is $N(0, I_{K_3})$ and since this is independent of $Z_3' y_2$ it is also the unconditional distribution. Moreover, we may write M as

$$M = C \begin{bmatrix} 0 & | & 0 \\ \hline 0 & | & I_{K_3-1} \end{bmatrix} C'$$

and then (3.120) takes the form:

$$(3.123) \quad t = \frac{z_1}{\left\{ \sum_{j=2}^{K_3} z_j^2 / (K_3 - 1) \right\}^{1/2}} \equiv \frac{N(0,1)}{\left\{ \chi_{K_3-1}^2 / (K_3 - 1) \right\}^{1/2}}.$$

Since numerator and denominator in (3.123) are independent, we deduce that this has a Student's t -distribution with $K_3 - 1$ degrees of freedom. In the case of 2SLS considered by Richardson and Rohr (1971) $K_3 - 1 = K_2 - 1 =$ degree of overidentification of the structural equation and it is assumed that $K_2 - 1 \geq 1$.

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 (3.109). This estimator relies on linear forms in the data such as $Z_3'X$ (compare (3.122) above) and does not involve the sample second moments $X'X$ directly, as do the more conventional estimators σ_{IV} and σ_{LIML} in (3.110)-(3.111). 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.

Identifiability Test Statistics

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

$$(3.124) \quad y_1 = Y_2\beta + Z_1\gamma_1 + Z_2\gamma_2 + u$$

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

$$(3.125) \quad H_0 : \gamma_2 = 0 .$$

It is usually assumed that $K_2 \geq n+1$ so that attention is focussed on the overidentifying restrictions in (3.124). 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 (3.107)-(3.109) studied in connection with structural variance estimators. Explicitly we have:

$$(3.126) \quad \hat{\phi}_{2SLS} = \frac{G_1(\beta_{2SLS}) - G_2(\beta_{2SLS})}{G_2(\hat{\beta}_{2SLS})} = \frac{Q(\beta_{2SLS})}{G_2(\beta_{2SLS})}$$

$$(3.127) \quad \hat{\phi}_{LIML} = \frac{G_1(\beta_{LIML}) - G_2(\beta_{LIML})}{G_2(\hat{\beta}_{LIML})} = \frac{Q(\beta_{LIML})}{G_2(\beta_{LIML})} .$$

If the identifiability hypothesis (3.125) is valid, the limiting distributions of $T\hat{\phi}_{2SLS}$ and $T\hat{\phi}_{LIML}$ are both $\chi^2_{K_2-n}$ as $T \rightarrow \infty$ under fairly general conditions. These asymptotic results were derived by Anderson and Rubin (1950), Hood and Koopmans (1953) and Basmann (1960a) and are reported by Hausmann (1982) in this volume of the Handbook. They lead to the common critical region (of rejection) in a large sample test of identifying restrictions:

$$(3.128) \quad T\hat{\phi} \geq \chi^2_{K_2-n}(\alpha)$$

where α is the chosen significance level and $\hat{\phi}$ denotes either $\hat{\phi}_{2SLS}$ or $\hat{\phi}_{LIML}$.

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

$$(3.129) \quad \frac{T-K}{K_2} \phi_{LIML} \geq F_{K_2, T-K}(\alpha) .$$

This may be justified on the argument that for fixed β in (3.127) the ratio $(T-K)\phi/K_2$ is indeed distributed as $F_{K_2, T-K}$ (compare (3.118) above). Basmann (1960) criticized this suggestion on the grounds that as $T \rightarrow \infty$

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

whereas

$$(T-K)F_{K_2, T-K} \xrightarrow{D} \chi_{K_2}^2 .$$

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

$$(3.130) \quad \frac{T-K}{K_2 - n} \phi_{2SLS} \geq \left. \begin{array}{l} \\ \\ \\ \end{array} \right\} F_{K_2-n, T-K}(\alpha)$$

$$(3.131) \quad \frac{T-K}{K_2 - n} \phi_{LIML} \geq \left. \begin{array}{l} \\ \\ \\ \end{array} \right\}$$

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 (3.130). 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 (3.130)-(3.131) 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 (3.126) 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 (1968)) 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, viz $\phi_{LIML} \leq \phi_{2SLS}$, that arises from the construction of the LIML estimator (see (3.83) and (3.127)).

Some of the results in the preceding paragraph 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 noncentrality parameter matrix which we will denote by \bar{M} , as in paragraph 3.5 above. In our notation

$$(3.132) \quad \bar{M} = \frac{1}{2} \Gamma^{-1} E(X') (P_2 - P_{Z_1}) E(X) .$$

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

$$(3.133) \quad \bar{M} = T \begin{bmatrix} \beta' \\ \hline I_n \end{bmatrix} \Pi'_{22} \Pi_{22} [\beta : I_n] .$$

Thus, under the null \bar{M} has one zero latent root and generally n non-zero roots. When the null is false, the simpler form of \bar{M} in (3.133) no longer holds and (3.132) normally has rank $n+1$ rather than n . Thus, the true power functions of tests such as (3.128), (3.129) or (3.131) 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 \bar{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 earlier, Rhodes calculated the non-zero latent roots of the relevant noncentrality matrix for Basmann's experiment and found the roots to be large, explaining in part why (3.131) proved to be quite accurate in those experiments.¹⁹

Since the exact distribution of ϕ_{LIML} is not amenable to computation, some steps have been taken to provide improvements on the critical regions (3.128), (3.129) and (3.131). 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 \bar{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 earlier paragraphs 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 paragraph 2.3 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.

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 (1976, 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 estimators²⁰ 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.²¹ 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 as shown in paragraph 2.4. Such approximating moments, or pseudo-moments (in the terminology of paragraph 2.4 where this 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 paragraph 3.4 is possible. The principles may be illustrated by considering FIML applied to a two equation system of the form (3.1) with

$$(3.134) \quad B = \begin{bmatrix} 1 & b_{21} \\ b_{12} & 1 \end{bmatrix} \quad \text{alternatively normalized as} \quad B = \begin{bmatrix} \cos \theta & \cos \psi \\ \sin \theta & \sin \psi \end{bmatrix}$$

and overidentifying restrictions imposed on each column of the exogenous variable coefficient matrix C . We may consider the null hypothesis in which $C = 0$ and hence $\Pi = 0$ in the reduced form (3.2). In this case, the sample data and, hence, the unrestricted reduced form coefficient and error covariance matrix estimators (Π^*, Ω^*) carry no information about (b_{12}, b_{21}) or (θ, ψ) in (3.134). As with the LIML estimator, the p.d.f. of (Π^*, Ω^*) (or any other relevant sample moment matrices on which FIML depends) now splits into the p.d.f. of $(\theta_{\text{FIML}}, \psi_{\text{FIML}})$ which is uniform over the relevant domain and the density of appropriate auxiliary variates. The density of the unrestricted coefficients (b_{12}, b_{21}) in the conventional normalization now becomes Cauchy under the transformations $b_{12} = \tan \theta$ and $b_{21} = \tan \psi$ (compare (3.94)-(3.95) given earlier). This result confirms that the FIML estimator of the structural coefficients (b_{12}, b_{21}) has no finite integral moments. Sargan (1970) originally established this result by using the fact that the FIML estimator is independent of the normalization of the structural equation. In the context of (3.134) 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 will normally imply that the distribution of such an estimator has no integral moments. In the present case we note that the distributions of $b_{12}^{\text{FIML}} = \tan \theta$ and $b_{11}^{\text{FIML}} = \cot \theta$ are both Cauchy. (In fact, if ξ is Cauchy then $1/\xi$ is Cauchy also

as we may simply verify).

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, $\hat{\Sigma}_{2SLS}$, arising in the conventional 3SLS formula. Sargan, therefore, considers two cases. The first case treats $\hat{\Sigma}_{2SLS}$ as nonrandom or, more generally, allows $\hat{\Sigma}_{2SLS}$ 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 $\hat{\Sigma}_{2SLS}$ 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 $\hat{\Sigma}_{2SLS}$ nonrandom also applies in this second case where $\hat{\Sigma}_{2SLS}$ is the conventional error covariance matrix estimator.

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.²² 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 here.

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), viz. $\Pi = -CB^{-1}$. 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, viz.

$\int \hat{\Pi} \text{pdf}(\hat{B}, \hat{C}) d\hat{B} d\hat{C}$, 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):

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 derivatives at p_0 , denoted by 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.

PROOF

We consider the k^{th} absolute moment

$$E(|\hat{\beta}|^k) = \int |\psi(p)|^k |\phi(p)|^{-k} \text{pdf}(p) dp .$$

We select a neighborhood N_ϵ of p_0 for which

$$\inf_{p \in N_\epsilon} \{\text{pdf}(p) |\psi(p)|^k\} = \delta > 0$$

and for which the transformation $p \rightarrow (\phi, q)$ is 1:1, where $\phi = \phi(p)$ and q is a vector of $n-1$ components of p selected so that the jacobian of the transformation from p to (ϕ, q) viz.

$$(\partial\phi(p)/\partial p_j)^{-1} = g(\phi, q), \text{ for some } j$$

is non-zero throughout N_ϵ and, consequently, throughout the image of this neighborhood, M_ϵ , in (ϕ, q) space. We deduce that

$$(3.135) \quad \begin{aligned} E(|\hat{\beta}|^k) &\geq \int_{N_\epsilon} |\psi(p)|^k |\phi(p)|^{-k} \text{pdf}(p) dp \\ &\geq \delta \int_{M_\epsilon} |\phi|^{-k} |g(\phi, q)| d\phi dq . \end{aligned}$$

Since M_ϵ contains $\phi(p_0) = 0$ and $g(\phi, q)$ is continuous and non-zero in M_ϵ , the integral (3.135) diverges for all values of k in the interval $1 \leq k < \infty$ (see, for example, Widder (1961, p. 338)). It follows that $\hat{\beta}$ has no finite integral moments. \square

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 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. In this case we have

$$(3.136) \quad B = \begin{bmatrix} 1 & b_{21} \\ b_{12} & 1 \end{bmatrix} \quad C = \begin{bmatrix} c_{11} & 0 \\ c_{12} & c_{22} \\ 0 & c_{23} \end{bmatrix} \begin{matrix} K_1 \\ K_2 \\ K_3 \end{matrix}$$

and the reduced form coefficients for y_1 are given by

$$(3.137) \quad \begin{bmatrix} \pi_{11} \\ \pi_{12} \\ \pi_{13} \end{bmatrix} = -(1 - b_{12}b_{21})^{-1} \begin{bmatrix} c_{11} \\ c_{12} - b_{12}c_{22} \\ -b_{12}c_{23} \end{bmatrix} .$$

Assuming that standardizing transformations have been carried out and that structural estimation is by 2SLS we deduce from (3.65)-(3.66) that

$$\hat{b}_{12} = (y_2' Z_3 Z_3' y_2)^{-1} (y_2' Z_3 Z_3' y_1)$$

$$\hat{b}_{21} = (y_1' Z_1 Z_1' y_1)^{-1} (y_1' Z_1 Z_1' y_2)$$

$$\hat{c}_{11} = T^{-1} Z_1' y_1 - T^{-1} Z_1' y_2 \hat{b}_{12}$$

and similar formulae hold for \hat{c}_{12} , \hat{c}_{22} and \hat{c}_{23} . The reduced form coefficient vector π_{11} is now estimated by $\hat{\pi}_{11} = \hat{c}_{11} / (1 - \hat{b}_{12}\hat{b}_{21})$. The joint density pdf($\hat{b}_{12}, \hat{b}_{21}, \hat{c}_{11}$) is continuous and nonzero over \mathbb{R}^3 when the equation errors are normally distributed. Moreover, there exists a vector $(\hat{b}_{12}, \hat{b}_{21}, \hat{c}_{11})$ with $\hat{c}_{11} \neq 0$ for which $1 - \hat{b}_{12}\hat{b}_{21} = 0$. We may, for example, take $\hat{b}_{12} = \hat{b}_{21} = 1$ and then select any $\hat{c}_{11} \neq 0$ from its conditional normal distribution. In Theorem 3.9.1 $p = (\hat{b}_{12}, \hat{b}_{21}, \hat{c}_{11})$, $\phi(p) = 1 - \hat{b}_{12}\hat{b}_{21}$, $\psi(p) = \hat{c}_{11}$ and we deduce directly that $\hat{\pi}_{11}$ has no finite integral moments from the conclusion of the theorem.

Sargan (1976b) extends this framework of 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 $\det \hat{B} = \text{const}$. This can be most easily seen by considering in place of (3.136) and (3.137) the structural coefficient and reduced form matrices:

$$B = \begin{bmatrix} 1 & b_{21} \\ 0 & 1 \end{bmatrix}, \quad \Pi = \begin{bmatrix} \pi_{11} & \pi_{21} \\ \pi_{12} & \pi_{22} \\ \pi_{13} & \pi_{23} \end{bmatrix} = - \begin{bmatrix} c_{11} & -b_{21}c_{11} \\ c_{12} & c_{22} - b_{21}c_{12} \\ 0 & c_{23} \end{bmatrix}.$$

In this case reduced form coefficients in the second equation will possess moments to an order determined by the degree of overidentification of the second structural equation when the latter is estimated by 2SLS or 3SLS.

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

$$(3.138) \quad \ln[\det\{T^{-1}(Y - Z\Pi(\theta))'(Y - Z\Pi(\theta))\}].$$

From the criterion function (3.138) 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 (3.138). 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 (1976) 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).²³ 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:

$$(3.139) \quad \pi^* = \lambda \Pi_{\text{3SLS}} + (1-\lambda) \Pi_{\text{OLS}}$$

where

$$(3.140) \quad \lambda = \begin{cases} 1 & \text{if } \phi \leq C_\alpha \\ \left[\frac{C_\alpha}{\phi} \right]^{1/2} \quad \text{or} \quad \left[\frac{C_\alpha}{\phi} \right] & \text{if } \phi > C_\alpha \end{cases}$$

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)):

$$(3.141) \quad \phi = \text{tr}\{W^{-1}(\Pi_{\text{OLS}} - \Pi_{\text{3SLS}})'Z'Z(\Pi_{\text{OLS}} - \Pi_{\text{3SLS}})\}$$

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 (3.140) is the (asymptotic)

critical value of the test corresponding to a chosen significance level α . Note that the choice $\lambda = (C_\alpha/\phi)^{1/2}$ when $\phi > C_\alpha$ in (3.140) ensures that the test statistic, ϕ^* based on the new estimator Π^* satisfies

$$(3.142) \quad \phi^* = \text{tr}\{W^{-1}(\Pi_{OLS} - \Pi^*)'Z'Z(\Pi_{OLS} - \Pi^*)\} = \lambda^2\phi \leq C_\alpha.$$

The combined estimator Π^* can be additionally motivated by appealing to the principles of Stein-James estimation,²⁴ mixed regression²⁵ and minimum expected loss²⁶ (MELO) methodologies, all of which lead to estimators which can be expressed as combinations of restricted and unrestricted estimators. To highlight the similarity of (3.139) with Stein-like procedures (and in particular the positive rule estimator proposed by Sclove (1968, 1971)) we may write Π^* as

$$(3.143) \quad \Pi^* = \Pi_{3SLS} + I_{(C_\alpha, \infty)} \left(1 - \frac{C_\alpha}{\phi}\right) (\Pi_{OLS} - \Pi_{3SLS})$$

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).²⁷ 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 paragraph) 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 of the critical level C_α . The statistic ϕ in (3.141) has a limiting χ_N^2 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 (3.141) 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 paragraph 2.3 and this is an area of extensive current research in mathematical statistics (see the discussion and references given in paragraph 2.3 above). 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

$$(3.144) \quad \begin{bmatrix} \pi_{11} \\ \pi_{21} \end{bmatrix} = \begin{bmatrix} \pi_{12} \\ \pi_{22} \end{bmatrix} \beta + \begin{bmatrix} \gamma \\ 0 \end{bmatrix} = \left[\begin{array}{c|c} \pi_{12} & I \\ \hline \pi_{22} & 0 \end{array} \right] \begin{bmatrix} \beta \\ \gamma \end{bmatrix}$$

(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 side of (3.144)) is then

$$(3.145) \quad \begin{bmatrix} \hat{\pi}_{11} \\ \hat{\pi}_{21} \end{bmatrix} = \left[\begin{array}{c|c} (Z'Z)^{-1}Z'Y & \\ \hline & 0 \end{array} \right] \begin{bmatrix} \beta_{2SLS} \\ \gamma_{2SLS} \end{bmatrix}$$

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} , γ_{2SLS} are the usual structural 2SLS estimators. The small sample properties of the estimator (3.145) and associated forecasts have been studied by Knight (1977), Nagar and Sahay (1978) 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 (3.145) 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 (3.145) 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 paragraph 3.3. Their conclusion accords with the intuitive ideas discussed in the paragraph following (3.75) 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 (3.75)).

3.10. Improved Estimation of Structural Coefficients

The idea of developing alternative estimators which may improve on certain aspects of the small sample 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.²⁸ 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 appeared in Nagar (1959) who provided (pseudo-) moment expansions (see paragraph 2.4 above) 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 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 macroeconomic 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)²⁹ 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 paragraph 2.4 and footnote 6). 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 error³⁰ of the 2SLS-LIML combination and demonstrated its superiority over LIML according to this criterion.

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 paragraph 3.2 above, is as follows:

$$(3.146) \quad \beta_{21} = [A_{22}(I) - \ell\{A_{22}(I) - A_{22}(P_Z)\}]^{-1} [a_{21}(I) - \ell\{a_{21}(I) - a_{21}(P_Z)\}]$$

where

$$(3.147) \quad \ell = a\lambda + b, \quad a, b = \text{const's.}$$

and λ is the smallest latent root of the equation

$$(3.148) \quad \det[A(I) - \lambda\{A(I) - A(P_Z)\}] = 0$$

as in (3.22) above. The estimator β_M in (3.146) specializes: to LIML for $a = 1$, $b = 0$; to the fixed k class for $a = 0$, $b = \text{const.}$; 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 (3.147), β_M is a direct modification of β_{LIML} (compare (3.23) above). Fuller shows that when estimators of the class (3.146) 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 (3.146) 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 (3.146) 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 earlier (in paragraph 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 paragraphs 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, $0 \leq k \leq 1$*)

Integral moments of the estimator $\beta_{(k)}$ given by (3.19) for non-stochastic k in the structural equation (3.3) exist up to order M where

$$M = \begin{cases} T - K_1 - n & \text{for } 0 \leq k < 1 \\ K_2 - n & \text{for } k = 1 . \end{cases}$$

□

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 $0 \leq k \leq 1$ and developed asymptotic expansions for them in terms of the reciprocal of the concentration parameter, viz., $1/\mu^2$. Similar formulae were derived by Takeuchi (1970) for OLS, 2SLS and IV estimators in the two endogenous variable case.³¹ 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) 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 paragraph 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 we analyzed in paragraphs 3.5-3.6 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.³²

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 paragraph 3.3 above and the results for the coefficients of the included endogenous variables. Thus, in the case of the IV estimator we have from (3.66):

$$(3.149) \quad Y_{IV} = T^{-1}Z_1'Y_1 - T^{-1}Z_1'Y_2\beta_{IV}$$

(assuming standardizing transformations are performed). Then $Z_1'X = Z_1'[y_1 : Y_2]$ is statistically independent of β_{IV} (under error normality of course) and moments of Y_{IV} in (3.149) are defined to the same order as those of Y_{IV} (see paragraph 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 paragraphs 3.5-3.6. 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:

$$(3.150) \quad y_1 = \gamma_2 \beta + Z_1 \gamma_1 + Z_4 \gamma_4 + u .$$

We write the reduced form as

$$(3.151) \quad [y_1 : y_2] = X = [Z_1 : Z_4 : Z_5] \begin{bmatrix} \pi_{11} & \pi_{12} \\ \pi_{41} & \pi_{42} \\ \pi_{51} & \pi_{52} \end{bmatrix} + [v_1 : v_2]$$

$$= [Z_1 : Z_4 : Z_5] \begin{bmatrix} \pi_1 \\ \dots \\ \pi_4 \\ \dots \\ \pi_5 \end{bmatrix} + v$$

where the coefficients satisfy

$$(3.152) \quad \pi_{11} - \pi_{12}\beta = \gamma_1, \quad \pi_{41} - \pi_{42}\beta = \gamma_4, \quad \pi_{51} - \pi_{52}\beta = 0.$$

We define $Z_2 = [Z_4 : Z_5]$ and then

$$(3.153) \quad M = E[T^{-1/2}X'Z_2] = T^{-1/2}\Pi'Z_2 = T^{1/2}[\Pi'_4 : \Pi'_5].$$

The noncentrality parameter matrix is

$$(3.154) \quad MM' = T[\Pi'_4\Pi_4 + \Pi'_5\Pi_5] = TD, \quad \text{say.}$$

We may note that this reduces to

$$(3.155) \quad \begin{bmatrix} \beta' \\ I \end{bmatrix} (\Pi'_{42}\Pi_{42} + \Pi'_{52}\Pi_{52})(\beta : I) = \begin{bmatrix} \beta' \\ I \end{bmatrix} \Pi'_{22}\Pi_{22}(\beta : I)$$

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

As in paragraph 3.5, OLS and 2SLS estimators of β in (3.150) depend on the matrix

$$(3.156) \quad A = T^{-1}X'FF'X$$

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 paragraph 3.5, as shown by Maasoumi and Phillips (1982b). Thus, the exact p.d.f. of A in (3.156) is given by (3.97) and the analysis leading to (3.98) gives us in the present case

$$(3.157) \quad \text{pdf}(w, r, B) = \frac{\text{etr}\left[-\frac{T}{2}D\right]}{2^{f(n+1)/2} \Gamma_{n+1}\left(\frac{f}{2}\right) [\det(I+rr')]^{(f+1)/2}} \\ \cdot {}_0F_1\left(\frac{f}{2}; \frac{T}{4}AD\right) \exp(-w/2) \text{etr}(-B/2) w^{(f-n-2)/2} (\det B)^{(f-n)/2}$$

where

$$A = \left[\begin{array}{c|c} w + r'(I+rr')^{-1/2} B(I+rr')^{-1/2} r & r'(I+rr')^{-1/2} B(I+rr')^{-1/2} \\ \hline (I+rr')^{-1/2} B(I+rr')^{-1/2} r & (I+rr')^{-1/2} B(I+rr')^{-1/2} \end{array} \right].$$

Analysis for the case $n = 1$ where the equation (3.150) contains only two endogenous variables is now straightforward. Since

$$(3.158) \quad {}_0F_1(a; S) = \sum_{k=0}^{\infty} \frac{(\det S)^k}{\left[a - \frac{1}{2}\right]_k (a)_{2k} k!} \sum_{\ell=0}^{\infty} \frac{\{\text{tr}(S)\}^{\ell}}{(a+2k)_{\ell} \ell!}$$

(Muirhead (1975)) when S is a 2×2 matrix, (3.157) has the following explicit series representation (where we set B equal to the scalar b):

$$(3.159) \quad \text{pdf}(w, r, b) = \frac{\text{etr}(-TD/2)}{2^f \pi^{1/2} \Gamma\left(\frac{f}{2}\right) \Gamma\left(\frac{f-1}{2}\right) (1+r^2)^{(f+1)/2}} \\ \cdot \sum_{k=0}^{\infty} \frac{\{\det(TD/4)\}^k}{\left[\frac{f-1}{2}\right]_k \left[\frac{f}{2}\right]_{2k} k!} \sum_{h=0}^{\infty} \sum_{j=0}^{\infty} \frac{(T/4)^{h+j} d_{11}^j (r^2 d_{11} + 2rd_{12} + d_{22})^h}{\left[\frac{f}{2} + 2k\right]_{h+j} h! j! (1+r^2)^{k+h}} \\ \cdot w^{(f-3)/2+k+j} b^{(f-1)/2+k+h} \exp(-w/2) \exp(-b/2).$$

Integrating over the regions $0 \leq w < \infty$ and $0 \leq b < \infty$ and using (3.104)

we obtain

$$(3.160) \quad \text{pdf}(r) = \frac{\text{etr}(-TD/2)}{B\left(\frac{1}{2}, \frac{f}{2}\right) (1+r^2)^{(f+1)/2}} \sum_{k=0}^{\infty} \frac{(\det(TD/2))^k \left(\frac{f+1}{2}\right)_k}{\left(\frac{f}{2}\right)_{2k} k!}$$

$$\cdot \sum_{h=0}^{\infty} \frac{\left(\frac{T}{2}\right)_h \left(\frac{f+1}{2} + k\right)_h (r^2 d_{11} + 2rd_{12} + d_{22})^h}{\left(\frac{f}{2} + 2k\right)_h h! (1+r^2)^{k+h}} {}_1F_1\left(\frac{f-1}{2} + k, \frac{f}{2} + h + 2k; \frac{Tdu}{2}\right)$$

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 (3.106) when the structural equation is correctly specified ($\gamma_4 = 0$) as can be shown by rearrangement of the series.

Formula (3.160) was derived by Rhodes and Westbrook (1981)³³ 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 nonstochastic in the interval $0 \leq k \leq 1$). 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 have 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).^{34, 35, 36, 37}

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 seem 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 wavy 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 approximants¹ 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 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 paragraph 3.4), (ii) knowledge of its moments where these exist, (iii) leading terms in its Edgeworth expansion (see paragraph 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 zeroes 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:

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

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 $\text{pdf}(r) > 0$ and $\text{pdf}(r) \rightarrow 0$ as $r \rightarrow \pm\infty$ (already explicit in $s(r)$); or hard quantitative information, for example of the type (i) that $\text{pdf}(r)$ has moments up to a certain order or (ii) that $\text{pdf}(r)$ takes an especially simple form in an important and relevant leading case or (iii) the $\text{pdf}(r)$ 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. $m = n$).² 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 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 $r \rightarrow \pm\infty$.

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

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

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 $\gamma' = \{a_0, \dots, a_n; b_1, \dots, b_n\}$ in (4.1), this means that there exists a vector γ^* and a corresponding rational fraction $R'_{nn}(r)$ for which

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

given some continuous density $\text{pdf}(r)$; and, moreover, the rational fraction $R'_{nn}(r)$ with the property (4.3) is unique. As $n \rightarrow \infty$ $R'_{nn}(r)$ converges uniformly to $\text{pdf}(r)$. Hence, 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 paragraph 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.

4.3. An Example of Multiple Point Padé Approximation

We take as an example the 2SLS estimator, β_{2SLS} , of the scalar coefficient β in the structural equation (3.3) when $n = 1$ (the two endogenous variable case). The exact density of β_{2SLS} is given by (3.106) and we seek a simple rational function of the form (4.1) that can serve as an adequate approximation to it. In constructing such an approximant we note the following useful information about the distribution:

(i) Integral moments exist up to order $K_2 - 1$ (see paragraph 3.5 and also Theorem 3.101);

(ii) The leading term in the exact density (3.106) takes the form

$$(4.4) \quad s(r) = \left[B \left(\frac{1}{2}, \frac{K_2}{2} \right) \right]^{-1} (1+r^2)^{-(K_2+1)/2}$$

as can readily be deduced from a leading case analysis (see paragraph 3.4 and equation (3.82));

(iii) Writing

$$(4.5) \quad \text{pdf}(r) = s(r)\text{pdf}_s(r)$$

where $s(r)$ is defined by the leading case (4.4) we find from (3.106) the following Taylor expansions about the origin ($r = 0$) and infinity ($r = \pm\infty$)

$$(4.6) \quad \text{pdf}_s(r) \sim \beta_0 + \beta_1 r + \beta_2 r^2 + \beta_3 r^3 + \dots, \text{ as } r \rightarrow 0$$

$$(4.7) \quad \text{pdf}_s(r) \sim \alpha_0 + \alpha_1 r^{-1} + \alpha_2 r^{-2} + \alpha_3 r^{-3} + \dots, \text{ as } |r| \rightarrow \infty$$

in which the coefficients β_i and α_j are constants that depend on the parameters of the density (3.106).

By utilizing the leading case (4.4) we may construct a rational approximant of the family (4.1)

$$(4.8) \quad R_{nn}(r) = s(r) \frac{a_0 + a_1 r + \dots + a_n r^n}{1 + b_1 r + \dots + b_n r^n} = s(r) [n/n](r), \quad \text{say}$$

where numerator and denominator polynomials are of equal (even) degree to ensure asymptotic behavior in the approximant (4.8) compatible with that implied by the information in (i) and (ii) above. The coefficients in the (normalized) rational fraction $[n/n](r)$ of (4.8) can be chosen so that this function has the same local behavior at the origin and at infinity as (4.6) and (4.7). Explicitly these coefficients are determined by the equations

$$(4.9) \quad \text{pdf}_s(r) Q_n(r) - P_n(r) = O(r^{n+1}), \quad r \rightarrow 0$$

$$(4.10) \quad (\text{pdf}_s(r) Q_n(r) - P_n(r)) r^{-n} = O(r^{-n}), \quad r \rightarrow \pm\infty$$

where the order of equivalence in (4.9) and (4.10) is selected to provide sufficient information to uniquely determine the coefficients of $[n/n](r)$. Equations (4.9)-(4.10) lead to the system of $2n+1$ simultaneous equations

$$\begin{array}{c}
 \xrightarrow{n+1} \quad \quad \quad \xrightarrow{n} \\
 \begin{array}{c}
 \uparrow \\
 n+1 \\
 \downarrow
 \end{array}
 \begin{array}{|c|}
 \hline
 \begin{array}{cccc}
 1 & 0 & 0 & \dots & 0 \\
 0 & 1 & 0 & \dots & 0 \\
 \dots & \dots & \dots & \dots & \dots \\
 0 & 0 & 0 & \dots & 1
 \end{array} \\
 \hline
 \begin{array}{cccc}
 0 & 1 & 0 & \dots & 0 \\
 0 & 0 & 1 & \dots & 0 \\
 \dots & \dots & \dots & \dots & \dots \\
 0 & 0 & 0 & \dots & 1
 \end{array} \\
 \hline
 \end{array}
 \begin{array}{|c|}
 \hline
 \begin{array}{ccc}
 0 & 0 & 0 \\
 -\beta_0 & 0 & 0 \\
 -\beta_1 & -\beta_0 & 0 \\
 \dots & \dots & \dots \\
 -\beta_{n-1} & -\beta_{n-2} & \dots & -\beta_0
 \end{array} \\
 \hline
 \begin{array}{ccc}
 -\alpha_0 & -\alpha_1 & \dots & -\alpha_{n-1} \\
 0 & -\alpha_0 & \dots & -\alpha_{n-1} \\
 \dots & \dots & \dots & \dots \\
 0 & 0 & \dots & -\alpha_0
 \end{array} \\
 \hline
 \end{array}
 \begin{array}{|c|}
 \hline
 \begin{array}{c}
 a_0 \\
 a_1 \\
 a_2 \\
 \vdots \\
 a_n
 \end{array} \\
 \hline
 \begin{array}{c}
 b_1 \\
 b_2 \\
 \vdots \\
 b_n
 \end{array} \\
 \hline
 \end{array}
 =
 \begin{array}{|c|}
 \hline
 \begin{array}{c}
 \beta_0 \\
 \beta_1 \\
 \vdots \\
 \beta_n
 \end{array} \\
 \hline
 \begin{array}{c}
 0 \\
 \vdots \\
 0
 \end{array} \\
 \hline
 \end{array}
 \end{array}
 \quad (4.11)$$

which can be solved for the coefficients. The equations (4.11) are the multiple point Padé equations that relate the rational function $[n/n](r)$ to the Taylor expansions (4.6) and (4.7) at the origin and infinity. They can be generalized to allow for an arbitrary number of points of local expansion.³ The choice of $n+1$ equations to refer to the origin and n equations to refer to infinity in (4.11) is also somewhat arbitrary. Present experience tends to suggest that the first one or two coefficients in the tail expansion (4.7) are the most useful in density function approximation. After that, information about the shape of the main body of the distribution appears to be more valuable in determining the quality of the Padé approximant.

Once (4.11) or its generalization to a greater number of local expansions has been solved, the coefficients in $[n/n](r)$ are determined and the resulting approximant $R_{nn}(r)$ in (4.8) may be regarded as a crude Padé approximant to the density pdf(r). In some cases this may turn out to be a satisfactory approximant. In others, it will need to be modified to produce a good approximation to the exact density over a wide interval. This is because there is nothing in the crude Padé procedure

which prevents the occurrence of zeroes and poles in the rational fraction $[n/n](r)$. The poles in particular will need to be eliminated if the approximation is to perform adequately unless they occur in remote and (essentially) irrelevant regions of the distribution.

A procedure for modifying Pade approximants so that they are well behaved over the whole real axis is suggested in Phillips (1982c). The idea arises from the observation that (real) zeroes and poles of $[n/n](r)$ occur in pairs since n is an even integer. We may therefore replace the real roots of the associated quadratics by complex conjugate pairs selected to preserve as far as possible the known behavior of the function at the points of local expansion. These principles may be illustrated by the case of an approximant derived from (4.11) with $n = 4$. Suppose we have

$$(4.12) \quad [4/4](r) = \frac{\sum_{i=0}^4 a_i r^i}{\sum_{i=0}^4 b_i r^i} = \frac{a_4 (r - \gamma_1)(r - \gamma_2)(r - \gamma)(r - \bar{\gamma})}{b_4 (r - \delta_1)(r - \delta_2)(r - \delta)(r - \bar{\delta})}$$

where (γ_1, γ_2) and (δ_1, δ_2) are the real zeroes of the numerator and denominator polynomials. We rewrite (4.12) as

$$(4.13) \quad [4/4](r) = \frac{a_4 \{r^2 - (\gamma_1 + \gamma_2)r + \gamma_1 \gamma_2\} (x - \gamma)(x - \bar{\gamma})}{b_4 \{r^2 - (\delta_1 + \delta_2)r + \delta_1 \delta_2\} (x - \delta)(x - \bar{\delta})}$$

and modify the coefficients of the quadratics in braces in (4.13) so that $[4/4](r)$ is a bounded positive function on the real axis with the same behavior (but to a lower order) as (4.12) in the locality of $r = 0$ and $r^{-1} = 0$. To achieve this modification we define the family of related rational functions

$$(4.14) \quad [4/4](r; \theta) = \frac{a_4 \{cr^2 + dr + e\}(r-\gamma)(r-\bar{\gamma})}{b_4 \{fr^2 + gr + h\}(r-\delta)(r-\bar{\delta})}$$

where $\theta' = (c, d, e, f, g, h)$ is a vector of supplementary parameters. To ensure first order equivalent local behavior in (4.14) and (4.12) we must restrict our choice of θ so that $c/f = 1$ and $e/h = \gamma_1 \gamma_2 / \delta_1 \delta_2$. These conditions will ensure that $[4/4](r; \theta) / [4/4](r) \rightarrow 1$ as $x \rightarrow 0$ or $\pm\infty$. We must also confine our choice of θ so that the desired global behavior $0 < [4/4](r; \theta) < \infty$ is achieved. This restricts us to the set of θ for which the discriminants $d^2 - 4ce$ and $g^2 - 4fh$ are negative. Further degrees of freedom in the choice of θ may be used up by raising the order of local contact between $[4/4](r; \theta)$ and $[4/4](r)$ at the origin and infinity.

The procedure outlined above was successfully applied in the context of approximating the exact density (3.106) of the 2SLS estimator by the rational family $R_{nn}(r)$ in (4.8) with n set equal to 4. 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 (see paragraphs 2.3, 2.5) produced very poor results. In particular, the modified two-point Padé approximant (as above) 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.

4.4 Curve Fitting or Constructive Functional Approximation?

The above discussion and reported application present a favourable 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 $\{a_0, \dots, a_m, b_1, \dots, b_n\}$. Thus, Edgeworth expansions and saddlepoint approximations (see paragraphs 2.3, 2.5) are just sub-cases 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 $|r| \rightarrow \infty$. Additional information about the distribution can then be incorporated in the rational coefficients and in adjustments to $s(r)$ that ensure the same tail behaviour 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 above example (4.4) demonstrates. Moreover, experimental data about the distribution can be utilised in the choice of the rational coefficients by least squares or generalised least squares fitting to the empirical distribution in place of (or in addition to) exact solutions such as (4.11). 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 unmistakably 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 paragraphs 3.5 - 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 marshal 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 phenomena 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.¹

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 paragraphs 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.

FOOTNOTES

SECTION 1

1. The nature of local alternative hypotheses is discussed by Engle (1982) in this Handbook.
2. See, for example, Fisher (1921, 1922, 1924, 1928, 1935) and the treatment of exact sampling distributions by Cramér (1946).
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.

SECTION 2

1. See, for example, Appendix B of Sargan (1976a) and Phillips (1980a). These issues will be taken up further in paragraph 3.5.
2. See the discussion and references in Phillips (1978).
3. Sargan (1975), Phillips (1977b) and Sargan and Satchell (1981) provide rigorous justifications of the expansions that follow and a full set of conditions under which their validity is established.
4. The second equality of (2.16) is obtained by expanding the exponential factor $\exp[i\epsilon_{rs} \bar{m}_r \bar{m}_s / 2\sqrt{T} + \dots]$.
5. The same technique as that in footnote 4 is applied again.
6. In the argument that ensues we develop 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 $\sigma \rightarrow 0$. Anderson (1977) has recently discussed the relationship between these alternative parameter sequences in the context of the SEM.

7. Conditions under which (2.30) is valid (in the sense that the order of magnitude of the error on the moment approximation $E|\theta_{Tk(m)}|^j$ is accurately stated) are investigated and reported in Sargan (1974) and Sargan (1976a, Appendix A). The article by Srinivasan (1970) also addresses the issue of the validity of the Nagar moment approximations.
8. See, for example, Phillips (1978), Holly and Phillips (1979), Daniels (1980), Durbin (1980a, 1980b) and Barndorff-Nielson and Cox (1979).

SECTION 3

1. 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 SEM's) 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.

2. See also the useful discussion and graphical plots in Anderson (1982).
3. This density is given, for example, in Mariano and McDonald (1979).
4. This parameter is so called because as $\mu^2 \rightarrow \infty$ the commonly used single equation estimators all tend in probability to the true parameter. Thus, the distributions of these estimators all "concentrate" as $\mu^2 \rightarrow \infty$, even if the sample size T remains fixed. See Basmann (1963) and Mariano (1975) for further discussion of this point.
5. Similar situations arise with certain posterior p.d.f.'s in Bayesian analysis. Zellner (1978) provides an interesting analysis of such a case involving the posterior p.d.f. of the reciprocal of a mean with a diffuse prior.
6. See Basmann (1963) and Kabe (1963, 1964).
7. Some other specialized SEM models in which the distributions of commonly used estimators depend only on central Wishart matrices are discussed by Wegge (1971).
8. An example of this type of analysis for structural variance estimators is given in paragraph 3.7.
9. 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

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

when $\Pi_{22} = 0$. The first of these equations corresponds to (3.79) in the text when $\beta = 0$.

10. An alternative approach to the extraction of the exact density of β_{IV} from (3.98) is given in Appendix B of Phillips (1980a) 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 and b) and has recently been extended by Chikuse (1981) to matrix multinomial expansions of zonal polynomials of the sum of several matrices.
11. 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 (1980) and Chikuse (1981).
12. See Mariano and McDonald (1979) for a small correction.
13. The numbers in these tables have been selected from the extensive tabulations in Anderson and Sawa (1978, 1979) which are recommended to the reader for careful study. My thanks go to Professors Anderson and Sawa for their permission to quote from their tables.
14. We note that since β_{LIMLK} depends only on the noncentral 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} .
15. 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 (3.109)

$$(3.109)^{15} Q(\beta) = \beta'_{\Delta} X' (P_H - P_{Z_1}) X \beta_{\Delta} = \beta'_{\Delta} A (P_H) \beta_{\Delta} .$$

16. Direct integration of this expression for the density yields

$$\int_0^{\infty} \text{pdf}(\sigma_{IV}) d\sigma_{IV} = \frac{B\left(\frac{1}{2}, \frac{T-K_1+K_3}{2}\right)}{B\left(\frac{1}{2}, \frac{K_3}{2}\right)} \sum_{j=0}^{\infty} \frac{\left(\frac{1}{2}\right)_j \left(\frac{T-K_1}{2}\right)_j}{\left(\frac{T-K_1+K_3+1}{2}\right)_j j!}$$

$$= \frac{B\left(\frac{1}{2}, \frac{T-K_1+K_3}{2}\right)}{B\left(\frac{1}{2}, \frac{K_3}{2}\right)} {}_2F_1\left(\frac{1}{2}, \frac{T-K_1}{2}; \frac{T-K_1+K_3+1}{2}; 1\right)$$

which is seen to be unity upon application of the result (Erdélyi (1953), p. 104):

$${}_2F_1(a, b; c; 1) = \frac{\Gamma(c)\Gamma(c-a-b)}{\Gamma(c-a)\Gamma(c-b)}$$

for $c \neq 0, -1, -2, \dots$ and $\text{Re}(c) > \text{Re}(a+b)$.

17. See, in particular, formula (25) on page 228 of Basmann (1974). This formula seems to involve a printing error in that $\Gamma\left(\frac{n+v+m}{2}\right)$ should be replaced by $\Gamma\left(\frac{n+v+m+1}{2}\right)$ for correct normalization, in the notation of that article.

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

19. Strictly speaking Basmann (1960) reports in detail these results for (3.130) i.e. the 2SLS statistic. However, another article detailing the results for (3.131) is referenced and the extensive recent experiments by Rhodes and Westbrook (1982) suggest the same conclusion, at least for low degrees of overidentification. In fact, Rhodes and Westbrook find that the LIML statistic is always more closely approximated

- by the $F_{K_2-n, T-K}$ distribution than is the 2SLS statistic in their experimental outcomes.
20. See Anderson and Sawa (1978), Holly and Phillips (1979), 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).
 21. Recent work in the same framework has been published by Maekawa (1980) for t ratio type test statistics.
 22. 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.
 23. 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).
 24. See, for example, James and Stein (1961), Zellner and Vandaele (1975) and Judge and Bock (1982) in this Handbook.
 25. See Goldberger (1973).
 26. See Zellner (1978) and Zellner and Park (1979).
 27. The finite sample properties of Stein-like improved estimators in the context of the linear regression model have been studied by Ullah (1974, 1980).

28. 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).
29. See also Morimune and Kunitomo (1980).
30. That is, the mean squared error of the asymptotic expansion of the distribution up to a certain order.
31. 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.
32. The issues raised in this paragraph 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-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 to have been rarely discussed in the literature in this field.
33. 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 $|T_n/2|^k$.
34. 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.

35. 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).
36. 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.
37. 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.

SECTION 4

1. Padé 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).
2. This is discussed at some length in Phillips (1982c, 1982d).
3. The corresponding system of equations is given in Phillips (1982c).

SECTION 5

1. Some further reflections on the problems inherent in asymptotic theory are given in Phillips (1982b)

REFERENCES

- Akahira, M. and K. Takeuchi (1979), "The Concept of Asymptotic Efficiency and Higher Order Asymptotic Efficiency in Statistical Estimation Theory," Lecture Notes, Faculty of Economics, University of Tokyo.
- 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. (1958). An Introduction to Multivariate Statistical Analysis. New York: J. Wiley & Sons.
- Anderson, T. W. (1976), "Estimation of Linear Functional Relationships: Approximate Distribution and Connections with Simultaneous Equations in Econometrics," Journal of Royal Statistical Society, B 38, 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 (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 #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," The 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-Nielsen, O. and D. R. Cox (1979), "Edgeworth and Saddle-point Approximations with Statistical Applications," Journal of the Royal Statistical Society, Series B, 41, pp. 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. (1963b), "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 of 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," Chapter 4 in Frontiers of Quantitative Economics, Vol. 2 (M. D. Intriligator and D. A. Kendrick, eds.). North-Holland.
- Basmann, R. L. and R. H. Richardson (1969), "The Finite Sample Distribution of a Structural Variance Estimator," Research Papers in Theoretical and Applied Economics, 24, University of Kansas, mimeographed.
- Basmann, R. L. and D. H. Richardson (1973), "The Exact Finite Sample Distribution of a 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: Wiley).
- Bleistein, N. and R. A. Handelsman (1976). Asymptotic Expansions of Integrals, Hold Rinehart and Winston: New York.
- Byron, R. P. (1974), "Testing Structural Specification Using the Unrestricted Reduced Form," Econometrica 42, 869-883.
- Chambers, J. M. (1967), "On Methods of Asymptotic Approximation for Multivariate Distributions," Biometrika 54, 367-383.
- Chikuse, Y. (1981), "Invariant Polynomials with Matrix Arguments and Their Applications." R. P. Gupta (ed.) Multivariate Statistical Analysis, North-Holland.
- 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 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.

- Davis, A. W. (1980b), "Invariant Polynomials with Two Matrix Arguments Extending the Zonal Polynomials: Applications to Multivariate Distribution Theory," Annals of Institute of Statistical Mathematics, forthcoming.
- De Bruijn, N. G. (1958), Asymptotic Methods in Analysis, Amsterdam: North-Holland.
- Dhrymes, P. J. (1969), "Alternative Asymptotic Tests of Significance and Related Aspects of 2SLS and 3 SLS Estimated Parameters," Review of Economic Studies, 36, 213-226.
- Dhrymes, P. J. (1973), "Restricted and Unrestricted Reduced Forms: Asymptotic Distribution and Relative Efficiency," Econometrica, Vol. 41, 119-134.
- Durbin, J. (1980a), "Approximations for Densities of Sufficient Estimators," Biometrika, 67, 311-33.
- Durbin, J. (1980b), "The Approximate Distribution of Partial Serial Correlation Coefficients Calculated from Residuals from Regression on Fourier Series," Biometrika, 67, 335-49.
- 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.
- Engle, R. F. (1982), "Wald, Likelihood Ratio and Lagrange Multiplier Tests in Econometrics," M. D. Intriligator and Z. Griliches (Eds.) Handbook of Econometrics, Vol. 2, Amsterdam: North-Holland.

- Erdélyi, A. (1953), Higher Transcendental Functions, Vol. 1 (New York: McGraw-Hill).
- Feller, W. (1970), An Introduction to Probability Theory and Its Applications Vol. II (Wiley: New York).
- Fieiller, 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, 329-32.
- Fisher, R. A. (1925), Statistical Methods for Research Workers, Edinburgh: Oliver and Boyd.
- Fisher, R. A. (1928), "On a Distribution Yielding the Error Functions of Several Well Known Statistics," Proceedings of the International Congress of Mathematics, Toronto, 805-13.

- Fisher, R. A. (1928), "The General Sampling Distribution of the Multiple Correlation Coefficient," Proceedings of the Royal Statistical Society, 121, 654-73.
- Fisher, R. A. (1935), "The Mathematical Distributions Used in the Common Tests of Significance," Econometrica 3, 353-65.
- Fukikoshi, Y. (1973), "Asymptotic Formulas for the Distribution of Three Statistics for Multivariate Linear Hypotheses," Annals of the Institute of Statistical Mathematics, 25, 423-437.
- Fujikoshi, Y. (1977), "Asymptotic Expansions for the Distribution of Some Multivariate Tests," in P. R. Krishnaiah (Ed.), Multivariate Analysis - IV, North-Holland.
- Fuller, W. A. (1977), "Some Properties of a Modification of the Limited Information Estimator," Econometrica, 45, 939-953.
- Geary, R. C. (1944), "Extension of a Theorem by Harald Cramér on the Frequency Distributions of a Quotient of Two Variables," Journal of the Royal Statistical Society, 17, 56-67.
- Gil-Pelaez, J. (1951), "Note on the Inversion Theorem," Biometrika, 38, 481-482.
- Goldberger, A. (1973), "Efficient Estimation in Overidentified Models: An Interpretive Analysis," A. S. Goldberger and O. D. Duncan (Eds.) Structural Equation Models in the Social Sciences, Seminar Press.
- Gurland, J. (1948), "Inversion Formulae for the Distribution of Ratios," Annals of Mathematical Statistics 19, 228-237.
- Haavelmo, T. (1947), "Methods of Measuring the Marginal Propensity to Consume," Journal of the American Statistical Association, 42, 105-122.

- 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.
- Hannan, E. J. (1970), Multiple Time Series (New York: Wiley).
- 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. (1982), "Specification and Estimation of Simultaneous Equation Models," M. D. Intriligator and Z. Griliches (Eds.) Handbook of Econometrics, Vol. 2, Amsterdam: North-Holland.
- 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.
- Hendry, D. F. (1982), "Monte Carlo Experimentation in Econometrics," in M. D. Intriligator and Z. Griliches (Eds.) Handbook of Econometrics, Vol. 2, Amsterdam: North Holland.
- 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 Endogeneous Variable in a General Structural Equation," mimeographed, Monash University.
- 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 Hood, Wm. C. and T. C. Koopmans (Eds.) Studies in Econometric Method, Cowles Commission, New York: J. Wiley & Sons.
- Hsiao, C. (1981), "Identification," in M. D. Intriligator and Z. Griliches (Eds.) Handbook of Econometrics, Amsterdam: North Holland.
- Hurwicz, L. (1950), "Least Squares Bias in Time Series," in T. C. Koopmans (ed.), Statistical Inference in Dynamic Economic Models, New York: Wiley.
- Ibragimov, I. A. and V. V. Linnik (1971), Independent and Stationary Sequences of Random Variables (Groningen: Wolter-Noordhoff).
- Imhof, J. P. (1961), "Computing the Distribution of Quadratic Forms in Normal Variables," Biometrika 48, 419-26.
- Intriligator, M. D. and D. A. Kendrick (1974) Eds., Frontiers of Quantitative Economics, Vol. 11, North-Holland.
- 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, 475-50.
- James, A. T. (1968), "Calculation 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, 361-379, Berkeley, University of California Press.

- 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: Wiley.
- Judge, G. G., and M. E. Bock (1982), "Biased Estimation," in M. D. Intriligator and Z. Griliches (Eds.) Handbook of Econometrics, Vol. 2, Amsterdam: North-Holland.
- Kabe, D. G. (1964), "On the Exact Distributions of the GCL Estimators in a Leading Three-Equation Case," Journal of the American Statistical Association, Vol. 58, 535-537.
- 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, Vol. 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.

- 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," Mimeograph, School of Economics, The University of New South Wales.
- 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. (1965), Special Functions and Their Application (Englewood Cliffs, N.J.: Prentice Hall).
- Lebedev, N. N. (1972), Special Functions and Their Application, Prentice Hall.
- Maasoumi, E. (1977), "A Study of Improved Methods of Estimating Reduced Form Coefficients Based Upon 3SLS," unpublished Ph.D. Thesis, London School of Economics, 1977.
- Maasoumi, E. (1978), "A Modified Stein-Like Estimator for the Reduced Form Coefficients of Simultaneous Equations," Econometrica, 46, 695-703.
- Maasoumi, E. (1981), "Uncertain Structural Models and Generic Reduced Form Estimation," Iowa University (mimeographed).
- 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," Mimeographed, Yale University.

- 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.
- Magnus, J. R. and H. Neudecker (1979), "The Commutation Matrix: Some Properties and Applications," The Annals of Statistics, 7, 381-394.
- Magnus, J. R. and H. Neudecker (1980), "The Elimination Matrix: Some Theorems and Applications," SIAM Journal on Algebraic and Discrete Methods, 1, 422-4491
- Malinvaud, E. (1980), Statistical Methods of Econometrics, North-Holland.
- Mariano, R. S. (1973), "Approximations to the Distribution Functions of the Ordinary Least-Squares and Two-state 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. 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 American Statistical Association, 67, 159-163.

- Marsaglia, G. (1965), "Ratios of Normal Variables and Ratios of Sums of Uniform Variables," Journal of the American Statistical Association, 60, 193-204.
- 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, 1969.
- Mikhail, W. M. (1972), "The Bias of the Two-Stage Least Squares Estimator," Journal of the American Statistical Association, 67, 625-627.
- 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 American Statistical Association, 75, 230-237.
- Muirhead, R. J. (1975), "Expressions for some Hypergeometric Functions of Matrix Argument with Applications," Journal of Multivariate Analysis, 5, 283-293.
- 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.
- Olmsted, J. M. H. (1961), Advanced Calculus, New York: Appleton-Century-Crofts, Inc.
- Pan Jie-Jian (1968), "Distribution of the Noncircular Serial Correlation Coefficients." Am. Math. Soc. and Inst. Math. Statist. Selected Translations in Probability and Statistics, 7, 281-91.
- Petrov, V. V. (1975), Sums of Independent Random Variables (Berlin: Springer-Verlag).
- Pfanzagl, J. (1973), "Asymptotic Expansions Related to Minimum Contrast Estimators," Annals of Statistics, 1, 993-1026.
- Pfanzagl, J. (1980), comment on "Minimum Chi-Square, Not Maximum Likelihood!", The Annals of Statistics, 8, 479-481.

- 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, 91-8.
- Phillips, P. C. B. (1980a), "The Exact Finite Sample Density of Instrumental Variable Estimators in an Equation with $n+1$ Endogenous Variables," Econometrica, Vol. 48, No. 4, May 1980, pp. 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, Vol. 47, No. 1, January 1980, pp. 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," mimeographed, Cowles Foundation, Yale University.
- Phillips, P. C. B. (1982d), "Best Uniform Approximation of Probability Densities in Econometrics," in W. Hildenbrand (Ed.) Advances in Econometrics, Cambridge University Press.
- Phillips, P. C. B. and M. R. Wickens (1978), Exercises in Econometrics, Volumes I and II, Ballinger & Philip Allan.
- Pitman, E. J. G. (1937), "The 'Closest' Estimates of Statistical Parameters," Proceedings of the Cambridge Philosophical Society, 33, 212-22.
- Rainville, E. D. (1963), Special Functions, New York: Macmillan.
- Rao, C. R. (1973), "Linear Statistical Inference and its Applications," New York: Wiley.
- 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 "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. and DeMin 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 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," Mimeographed, Brown University.
- 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 (1970), "Least Squares and Grouping Method Estimators in the Errors in Variables Model," Journal of 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," mimeographed, University of California, Berkeley.
- Rothenberg, T. J. (1980), "Comparing Alternative Asymptotically Equivalent Tests" presented to World Congress of Econometric Society, Aix-en-Provence, 1980.

- Rothenberg, T. J. (1982), "Approximating the Distribution of Econometric Estimators and Test Statistics," in M. D. Intriligator and Z. Griliches (Eds.) Handbook of Econometrics, Vol. 2. Amsterdam: North Holland.
- 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. (1974), "The Validity of Nagar's Expansion for the Moments of Econometric Estimators." Econometrica, 42, 169-176.
- Sargan, J. D. (1975), "Gram-Charlier Approximations Applied to t-Ratios of k-Class Estimators," Econometrica 43, 326-346.
- Sargan, J. D. (1976a), "Econometric Estimators and the Edgeworth Approximation," Econometrica 44, 421-428, and (1977) "Erratum," Econometrica, 45, 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 Approximations 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.

- Sargan, J. D. and Y. K. Tse (1980), "Some Experience of Numerical Computation of Edgeworth Approximations," Discussion Paper No. A21, London School of Economics, mimeographed.
- 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, Series A, 61-66.
- Slater, L. J. (1965), "Confluent Hypergeometric Functions," in M. Abramowitz and I. A. Stegun, 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.
- Takeuchi, K. (1970), "Exact Sampling Moments of Ordinary Least Squares, Instrumental Variable and Two Stage Least Squares Estimators," International Economic Review, 11, 1-12.
- Tan, W. Y. (1969), "Note on the Multivariate and the Generalized Multivariate Beta Distributions," Journal of the American Statistical Association, 64, 230-241.

- Tse, Y. K. (1981), "Edgeworth Approximations in First Order Stochastic Difference Equations with Exogenous Variables," London School of Economics, mimeographed.
- 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 State 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.
- Wilks, S. S. (1962), Mathematical Statistics. J. Wiley & Sons.
- Williams, J. D. (1941), "Moments of the Ratio of the Mean Square Successive Difference to the Mean Square Difference in Samples from a Normal Universe," Annals of Mathematical Statistics, 12, 239-241.

- 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.
- 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, 185-183.
- 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.