ON THE EXACT DISTRIBUTION OF LIML

Peter C. B. Phillips

March 26, 1982
ON THE EXACT DISTRIBUTION OF LIML

by

P. C. B. Phillips*

Cowles Foundation
Yale University

I. ABSTRACT

It is shown that the exact distribution of the LIML estimator in a general and leading single equation case is multivariate Cauchy. The corresponding result for the IV estimator is a form of multivariate t density where the degrees of freedom depend on the number of instruments.

*My thanks go to Glena Ames for skill and effort in typing the manuscript of this paper and to the NSF for support under Grant Number SES 8007571.
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 by Herz (1955), Constantine (1963), James (1964) and Davis (1980a, 1980b), all of which have substantially facilitated the development of econometric small sample theory in recent years. A detailed account of the theoretical developments that 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.
2. **LEADING CASES AND THE INSTRUMENTAL VARIABLE (IV) ESTIMATOR**

We will work with the structural equation

(1) \[ y_1 = Y_2 \beta + Z_1 \gamma + u \]

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) \[ [y_1 : y_2] = [Z_1 : Z_2] \begin{bmatrix} \Pi_{11} & \Pi_{12} \\ \Pi_{21} & \Pi_{22} \end{bmatrix} + [v_1 : v_2] = Z\Pi + V, \]

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 rows of \( V \) is the identity matrix and \( T^{-1}Z'Z = I_K \) where \( K = K_1 + K_2 \). We also assume that \( K_2 > n \) 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 \( \beta \) 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) \[ \beta_{ILS} = [Z_2'Y_2]^{-1}[Z_2'Y_1] \]
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 Fieiller (1932) and takes the following form here (see Mariano and McDonald (1979))

\[
\text{pdf}(r) = \frac{\exp \left( -\frac{\mu^2}{2} (1 + \beta^2) \right)}{\pi (1 + r^2)} \text{Hypergeometric}_1 \left[ \frac{1}{2}, \frac{1}{2}; \frac{\mu^2}{2}, \frac{(1 + \beta r)^2}{1 + r^2} \right]
\]

where $\mu^2 = \text{tr} \Sigma_{22}^1$ 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 $\text{Hypergeometric}_1$ 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 $\beta_{IV} = [W'Y_2]^{-1}[W'Y_1]$ 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

\[(2') \quad [y_1^*; Y_2] = Z_1[\Pi_{11}^*; \Pi_{12}^*] + [v_1^*; V_2].\]

The statistical analysis of this leading case can now be simply illustrated in terms of the following IV estimator of $\beta$:

\[(5) \quad \beta_{IV} = [Y_2^*Z_3^*Z_3^*Y_2]^{-1}[Y_2^*Z_3^*Z_3^*Y_1]\]

where $Z_3(T \times K_3)$ is a submatrix of $Z_2$ selected as instruments additional to $Z_1$ and where it is assumed that $K_3 > n$. We note that the conditional distribution of $T^{-1/2}[Y_2^*Z_3^*Z_3^*Y_2]$ 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, $T^{-1}Y_2'Z_3Z_2'y_2$ has a central Wishart distribution of order $n$ with degrees of freedom $K_3$ and covariance matrix $I_n$. We may therefore write $\beta_{IV}$ in the form

$$\beta_{IV} = \left[(T^{-1}Y_2'Z_3Z_2'y_2)^{-1/2}\right] \left[(T^{-1}Y_2'Z_3Z_2'y_2)^{-1/2}(T^{-1}Y_2'Z_3Z_3'y_1)\right]$$

$$\equiv \left[\hat{\theta}_n(K_3, I_n)^{-1/2}\right] N(0, I_n)$$

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

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

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

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 (1982b), (7) is in fact the leading term in the multiple series representation of the exact density of $\beta_{IV}$ in the general single equation case where $\Pi_{22}$ is not necessarily the zero matrix. Moreover, the leading marginal densities can be readily deduced from (7) (see Phillips (1982c)) and standard properties of the multivariate $t$ confirm that integer moments exist up to the order $L$ (i.e. the number of surplus instruments).
3. **THE DISTRIBUTION OF LIML**

The LIML estimator $\hat{\beta}_{\text{LIML}}$ of $\beta$ minimizes the ratio $\hat{\beta}'_\Delta W_\Delta^\Delta / \hat{\beta}'_\Delta S_\Delta^\Delta$

where $\hat{\beta}'_\Delta = (1, -\beta')$, $W = X'(P_\Delta - P_\Delta 1)X$, $S = X'(I - P_\Delta)X$ and where $X = [y_1 y_2]$ and $P_\Delta = A(A'A)^{-1}A'$. Under the null hypothesis that $\pi_{22} = 0$ in (2) we deduce that $W$ and $S$ have central Wishart distributions $W_m(K_2, I)$ and $W_m(T-K, I)$ respectively, where $m = n+1$. Since $(P_\Delta - P_\Delta 1)(I - P_\Delta) = 0$ , $W$ and $S$ are independent with joint p.d.f. given by

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

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

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

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

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

for $\text{Re}(Z) > 0$ and $\text{Re}(a) > (m-1)/2$ we deduce from (9) the marginal matrix density of $F$. 

\[
\text{pdf}(F) = \frac{\Gamma\left[\frac{T-K_1}{2}\right]}{\Gamma\left[\frac{K_2}{2}\right]} \left(\frac{(K_2-m-1)/2}{\Gamma_m}\right)^{-(T-K_1)/2} (K_2 - m - 1)/2 \\
\left(\det F\right)^{-(T-K_1)/2} \left[\det(I + F)\right]^{-(T-K_1)/2} \\
= \left[b\left[\frac{K_2}{2}, \frac{T-K_1}{2}\right]\right]^{-1} \left(\frac{K_2 - m - 1}{2}\right)^{-(T-K_1)/2} (K_2 - m - 1)/2 \\
\left(\det F\right)^{-(T-K_1)/2} \left[\det(I + F)\right]^{-(T-K_1)/2}
\]

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

Expression (11) now leads simply to the p.d.f. of the LIML estimator \( \beta_{\text{LIML}} \). We note that if \( \lambda_m \) is the smallest latent root of \( F = S^{-1/2}WS^{-1/2} \) then \( \beta_{\text{LIML}} \) satisfies the system

\[
(S^{-1/2}WS^{-1/2} - \lambda_m I)\beta_{\Delta} = (F - \lambda_m I)\beta_{\Delta} = 0
\]

where \( \beta_{\Delta} = (\beta_{\Delta 1}, \beta_{\Delta 2})' \) is the latent vector associated with the smallest latent root \( \lambda_m \) and \( \beta_{\text{LIML}} = -\beta_{\Delta 2}/\beta_{\Delta 1} \).

In order to extract the exact distribution of \( \beta_{\text{LIML}} \) we introduce the orthogonal transformation \( H \) by which \( F \) is diagonalized so that \( H'FH = \Lambda = \text{diag}(\lambda_1, \ldots, \lambda_n) \). This transformation is unique if we specify that \( \lambda_1 > \lambda_2 > \ldots > \lambda_m \) 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 (11) we can deduce the joint distribution of \( (\Lambda, H) \).

It is most convenient in the general multivariate case to work with the probability element \( \text{pdf}(F)dF \). Under the transformation \( F \to (\Lambda, H) \) we then have
\[
\text{pdf}(\Lambda, \mathcal{H})d\Lambda(d\mathcal{H}) = \text{pdf}(F)dF
\]

(13) 
\[
= \text{pdf}(F) \prod_{i<j} (\lambda_i - \lambda_j)(\prod_{i=1}^{m} d\lambda_i)(d\mathcal{H})
\]

(14) 
\[
= \left[ B_n \left( \frac{K_2}{2}, \frac{T-K_1}{2} \right) \right]^{-1} \prod_{i=1}^{m} (\lambda_i)^{(K_2-m-1)/2} 
\]
\[
\cdot \prod_{i=1}^{m} (1+\lambda_i)^{-\frac{T-K_1}{2}} \prod_{i<j}^{m} (\lambda_i - \lambda_j)(\prod_{i=1}^{m} d\lambda_i)(d\mathcal{H}).
\]

Line (13) follows from Constantine (1963, equation (43), p. 1280) and line (14) follows directly from (11). \((d\mathcal{H})\) 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 \(h_{1j} > 0\)) is unity. The distribution of \(\Lambda\) is called the conditional Haar invariant distribution (see Anderson (1958, p. 322)) and we see from (14) that \(\Lambda\) is distributed independently of the latent roots that form the diagonal elements of \(\Lambda\). To find the distribution of \(\Lambda_{\text{LIML}}\) we now need to concentrate on the final column of \(\mathcal{H}\) which we write as the \((n+1)\)-vector \(\mathbf{h}\) or in partitioned form as \(\mathbf{h}' = (h_1, h_2)\). The invariant distribution of \(\mathcal{H}\) implies an invariant measure over the Stiefel manifold defined by 
\[
\mathbf{h}' \mathbf{h} = h_1^2 + h_2^2 = 1.
\]
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

(15) 
\[
(d\mathbf{h}) = \sum_{j=1}^{n} b_j^\dagger d\mathbf{h}
\]

where \(b_1, b_2, \ldots, b_n\) are orthonormal column vectors orthogonal to \(\mathbf{h}\).
(see equation (5.1) of James (1954)). 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 (15) can be written in the alternative form

\[
(dh) = \frac{k dh_2}{(1 - h_2^2 h_2')^{1/2}}
\]

(16)

(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 \( \mathbb{R}^{n+1} \) 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 (16) over the restricted region is (ignoring questions of sign in (16) since we are working with positive probability measures):

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

(17)

We now renormalize the latent vector \( h \) to yield the LLML estimator. This involves the transformation \( h_2 = -r / (1 + r' r)^{1/2} \) with \( h_1 = (1 - h_2^2 h_2')^{1/2} = (1 + r' r)^{-1/2} \). Taking differentials we deduce that

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

and the modulus of the jacobian of the transformation is \( (1 + r' r)^{- (n+2)/2} \). Thus, the invariant measure (16) defined over the appropriately restricted region (for which \( h_1 > 0 \)) of the unit sphere in \( \mathbb{R}^{n+1} \) transforms as follows:
\[ (dh) = \frac{\Gamma \left[ \frac{n+1}{2} \right] dr}{\pi^{(n+1)/2} \left( 1 + \mathbf{r}^\mathbf{r} \right)^{(n+1)/2}}. \]

The p.d.f. of \( \beta_{\text{LIML}} \) then takes the form

\[ \text{pdf}(r) = \frac{\Gamma \left[ \frac{n+1}{2} \right]}{\pi^{(n+1)/2} \left( 1 + \mathbf{r}^\mathbf{r} \right)^{(n+1)/2}}, \]

that is, a multivariate Cauchy distribution.

In the two endogenous variable case \((n = 1)\), (19) reduces to the univariate Cauchy, which provides the leading term in the multiple series representation of the exact density given by Mariano and McDonald (1979) in the general case. We deduce directly from (18) that \( \beta_{\text{LIML}} \) has no finite moments of integral order, as was originally shown by Mariano and Sawa (1972) and Sargan (1970).

4. CONCLUDING REMARKS

The exact distributions obtained above apply when \( \pi_{22} = 0 \) and correspond, therefore, to a particular structure of the model in which the true coefficient vector \( \beta \) is not identifiable. Neither of the exact densities (7) or (19) actually involve \( \beta \). Both are in fact centered around the origin. When \( \beta \) 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 consistent. In both this case \((\beta = 0)\) and for other values of \( \beta \) the exact densities of LIML and 2SLS (the latter with \( K_3 = K_2 \) in (7)) are invariant to changes in the sample size \( T \). Thus, as \( T \to \infty \) these distributions continue to demonstrate the uncertainty about \( \beta \) due to the lack of identification. It is of interest that in the more general case
where $\Pi_{22} \neq 0$ the exact distributions of 2SLS and LIML retain certain important properties (such as their tail area behavior) which apply in the primitive forms of these densities given in (7) and (19) above.
REFERENCES


