

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
AT YALE UNIVERSITY

Box 2125, Yale Station
New Haven, Connecticut 06520

COWLES FOUNDATION DISCUSSION PAPER NO. 658*

*This is a revised and extended version of my CFDP No. 626.

Limited Circulation Only

This Cowles Foundation Discussion Paper is not for general circulation and is being privately circulated by the author alone. Its contents are strictly confidential and are not to be reproduced either in part or in whole without his authorization. Nor are the materials herein to be used in unpublished or published work without his explicit consent.

THE EXACT DISTRIBUTION OF LIML: I

by

Peter C. B. Phillips

December 16, 1982

THE EXACT DISTRIBUTION OF LIML: I

by

P. C. B. Phillips*

Cowles Foundation
Yale University

O. ABSTRACT

It is shown that the exact finite sample distribution of the limited information maximum likelihood (LIML) estimator in a general and leading single equation case is multivariate Cauchy. When the LIML estimator utilizes a known error covariance matrix (LIMLK) it is proved that the same Cauchy distribution still applies. The corresponding result for the instrumental variable (IV) estimator is a form of multivariate t density where the degrees of freedom depend on the number of instruments.

*The referees made helpful comments on the earlier version of this paper and, in particular, drew my attention to a deficiency in part of the original argument. My thanks, therefore, go first to them, then to Glenna Ames for her skill and effort in typing the manuscript of this paper and, finally, to the NSF for support under Grant Number SES 80007571.

1. INTRODUCTION

Improvements in the algebraic machinery of multivariate analysis have recently led to many advancements in our understanding of the finite sample properties of statistical methods in econometrics, particularly with regard to the simultaneous equations model. Modern multivariate methods provide a convenient stepping stone to the solution of exact sampling distribution problems through manageable algebraic representations of the joint density functions of the matrices of sample moments upon which most common econometric estimators depend. These matrix variates have, in general, noncentral multivariate distributions whose algebraic forms and properties have been intensively studied in mathematical statistics. Some of the most important contributions in this area have been made by Herz (1955), Constantine (1963), James (1954, 1960, 1964) and Davis (1980a, 1980b). All of these contributions have substantially facilitated the development of econometric small sample theory in recent years. A detailed account of the theoretical developments have taken place in econometrics, largely in conjunction with this analytic progress in multivariate methods, may be found in Mariano (1982) and Phillips (1980a, 1982a).

The purpose of the present paper is to focus on a simplified class of problems within the simultaneous equations setting where standard methods of multivariate analysis allow us to extract the exact distributions of econometric estimators with relative ease. Carefully chosen simplifications often enable us to work with central rather than noncentral distributions, thereby facilitating analytic derivations without sacrificing important elements of generality. The special models and leading cases we consider are discussed in Section 2 of the paper. Section 3 examines

the distribution of the IV estimator and Section 4 does the same for LIMLK, both for a leading case with an arbitrary number of included endogenous variables. The exact distribution of the LIML estimator is derived in Section 5 for the same leading case. The results are discussed in Section 6. A sequel to the paper (Phillips (1983)) gives the exact finite sample density of LIML in full generality using noncentral multivariate methods.

2. SPECIAL MODELS AND LEADING CASES

We will work with the structural equation

$$(1) \quad y_1 = Y_2\beta + Z_1\gamma + u$$

where $y_1(T \times 1)$ and $Y_2(T \times n)$ are an observation vector and observation matrix, respectively, of $n+1$ included endogenous variables, Z_1 is a $T \times K_1$ matrix of included exogenous variables, and u is a random disturbance vector. The reduced form of (1) is given by

$$(2) \quad [y_1 : Y_2] = [Z_1 : Z_2] \begin{bmatrix} \pi_{11} & \pi_{12} \\ \pi_{21} & \pi_{22} \end{bmatrix} + [v_1 : V_2] = Z\Pi + V,$$

where Z_2 is a $T \times K_2$ matrix of exogenous variables excluded from (1). The rows of the reduced form disturbance matrix V are assumed to be independent, identically distributed, normal random vectors. We assume that the usual standardizing transformations (see Phillips (1982a)) have been carried out so that the covariance matrix of each row of V is the identity matrix and $T^{-1}Z'Z = I_K$ where $K = K_1 + K_2$. We also assume that $K_2 \geq n$ so that the necessary order conditions for (1) to be identified are satisfied.

There are two special categories of models such as (1) and (2) in

which the exact density functions of the common single equation estimators of β in (1) can be extracted with relative ease. In the first category are the just identified structural models in which the usual consistent estimators all reduce to indirect least squares and take the form

$$(3) \quad \beta_{\text{ILS}} = [Z_2'Y_2]^{-1}[Z_2'y_1]$$

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 probability density function (p.d.f.) was first derived by Fieller (1932) and takes the following form here (see Mariano and McDonald (1979))

$$(4) \quad \text{pdf}(r) = \frac{\exp\left\{-\frac{\mu^2}{2}(1+\beta^2)\right\}}{\pi(1+r^2)} {}_1F_1\left(1, \frac{1}{2}; \frac{\mu^2}{2} \frac{(1+\beta r)^2}{1+r^2}\right)$$

where $\mu^2 = T\Pi_{22}'\Pi_{22}$ is the scalar concentration parameter. In the general case of $n+1$ included endogenous variables the density (4) is replaced by a multivariate analogue in which the ${}_1F_1$ function has a matrix argument (see Sargan (1976) and Phillips (1980b)).

The category of estimators that take the generic form of a matrix ratio of normal variates, as in (3), also includes the general IV estimator in the overidentified case provided the instruments are non-stochastic, that is, if $\beta_{\text{IV}} = [W'Y_2]^{-1}[W'y_1]$ 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), occurs in random coefficient models 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 (4) to the multivariate case in much the same way as the exact distribution theory for the IV estimator in the $n+1$ endogenous variable case.

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 simultaneous equations model. 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 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) noncentral distributions. 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 a more general setting.

It is with such leading cases that the present paper is concerned. We will consider, in particular, the leading subcase of (1) and (2) in which $\Pi_{22} = 0$. Under this hypothesis the reduced form (2) becomes

$$(2') \quad [y_1 : Y_2] = Z_1 [\pi_{11} : \pi_{12}] + [v_1 : V_2] .$$

3. THE IV ESTIMATOR

Statistical analysis of the leading case arising from (2') can be simply illustrated in terms of the IV estimator of β :

$$(5) \quad \beta_{IV} = [Y_2' Z_3 Z_3' Y_2]^{-1} [Y_2' Z_3 Z_3' y_1]$$

where $Z_3 (T \times K_3)$ is a submatrix of Z_3 forming instruments additional to Z_1 and where it is assumed that $K_3 \geq n$. This is the representation of the IV estimator given in equation (3) of Phillips (1980b). We note that the conditional distribution of $(T^{-1} Y_2' Z_3 Z_3' Y_2)^{-1/2} (T^{-1} Y_2' Z_3 Z_3' y_1)$ given $Z_3' Y_2$ is $N(0, I_n)$. This is independent of $Z_3' Y_2$ and is also, therefore, the unconditional distribution. Further, $T^{-1} Y_2' Z_3 Z_3' Y_2$ has a central Wishart distribution of order n with degrees of freedom K_3 and covariance matrix I_n . We may therefore write β_{IV} in the form

$$(6) \quad \beta_{IV} = [(T^{-1} Y_2' Z_3 Z_3' Y_2)^{-1/2}] [(T^{-1} Y_2' Z_3 Z_3' Y_2)^{-1/2} (T^{-1} Y_2' Z_3 Z_3' y_1)] \\ \equiv [W_n(K_3, I_n)]^{-1/2} N(0, I_n)$$

so that β_{IV} is proportional to a multivariate t variate (see, for example, Dickey (1967)). The p.d.f. of β_{IV} is therefore given by

$$(7) \quad \text{pdf}(r) = \frac{\Gamma\left(\frac{K_3+1}{2}\right)}{\pi^{n/2} \Gamma\left(\frac{K_3-n+1}{2}\right) (1+r'r)^{(K_3+1)/2}} = \frac{\Gamma\left(\frac{L+n+1}{2}\right)}{\pi^{n/2} \Gamma\left(\frac{L+1}{2}\right) (1+r'r)^{(L+n+1)/2}}$$

where $L = K_3 - n$ is the number of surplus instruments used in the estimation of β .

The density (7) specializes to the case of two stage least squares for $K_3 = K_2$ (where the result was given by Basmann (1974)) and to the case of ordinary least squares for $K_3 = T - K_1$ (where the result was given by Wegge (1971)). As shown in Phillips (1980b), (7) is in fact the leading term in the multiple series representation of the exact density of β_{IV} in the general single equation case where Π_{22} is not necessarily the zero matrix. Moreover, the leading marginal densities can be readily deduced from (7) (see Phillips (1982b)) and standard properties of the multivariate t confirm that integer moments exist up to the order L (i.e. the number of surplus instruments).

4. THE DISTRIBUTION OF LIMLK

In the overidentified case ($K_2 \geq n+1$), the LIML estimator, β_{LIML} , of β minimizes the ratio $\beta'_\Delta W \beta_\Delta / \beta'_\Delta S \beta_\Delta$ where $\beta'_\Delta = (1, -\beta')$, $W = X'(P_Z - P_{Z_1})X$, $S = X'(I - P_Z)X$ and where $X = [y_1 : y_2]$ and $P_A = A(A'A)^{-1}A'$. When the covariance matrix of the rows of X is known, the corresponding estimator is called LIMLK (see, for example, Anderson (1982)) and it will be denoted here by β_{LIMLK} . Since the model is already in canonical form, β_{LIMLK} minimizes the ratio $\beta'_\Delta W \beta_\Delta / \beta'_\Delta \beta_\Delta$ and satisfies the system

$$(8) \quad (W - \lambda_m I) \beta_\Delta = 0$$

where $\beta'_\Delta = (\beta_{\Delta 1}, \beta'_{\Delta 2})$ is the latent vector associated with the smallest latent root λ_m of W . This yields the estimator $\beta_{LIMLK} = -\beta_{\Delta 2} / \beta_{\Delta 1}$ by normalization. In the just identified case ($K_2 = n$), both LIML and LIMLK reduce to indirect least squares since the reduced form (2) is

unrestricted; and the analysis of the preceding section applies.

Under the null hypothesis that $\Pi_{22} = 0$ in (2), W has a central Wishart distribution $W_m(K_2, I)$, where $m = n+1 \leq K_2$, with density

$$(9) \quad \text{pdf}(W) = 2^{-mK_2/2} \left[\Gamma_m \left(\frac{K_2}{2} \right) \right]^{-1} \text{etr} \left(-\frac{1}{2}W \right) (\det W)^{(K_2-m-1)/2}.$$

In order to extract the exact distribution of β_{LIMLK} we introduce the orthogonal transformation H by which W is diagonalized so that $H'WH = \Lambda = \text{diag}(\lambda_1, \dots, \lambda_m)$. 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 (9) we can deduce the joint distribution of (Λ, H) . It is most convenient to work with the probability element $\text{pdf}(W)dW$. Under the transformation $W \rightarrow (\Lambda, H)$ we have

$$\text{pdf}(\Lambda, H)d\Lambda(dH) = \text{pdf}(W)dW$$

$$(10) \quad = \text{pdf}(W) \prod_{i < j} (\lambda_i - \lambda_j) \left(\prod_{i=1}^m d\lambda_i \right) (dH)$$

$$(11) \quad = 2^{-mK_2/2} \left[\Gamma_m \left(\frac{K_2}{2} \right) \right]^{-1} \exp \left(-\frac{1}{2} \sum_{i=1}^m \lambda_i \right) \left(\prod_{i=1}^m \lambda_i \right)^{(K_2-m-1)/2} \\ \cdot \prod_{i < j} (\lambda_i - \lambda_j) \left(\prod_{i=1}^m d\lambda_i \right) (dH).$$

Line (10) involves the jacobian of the transformation and is given, for example, in Constantine (1963, equation (43), p. 1280). (dH) is the invariant measure on the orthogonal group $O(m)$ (the group of orthogonal $m \times m$ matrices) that is normalized so that the measure over the whole group

(restricted so that $h_{1j} \geq 0$) is unity. The distribution of H is called the conditional Haar invariant distribution (see Anderson (1958, p. 322)). We see from (11) that H is distributed independently of the latent roots that form the diagonal elements of Λ , a result obtained by Anderson (1951).

To find the distribution of β_{LIMLK} we concentrate on the final column of H . We write this m -vector h , say, in partitioned form as $h' = (h_1, h_2')$. The invariant distribution of H implies an invariant measure for h over the Stiefel manifold defined by $h'h = h_1^2 + h_2'h_2 = 1$ and denoted by $V_{1,m}$. The latter is the unit sphere in $(n+1)$ -dimensional Euclidean space and the invariant measure on this manifold is given by the exterior differential form

$$(12) \quad (dh) = \bigwedge_{j=1}^n b_j' dh$$

(see equation (5.1) of James (1954)) where b_1, b_2, \dots, b_n are orthogonal column vectors all of which are orthogonal to h and \bigwedge denotes the wedge product. Using the parameterization of the manifold in which $h_1 = (1 - h_2'h_2)^{1/2}$ and restricting the region so that $h_1 > 0$, the invariant measure (12) can be written in the alternative form

$$(13) \quad (dh) = \frac{k dh_2}{(1 - h_2'h_2)^{1/2}}$$

(see Farrell (1976), equations (7.7.3-4)) where the constant k is selected so that the measure over the restricted ($h_1 > 0$) region of the unit sphere is unity. Since the measure over the entire unit sphere in R^{n+1} is $2\pi^{(n+1)/2} / \Gamma\left(\frac{n+1}{2}\right)$, that is the surface area of the sphere (see James (1954) equation (5.9)), the normalizing constant in the invariant measure (13)

over the restricted region is (ignoring questions of sign in (13) since we are working with a positive probability measure):

$$(14) \quad k = \pi^{-(n+1)/2} \Gamma\left(\frac{n+1}{2}\right).$$

We now renormalize the latent vector h to yield the LIMLK estimator.

Setting $\beta_{\text{LIMLK}} = r$, we find that this involves the transformation

$h_2 \rightarrow -r/(1+r'r)^{1/2}$ with $h_1 = (1-h_2'h_2)^{1/2} = (1+r'r)^{1/2}$. Taking differentials we deduce that

$$dh_2 = -(1+r'r)^{-1/2} [I+rr']^{-1} dr$$

and the modulus of the jacobian of the transformation is $(1+r'r)^{-(n+2)/2}$.

Thus, the invariant measure (13) defined over the appropriately restricted region (for which $h_1 > 0$) of the unit sphere in \mathbb{R}^{n+1} transforms as follows:

$$(dh) = \frac{\Gamma\left(\frac{n+1}{2}\right) dr}{\pi^{(n+1)/2} (1+r'r)^{(n+1)/2}}.$$

The p.d.f. of β_{LIMLK} then takes the form

$$(15) \quad \text{pdf}(r) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\pi^{(n+1)/2} (1+r'r)^{(n+1)/2}};$$

that is, a multivariate Cauchy distribution.

In the two endogenous variable case ($n = 1$), (15) reduces to the univariate Cauchy. It is an interesting exercise (that is left for the reader) to verify that this distribution provides the leading term in the multiple series representation of the exact density given by Mariano and McDonald (1979).

It is instructive to illustrate the workings of the argument leading to (15) in the restricted setting where $n = 1$. The orthogonal transformation H can be given the explicit form:

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

$$= \begin{bmatrix} -\sin \theta & \cos \theta \\ \cos \theta & \sin \theta \end{bmatrix}, \quad -\frac{\pi}{2} \leq \theta \leq 0$$

There is only one free variable, θ , in this representation of H .

Note that the domain of θ is restricted and the form of $H = (h_{ij})$ is defined to conform with the condition that $h_{ij} \geq 0$ and to allow a full range of values for the ratio h_{22}/h_{12} . The invariant measure (12), normalized and signed so that the measure over the manifold is unity, is

$$(dh) = \begin{cases} \left(-\frac{1}{\pi}\right) (\cos \theta, \sin \theta) \begin{pmatrix} -\cos \theta \\ -\sin \theta \end{pmatrix} d\theta = \frac{1}{\pi} d\theta, & \frac{3\pi}{2} \leq \theta \leq 2\pi \\ \left(\frac{1}{\pi}\right) (-\sin \theta, \cos \theta) \begin{pmatrix} -\sin \theta \\ \cos \theta \end{pmatrix} d\theta = \frac{1}{\pi} d\theta, & -\frac{\pi}{2} \leq \theta \leq 0. \end{cases}$$

We deduce that

$$\beta_{\text{LIMLK}} = -h_{22}/h_{12} = \begin{cases} \frac{\cos \theta}{\sin \theta} = \cot \theta, & \frac{3\pi}{2} \leq \theta \leq 2\pi \\ -\frac{\sin \theta}{\cos \theta} = -\tan \theta, & -\frac{\pi}{2} \leq \theta \leq 0 \end{cases}$$

and the density of β_{LIMLK} is:

$$\text{pdf}(r) = \begin{cases} \frac{1}{\pi} \left| \frac{d}{dr} (\cot^{-1}(r)) \right|, & r \leq 0 \\ \frac{1}{\pi} \left| \frac{d}{dr} (-\tan^{-1}(r)) \right|, & r \geq 0 \end{cases}$$

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

This specializes (15) to the univariate case.

5. THE DISTRIBUTION OF LIML

The LIML estimator satisfies the system

$$(16) \quad (W - \lambda S)\beta_{\Delta} = 0$$

where λ is the smallest latent root of the matrix $S^{-1/2}WS^{-1/2}$. Manipulation of (16) shows that β_{Δ} also satisfies

$$(17) \quad [S - f(W+S)]\beta_{\Delta} = 0$$

where $f = (1+\lambda)^{-1}$ is the largest latent root of $(W+S)^{-1/2}S(W+S)^{-1/2}$.

Let the $m = n+1$ roots of the equation

$$(18) \quad \det[S - f(W+S)] = 0$$

be ordered $f_1 > f_2 > \dots > f_m > 0$ (the probability of equal roots being zero) and assembled in the matrix $F = \text{diag}(f_1, f_2, \dots, f_m)$. Let the corresponding vectors g_i satisfying

$$(19) \quad [S - f_i(W+S)]g_i = 0$$

be normalized by

$$g_i'(W+S)g_i = 1$$

and assembled into the matrix $G = [g_1, \dots, g_m]$. It follows that

$$(20) \quad G'(W+S)G = I_m$$

(see, for example, Anderson (1958, p. 309). We now define $E = G^{-1}$ and consider the transformation $(S,W) \rightarrow (E,F)$ arising from the equations

$$(21) \quad S = E'FE, \quad W = E'(I-F)E.$$

This transformation is one to one if we impose a unique sign on a particular column of $E = (e_{ij})$. Let us take the final column and set $e_{i,n+1} \geq 0$ for all i . This requirement removes the indeterminacy in G associated with the fact that columns of G can be multiplied by -1 without disturbing the validity of (19) and (20).

The transformation (21) is discussed at length by Anderson (1958, pp. 310-313) where it is shown that the jacobian is

$$(22) \quad 2^m |\det E|^{m+2} \prod_{i < j} (f_i - f_j).$$

Under the null hypothesis that $\Pi_{22} = 0$ in (2), W and S have independent central Wishart distributions $W_m(K_2, I)$ and $W_m(T-K, I)$ respectively, with joint p.d.f.

$$(23) \quad \text{pdf}(W,S) = \frac{\text{etr} \left[-\frac{1}{2}(W+S) \right] (\det W)^{(K_2-m-1)/2} (\det S)^{(T-K-m-1)/2}}{2^{m(T-K_1)/2} \Gamma_m \left(\frac{K_2}{2} \right) \Gamma_m \left(\frac{T-K}{2} \right)}.$$

We deduce that

$$(24) \quad \text{pdf}(E,F) = \frac{\text{etr} \left[-\frac{1}{2}E'E \right] (\det E'E)^{(T-K_1)/2-(m+1)} |\det E|^{m+2}}{2^{m(T-K_1)/2-m} \Gamma_m \left(\frac{K_2}{2} \right) \Gamma_m \left(\frac{T-K}{2} \right)} \\ \cdot (\det F)^{(T-K-m-1)/2} (\det(I-F))^{(K_2-m-1)/2} \prod_{i < j} (f_i - f_j).$$

In order to integrate out F we will use the multivariate beta integral

$$(25) \quad \int_0^I (\det R)^{a-(m+1)/2} (\det(I-R))^{b-(m+1)/2} dR = \frac{\Gamma_m(a)\Gamma_m(b)}{\Gamma_m(a+b)} ; \quad a, b > (m-1)/2$$

(see, for example, Tan (1969)). Let H be an orthogonal matrix for which $H'RH = \text{diag}(r_1, r_2, \dots, r_m)$. To ensure the transformation is unique we order the roots as $r_1 > r_2 > \dots > r_m$ and specify the first element in each column of H to be positive. Under this transformation, as in (10) above, we have

$$(26) \quad dR = \prod_{i < j} (r_i - r_j) \prod_i dr_i (dH)$$

where (dH) is the invariant measure over the orthogonal group or, strictly speaking, that part of which is defined by the stated restriction on H . We know that

$$(27) \quad \int_{O(m)} (dH) = \frac{2^m \pi^{m^2/2}}{\Gamma_m\left(\frac{m}{2}\right)}$$

(James (1954)) so that taking the fraction $1/2^m$ of (27) corresponding to the restriction on H and using (25) and (26) we deduce that

$$(28) \quad \int_0^I (\prod_i r_i)^{a-(m+1)/2} (\prod_i (1-r_i))^{b-(m+1)/2} \prod_{i < j} (r_i - r_j) \prod_i dr_i \\ = \frac{\Gamma_m(a)\Gamma_m(b)\Gamma_m\left(\frac{m}{2}\right)}{\Gamma_m(a+b)\pi^{m^2/2}}.$$

It follows that

$$(29) \quad \text{pdf}(E) = \frac{\Gamma_m\left(\frac{m}{2}\right)}{2^{\frac{m(T-K_1)}{2}-m} \pi^{\frac{m^2}{2}} \Gamma_m\left(\frac{T-K_1}{2}\right)} \text{etr}\left(-\frac{1}{2}E'E\right) (\det E'E)^{\frac{(T-K_1)}{2}-(m+1)} |\det E|^{m+2}.$$

It is now convenient to partition the inverse of E , viz. G , as follows

$$(30) \quad G = \begin{matrix} & \begin{matrix} 1 & n \end{matrix} \\ \begin{matrix} 1 \\ n \end{matrix} & \begin{bmatrix} g_{11} & g'_{12} \\ g_{21} & G_{22} \end{bmatrix} \end{matrix} = \begin{bmatrix} g_{11} & g'_{12} \\ -g_{11}^r & G_{22} \end{bmatrix}$$

where we use the fact that

$$(31) \quad \beta_{\text{LIML}} = -g_{21}/g_{11} = r, \text{ say.}$$

The corresponding partition of E is

$$E = \left[\begin{array}{c|c} \begin{matrix} -1 & -1 \\ g_{11} & -g_{11}g'_{12}(G_{22} + rg'_{12})^{-1}r \end{matrix} & \begin{matrix} -g_{11}g'_{12}(G_{22} + rg'_{12})^{-1} \\ -g_{11}g'_{12}(G_{22} + rg'_{12})^{-1} \end{matrix} \\ \hline \begin{matrix} (G_{22} + rg'_{12})^{-1}r \\ (G_{22} + rg'_{12})^{-1} \end{matrix} & \begin{matrix} (G_{22} + rg'_{12})^{-1} \\ (G_{22} + rg'_{12})^{-1} \end{matrix} \end{array} \right]$$

$$= \begin{bmatrix} e_{11} & e'_{12} \\ E_{22}^r & E_{22} \end{bmatrix}, \text{ say.}$$

Thus, from (29) and (32) we find

$$\begin{aligned}
(33) \quad \text{pdf}(e_{11}, e_{12}, r, E_{22}) &= \frac{\Gamma_m\left(\frac{m}{2}\right)}{2^{\frac{m(T-K_1)}{2}-m} \pi^{\frac{m^2}{2}} \Gamma_m\left(\frac{T-K_1}{2}\right)} \\
&\cdot \exp\left[-\frac{1}{2}(e_{11}^2 + e_{12}'e_{12})\right] \text{etr}\left[-\frac{1}{2}(I+rr')E_{22}'E_{22}\right] \\
&\cdot [\det(E_{22}'E_{22})]^{(T-K_1-m+1)/2} |e_{11} - e_{12}'r|^{T-K_1-m}.
\end{aligned}$$

E_{22} occurs in the density (33) only in the form $E_{22}'E_{22}$. Moreover, when E_{22} is integrated out of the density the value of the integral will be unchanged if we relax the (positive) sign requirement on the final column of E_{22} and multiply the integral by 2^{-n} . To assist in the integration we transform $E_{22} \rightarrow (H, D)$ where H is orthogonal and $D = (E_{22}'E_{22})^{1/2}$ according to the unique decomposition $E_{22} = HD$. The measure changes in accordance with the relation

$$(34) \quad dE_{22} = 2^{-n}(\det D)^{-1/2} dD(dH)$$

proved by James (1954) (see also Muirhead (1982)) where (dH) is the invariant measure on $O(n)$.

From (33) and (34) we deduce that

$$\begin{aligned}
(35) \quad \text{pdf}(e_{11}, e_{12}, r, D) &= \frac{\Gamma_m\left(\frac{m}{2}\right) \pi^{n^2/2}}{2^{\frac{m(T-K_1)}{2}-1} \pi^{\frac{m^2}{2}} \Gamma_m\left(\frac{T-K_1}{2}\right) \Gamma_n\left(\frac{n}{2}\right)} \\
&\cdot \exp\left[-\frac{1}{2}(e_{11}^2 + e_{12}'e_{12})\right] \text{etr}\left[-\frac{1}{2}(I+rr')D\right] \\
&\cdot (\det D)^{(T-K_1-m)/2} |e_{11} - e_{12}'r|^{T-K_1-m}.
\end{aligned}$$

Integrating out D in (35) and simplifying the constant we obtain

$$(36) \quad \text{pdf}(e_{11}, e_{12}, r) = \frac{\Gamma\left(\frac{n+1}{2}\right) \exp\left[-\frac{1}{2}(e_{11}^2 + e_{12}'e_{12})\right] |e_{11} - e_{12}'r|^{T-K_1-m}}{2^{\frac{(T-K_1)/2-1}{2}} \pi^{\frac{n+1}{2}} \Gamma\left(\frac{T-K_1}{2} - \frac{n}{2}\right) (1+r'r)^{\frac{(T-K_1)/2}{2}}}$$

Henceforth, we assume $T - K_1 - m$ to be an even integer, facilitating the reduction of (36). In particular, let $T - K_1 - m = 2k$ for some integer k . Define $q' = (e_{11}, e_{12}')$ and $a' = (1, r')$. We need to evaluate

$$(37) \quad \int e^{-q'q/2} (a'q)^{2k} dq$$

where the integral is over $0 < q_m < \infty$, $-\infty < q_i < \infty$, $i \neq m$. We will transform $q \rightarrow H'q = p$ for H orthogonal and integrate over the orthogonal group, normalized so that the measure over the whole group is unity. The latter measure will be denoted (dH) .

In view of the fundamental relationship (James (1964, equation (22)):

$$(38) \quad \int_{O(m)} (\text{tr}(XH))^{2k} (dH) = \sum_{\kappa} \frac{\left(\frac{1}{2}\right)_{\kappa}}{\left(\frac{m}{2}\right)_{\kappa}} C_{\kappa}(XX')$$

where κ is a partition of k into not more than m parts, we deduce that

$$(39) \quad \int_{O(m)} (\text{tr}(pa'H))^{2k} (dH) = \frac{\left(\frac{1}{2}\right)_{\kappa} (a'a)^k (p'p)^k}{\left(\frac{m}{2}\right)_{\kappa}}$$

It follows that (37) is equal to

$$(40) \quad \frac{\left(\frac{1}{2}\right)_{\kappa} (a'a)^k}{2 \left(\frac{m}{2}\right)_{\kappa}} \int e^{-p'p/2} (p'p)^k dp .$$

The additional factor of $(1/2)$ in the expression arises because the domain of p is taken to be unrestricted (i.e. $-\infty < p_i < \infty$ all i) while that of q in (37) satisfies $0 < q_m < \infty$.

We transform $p \rightarrow (s, h)$ according to the decomposition $p = (p'p)^{1/2}(p/(p'p)^{1/2}) = s^{1/2}h$. The measure changes in accordance with the relation

$$(41) \quad dp = 2^{-1} s^{(n-1)/2} ds(dh)$$

where (dh) denotes the invariant measure over the Stiefel manifold, $V_{1,m}$ (see James (1954, equation (8.19))). Under this transformation (40) becomes

$$\begin{aligned} & \frac{\left(\frac{1}{2}\right)_k (a'a)^k}{4 \left(\frac{m}{2}\right)_k} \int_0^\infty e^{-s/2} s^{k+(n-1)/2} ds \int_{V_{1,m}} (dh) \\ &= \frac{\left(\frac{1}{2}\right)_k (a'a)^k \Gamma\left(\frac{n+1}{2} + k\right)}{4 \left(\frac{m}{2}\right)_k \left(\frac{1}{2}\right)_k^{(n+1)/2+k}} \cdot \frac{2\pi^{(n+1)/2}}{\Gamma\left(\frac{n+1}{2}\right)}. \end{aligned}$$

We deduce from (36) and (42) that

$$\text{pdf}(r) = \frac{\left(\frac{1}{2}\right)_k \pi^{(n+1)/2} \Gamma\left(\frac{n+1}{2} + k\right) \Gamma\left(\frac{n+1}{2}\right) (1+r'r)^k}{2^{(T-K_1)/2} \left(\frac{1}{2}\right)_k^{(n+1)/2+k} \pi^{n+1/2} \Gamma\left(\frac{T-K_1}{2} - \frac{n}{2}\right) \left(\frac{m}{2}\right)_k \Gamma\left(\frac{n+1}{2}\right) (1+r'r)^{(T-K_1)/2}}$$

Since $k = (T - K_1 - m)/2 = (T - K_1 - (n+1))/2$, it follows that

$$(43) \quad \text{pdf}(r) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\pi^{(n+1)/2} (1+r'r)^{(n+1)/2}}$$

Thus, β_{LIML} has a multivariate Cauchy distribution.

6. DISCUSSION

The exact distributions obtained above apply when $\Pi_{22} = 0$ and correspond, therefore, to a particular structure of the model (1) and (2) in which the true coefficient vector β is not identifiable. None of the exact finite sample densities (7), (15) or (43) actually involve β . All are, in fact, centered on the origin. When β is itself zero, there is an absence of simultaneity in the model and, in this case, OLS (with $K_3 = T - K_1$ in (7)) is a consistent estimator. We observe, on the other hand, that the exact densities of LIML, LIMLK and 2SLS (the latter with $K_3 = K_2$ in (7)) are invariant to changes in the sample size T . Thus, as $T \rightarrow \infty$ these distributions (unlike that of OLS) continue to demonstrate the uncertainty about β that is due to its lack of identification.

Our results show that LIML and LIMLK have identical finite sample Cauchy distributions in this leading case. Knowledge of the error covariance matrix therefore adds nothing to the precision of the LIML estimator. While the lack of identification of β has a role in explaining this fact, the crucial factor behind the result is that an entire block of reduced form coefficients, viz. $[\pi_{21} : \Pi_{22}]$, is zero under the null hypothesis. This implies that the system of equations $\pi_{21} - \Pi_{22}\beta = 0$ which usually define β , also carry no real restrictions on the reduced form. Estimation by LIML or LIMLK under these conditions is essentially (that is, from the point of view of their distributional properties) equivalent to estimation under conditions of (apparent) just identification. This explains not only why LIML and LIMLK have identical distributions; it also explains why these distributions, viz. (15) and (43), are invariant to changes in the (apparent)

degree of overidentification. Readers who find these intuitive arguments convincing can rely directly on the elementary proof given in Section 3 of the exact distribution of the IV estimator for the (apparent) just identified case to deduce the results for LIML and LIMLK obtained by more sophisticated methods in Sections 4 and 5.

Finally, it is of interest that the exact distributions studied here in primitive forms retain certain important properties, notable their tail area behavior, when the null hypothesis $\Pi_{22} = 0$ is relaxed. This is already confirmed in the two endogenous variable case where the exact distribution of the LIML estimator was derived under the alternative $\Pi_{22} \neq 0$ by Mariano and Sawa (1972). A sequel to the present paper will provide the exact distribution of LIML under the same general alternative $\Pi_{22} \neq 0$ but for a structural equation containing an arbitrary number of endogenous variables.

REFERENCES

- Anderson, T. W. (1951). "The Asymptotic Distribution of Certain Characteristic Roots and Vectors," Proceedings of Second Berkeley Symposium on Statistics and Probability, University of California Press, pp. 103-130.
- Anderson, T. W. (1958). An Introduction to Multivariate Statistical Analysis. New York: J. Wiley & Sons.
- 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.
- Constantine, A. G. (1963), "Some Noncentral Distribution Problems in Multivariate Analysis," Annals of Mathematical Statistics, 34, 1270-1285.
- 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.
- Dickey, J. M. (1967), "Matricvariate Generalisations of the Multivariate t Distribution and the Inverted Multivariate t Distribution," Annals of Mathematical Statistics, 38, 511-518.
- Farrell, R. H. (1976), Techniques of Multivariate Calculation, Springer: New York.
- Pieiller, E. C. (1932), "The Distribution of the Index in a Normal Bivariate Population," Biometrika, 24, 428-440.
- Herz, C. W. (1955), "Bessel Functions of Matrix Argument," Annals of Mathematics 61, 474-523.

- James, A. T. (1954), "Normal Multivariate Analysis and the Orthogonal Group," Annals of Mathematical Statistics, 25, 40-75.
- James, A. T. (1960), "The Distribution of the Latent Roots of the Covariance Matrix, Annals of Mathematical Statistics, 31, 151-158.
- James, A. T. (1964), "Distribution of Matrix Variates and Latent Roots Derived from Normal Samples," Annals of Mathematical Statistics, 35, 475-500.
- Kelejian, H. H. (1974), "Random Parameters in a Simultaneous Equation Framework: Identification and Estimation," Econometrica, 42, 517-527.
- 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.
- Muirhead, R. J. (1982), Aspects of Multivariate Statistical Theory, Wiley: New York.
- Phillips, P. C. B. (1980a), "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. (1980b), "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. 861-868.
- Phillips, P. C. B. (1982a), "Exact Small Sample Theory in the Simultaneous Equations Model," in M. D. Intriligator and Z. Griliches (Eds.), Handbook of Econometrics, North Holland: Amsterdam.
- Phillips, P. C. B. (1982b), "Marginal Densities of Instrumental Variable Estimators in the General Single Equation Case," Advances in Econometrics (forthcoming).
- Phillips, P. C. B. (1983), "The Exact Distribution of LIML: II" (in preparation).
- 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. (1976), "Econometric Estimators and the Edgeworth Approximation," Econometrica 44, 421-428, and (1977) "Erratum," Econometrica, 15, 272.
- Tan, W. Y. (1969), "Note on the Multivariate and the Generalized Multivariate Beta Distributions," Journal of the American Statistical Association, 64, 230-241.
- Wegge, L. L. (1971), "The Finite Sampling Distribution of Least Squares Estimators with Stochastic Regressors," Econometrica, 38, 241-251.