Institute of Economic Development and Research
SCHOOL OF ECONOMICS
University of the Philippines

Discussion Paper No. 70-9 April 22, 1970

# EXACT FINITE-SAMPLE DISTRIBUTION OF THE LIMITED-INFORMATION MAXIMUM LIKELIHOOD ESTIMATOR IN THE CASE OF TWO INCLUDED ENDOGENOUS VARIABLES

By

1944-
*ROBERTO S. MARIANO* and *TAKAMITSU SAWA*

NOTE: This paper has been prepared for submission to an academic journal. Critical comments will be appreciated by the authors.

Summary

This paper presents the exact finite-sample distribution of the limited-information maximum likelihood (LIML) estimator when the structural equation being estimated contains two endogenous variables and is identifiable in a complete system of linear stochastic equations. Since the density function obtained is an infinite series of a complicated form it is difficult to deduce meaningful conclusions about the exact distribution of the LIML estimator. However, it reveals the important fact that for arbitrary values of the parameters in the model, the LIML estimator does not possess moments of order greater than or equal to one.

# Exact Finite-Sample Distribution of the Limited-Information Maximum Likelihood Estimator in the Case of Two Included Endogenous Variables[1]

ROBERTO S. MARIANO[2] and TAKAMITSU SAWA[3]
University of the Philippines and Kyoto University

## 1. INTRODUCTION

The exact finite-sample distributions of the ordinary least squares and the two-stage least squares estimators have been derived in a fairly general model. For example, see Richardson (10) and Sawa (11). In this paper we will further derive the exact distribution of the limited-information maximum likelihood (LIML) estimator in the same frame-work as in Sawa (11), which will be briefly described in Section 2.

Our derivation of the distribution of the LIML estimator essentially depends upon Anderson and Girshick (5)'s non-central Wishart distribution. The well-known Wishart distribution is the distribution of the sample covariance matrix from a multivariate normal population with constant mean vector and constant covariance matrix. The non-central Wishart distribution is the distribution of the sample covariance matrix when the observations arise from a set of normal multivariate populations with constant covariance matrix and expected values that vary from observation to observation.

The stochastic process underlying a complete system of linear stochastic equations may be regarded as a non-central situation in the sense that the endogenous variables are supposed to be normal variates with constant covariance matrix and expected values that vary corresponding to the values of the exogenous variables. Thus, the non-central Wishart distribution theory provides a starting point for finding the distribution of various estimators of parameters in a complete system of linear stochastic equations. Some important theorems about the non-central Wishart distribution and other related theorems are summarized without proof in the Appendix.

## 2. SPECIFICATIONS AND NOTATION

Suppose we are interested in the distribution of the LIML estimator of $\beta$ in the structural equation

$$y_2 = y_1\beta + Z_1\gamma_1 + Z_2\gamma_2 + u \tag{2.1}$$

in a complete system of linear stochastic equations, where $y_1$ and $y_2$ are N-component vectors of independent observations on two endogenous variables; $Z_1$ is a N by $K_1$ matrix of N observations on $K_1$ exogenous variables and $Z_2$ is a N by $K_2$ matrix of N observations on $K_2( = K-K_1)$ exogenous variables; $\beta, \gamma_1$ and $\gamma_2$ are unknown structural parameters; and u is a N-component vector of disturbance terms.

Our system is assumed to include no lagged endogenous variables and to be composed of at least two equations. The

$K_2(\geq 1)$ structural parameters $\gamma_2$ are supposed to be zero for the identifiability of the structural equation (2.1).

The reduced form equations for $y_1$ and $y_2$ are

$$Y = Z_1\Pi_1 + Z_2\Pi_2 + V, \qquad (2.2)$$

where $Y = (y_1\ y_2)$; $\Pi_1 = (\Pi_{11}\ \Pi_{12})$ and $\Pi_2 = (\Pi_{21}\ \Pi_{22})$ are respectively $K_1$ by 2 and $K_2$ by 2 matrices of unknown coefficients; $V$ is a $N$ by 2 matrix of random disturbance terms. The rows of $V$ are assumed to be mutually independent and identically distributed as normal with mean vector 0 and 2 by 2 positive definite covariance matrix

$$\Sigma = \|\sigma_{ij}\| . \qquad (2.3)$$

For the sake of the existence of the LIML estimators, we further assume that $N \geq K+2$.

As is well known, the LIML method makes use of all the exogenous variables in the system in order to estimate the parameters of a single equation, but does not require a detailed specification of the other structural equations in the system. Therefore the specifications described above are sufficient for us to derive the sampling distribution of the LIML estimator. That is to say, our following results hold true whatever the remaining part of the complete system may be.

In the following, we assume for convenience that

$$N^{-1}Z_1'Z_1 = I_{K_1},\ N^{-1}Z_2'Z_2 = I_{K_2},\ Z_1'Z_2 = 0, \qquad (2.4)$$

where $I_n$ is the identity matrix of dimension $n$. It can be verified that these assumptions cause no loss of generality. For example, see Anderson and Rubin (6).

## 3. JOINT DISTRIBUTION OF THE CHARACTERISTIC ROOTS AND VECTORS

Define the following 2 by 2 matrices:

$$A^* = (Y - Z_1\hat{\Pi}_1)'(Y - Z_1\hat{\Pi}_1) = Y'Y - N\hat{\Pi}_1'\hat{\Pi}_1, \qquad (3.1)$$

$$G^* = N\hat{\Pi}_2'\hat{\Pi}_2, \qquad (3.2)$$

$$W^* = A^* - G^* = Y'Y - N\hat{\Pi}_1'\hat{\Pi}_1 - N\hat{\Pi}_2'\hat{\Pi}_2, \qquad (3.3)$$

where $\hat{\Pi}_1$ and $\hat{\Pi}_2$ are the ordinary least squares estimators of the reduced form coefficients $\Pi_1$ and $\Pi_2$ respectively. Obviously $W^*$ is the residual sum of squares matrix of the bivariate normal regression model (2.2). Hence from Lemma A.2 in the Appendix, we can immediately see that $W^*$ is distributed according to the central Wishart distribution $W(\Sigma, N - K + 1)$. Referring to the usual multivariate normal regression theory, it is obvious that $\hat{\Pi}_2$ is independent of $W^*$ and each element of $\hat{\Pi}_2$, i.e. $\hat{\pi}_{ik}^{(2)}$, is normally distributed with mean value.

$$E(\pi_{ik}^{(2)}) = \pi_{ik}^{(2)} \qquad (i = 1,\ldots K_2;\ k = 1,2), \qquad (3.4)$$

and variance-covariance

$$E(\hat{\pi}_{ik}^{(2)} - \hat{\pi}_{ik}^{(2)})(\hat{\pi}_{jh}^{(2)} - \pi_{jh}^{(2)}) = \frac{1}{N}\sigma_{ij}\delta_{kh}(i,j=1,\ldots,K_2;\ k,h=1,2), \tag{3.5}$$

where the $\delta_{kh}$'s are Kronecker's deltas and $\pi_{ik}^{(2)}$ is the (i,k) element of $\Pi_2$. Hence the $\sqrt{N}(\hat{\pi}_{1k}^{(2)}\ \pi_{2k}^{(2)})$'s are mutually independent and normally distributed with non-identical mean vector $\sqrt{N}(\pi_{1k}^{(2)}\ \pi_{2k}^{(2)})$ and identical covariance matrix $\Sigma$.

From (2.1), (2.2), (2.4) and the identifiability restrictions, it follows that

$$\beta\Pi_{21} = \Pi_{22}. \tag{3.6}$$

Then the means sigma matrix (see Lemma A.1 in the Appendix) of $\sqrt{N}(\hat{\pi}_{1k}^{(2)}\ \pi_{2k}^{(2)})$ is

$$T^* = N\Pi_2'\Pi_2 = (N\Pi_{21}'\Pi_{21})\begin{pmatrix}1 & \beta\\ \beta & \beta^2\end{pmatrix}, \tag{3.7}$$

The rank of $T^*$ is clearly equal to 1. Hence, from Lemma A.2 and the above remarks, $G^*$ is seen to be independent of $W^*$ and distributed according to the non-central Wishart distribution $W'(\Sigma,T^*,1,K_2 + 1)$.

Now, consider the determinantal equation

$$|W^* - \lambda A^*| = |W^* - \lambda(W^* + G^*)| = 0. \tag{3.8}$$

Clearly, the roots of this equation are positive and less than unity. Let the roots be ordered such that $0<\lambda_1<\lambda_2<1$ (since the probability of two roots being equal is 0). Let $u_1^{*\prime} = (u_{11}^* \; u_{21}^*)$ and $u_2^{*\prime} = (u_{12}^* \; u_{22}^*)$ be vectors such that

$$(W^* - \lambda_i A^*)u_i^* = 0, \; i = 1,2. \tag{3.9}$$

Then the LIML estimate of $\beta$, i.e. $\hat{\beta}$, is given by[4]

$$\hat{\beta} = - \frac{u^*_{12}}{u^*_{22}} , \tag{3.10}$$

(see Anderson and Rubin (7)).

Let $\tau^2$ be the positive root of the determinantal equation

$$|T^* - \lambda\Sigma| = 0. \tag{3.11}$$

Then there exists a nonsingular 2 by 2 matrix $\Phi = \|\phi_{ij}\|$ such that

$$\Phi' \Sigma \Phi = I_2, \tag{3.12}$$

and

$$\Phi' T^* \Phi = \begin{pmatrix} \tau^2 & 0 \\ 0 & 0 \end{pmatrix} = T. \tag{3.13}$$

By using (3.7), it can be shown that

$$\tau^2 = \frac{(\sigma_{22} - 2\beta\sigma_{12} + \beta^2\sigma_{11})}{\sigma_{11}\sigma_{22} - \sigma_{12}^2} (N\Pi_{21}'\Pi_{21}) \tag{3.14}$$

$$= \frac{\nu^2}{|\Sigma|}(N\Pi_{21}'\Pi_{21}).$$

where

$$\nu^2 = (\sigma_{22} - 2\beta\sigma_{12} + \beta^2\sigma_{11}) = (\beta \;\; 1)\Sigma\begin{pmatrix}\beta\\1\end{pmatrix} . \tag{3.15}$$

And from (3.12) and (3.13), it can be deduced that

$$\Phi^{-1} = \frac{1}{\nu}\begin{bmatrix} |\Sigma|^{\frac{1}{2}} & \beta|\Sigma|^{\frac{1}{2}} \\ \beta\sigma_{11}-\sigma_{12} & \beta\sigma_{12}-\sigma_{22}\end{bmatrix} . \tag{3.16}$$

If we now let

$$W = \Phi' \, W^* \, \Phi,$$

$$G = \Phi' \, G^* \, \Phi,$$

then it follows from Lemma A.3 that $W$ and $G$ are independently distributed according to $W(I \;, N-K+1)$ and $W'(I_2, T, 1, K_2+1)$ respectively. Thus the joint probability density function of $W$ and $G$ is given by

$$C_1 e^{-\frac{1}{2}\tau^2} |W|^{\frac{1}{2}(N-K-3)} \, |G|^{\frac{1}{2}(K_2-3)} \exp\left[-\tfrac{1}{2}\mathrm{tr}\,(W+G)\right] H_{\frac{1}{2}(K_2-2)}(\tau\sqrt{g_{11}}) \tag{3.17}$$

where $g_{11}$ is the (1,1)th element of the matrix $G$ and

$$H_m(x) = \sum_{\alpha=0}^{\infty} \frac{1}{\alpha!\,\Gamma(m+\alpha+1)}\left(\frac{x}{2}\right)^{2\alpha} , \tag{3.18}$$

$$C_1^{-1} = 2^{N-K_1} \, \pi\Gamma\left[\tfrac{1}{2}(N-K)\right] \Gamma\left[\tfrac{1}{2}(N-K-1)\right] \Gamma\left[\tfrac{1}{2}(K_2-1)\right] \Gamma\left[\tfrac{1}{2}K_2\right]. \tag{3.19}$$

Since

$$|W - \lambda(W + G)| = |\Phi'| \cdot |W^* - \lambda(W^* + G^*)| \cdot |\Phi| , \tag{3.20}$$

the roots of (3.8) are equal to the roots of

$$| W - \lambda(W + G) | = 0. \tag{3.21}$$

The corresponding vector satisfying

$$\left[W - \lambda_i(W + G)\right] u_i = 0 , \quad i = 1,2 \tag{3.22}$$

is simply the linear transformation of $u_i^*$ such that

$$u_i = \Phi^{-1} u_i^* . \tag{3.23}$$

Let $U = (u_1 \ u_2)$, $\Lambda = \|\lambda_i \ \delta_{ij}\|$ and $E = \|e_{ij}\| = U^{-1}$.

To remove the indeterminacy of the vectors $u_i$, let

$$U' (W + G)U = I_2 \tag{3.24}$$

and

$$e_{i1} \geq 0 \quad (i=1,2) . \tag{3.25}$$

Then the one-to-one correspondence between $(W,G)$ and $(E,\Lambda)$ can be asserted. From (3.22) and (3.24) it follows that

$$W = E' \Lambda E , \qquad G = E' (I - \Lambda)E. \tag{3.26}$$

The Jacobian of this transformation is

$$4|E|^4 \ (\lambda_2-\lambda_1) \ . \tag{3.27}$$

See, for example, Anderson (3). Hence after a little rearrangement and letting $p=\frac{1}{2}(N-K-3)$ and $q=\frac{1}{2}(K_2-3)$, we obtain the joint probability density function of $E$ and $\Lambda$:

$$4C_1 e^{-\frac{1}{2}\tau^2} |E'E|^{\frac{1}{2}(N-K_1-2)} \exp\left(-\tfrac{1}{2}\mathrm{tr}E'E\right)$$

$$\lambda_1^p \lambda_2^p \ (\lambda_2-\lambda_1) \sum_{i,j=0}^{\infty} \frac{\tau^{2(i+j)}(1-\lambda_1)^{i+q}(1-\lambda_2)^{j+q} \ e_{11}^{2i} \ e_{21}^{2j}}{i! \ j! \ \Gamma(\frac{K_2}{2} + i + j) \quad 2^{2(i+j)}} \tag{3.28}$$

for $e_{11} \geq 0$, $e_{21} \geq 0$, $E$ non-singular and $0<\lambda_1<\lambda_2<1$.

## 4. EXACT DISTRIBUTION OF THE LIML ESTIMATOR

Let $r = -\frac{e_{12}}{e_{11}}$. From (3.10), (3.23) and the definition of $E$, we have

$$r = -\frac{e_{12}}{e_{11}} = \frac{u_{12}}{u_{22}} = \frac{\phi^{11}\hat{\beta} - \phi^{12}}{\phi^{21}\hat{\beta} - \phi^{22}} \ , \tag{4.1}$$

where $\Phi^{-1} = \|\phi^{ij}\|$. In this section, we shall proceed to obtain the probability density function of $r$ and from this, use (4.1) to derive the density of $\hat{\beta}$.

Now, integrate (3.28) with respect to $\lambda_1$ and $\lambda_2$ and let $r = -\frac{e_{12}}{e_{11}}$ be a transformation from $e_{12}$ to $r$.

Term by term integration of (3.28) is allowed and thus, the joint probability density function of $(e_{11}, e_{21}, e_{22}, r)$ for $e_{11} \geq 0$, $e_{21} \geq 0$, $-\infty < e_{22} < \infty$, $-\infty < r < \infty$, is

$$4C_1 e^{-\frac{1}{2}\tau^2} e^{-\frac{1}{2}\left[e_{11}^2(1+r^2) + e_{21}^2 + e_{22}^2\right]} e_{11}^{N-K_1-1} |e_{22} + re_{21}|^{N-K_1-2}$$

$$\times \sum_{i,j=0}^{\infty} \gamma(i,j)\, e_{11}^{2i}\, e_{21}^{2j} \quad , \qquad (4.2)$$

where

$$\gamma(i,j) = \frac{(\frac{\tau}{2})^{2(i+j)}}{i!\, j!\, \Gamma(\frac{K_2}{2} + i + j)} \omega(i,j), \qquad (4.3)$$

$$\omega(i,j) = \int_0^1 \int_0^{\lambda_2} (\lambda_1 \lambda_2)^p (1-\lambda_1)^{i+q} (1-\lambda_2)^{j+q} (\lambda_2 - \lambda_1)\, d\lambda_1\, d\lambda_2 . \qquad (4.4)$$

and, as in the previous section,

$$p = \frac{1}{2}(N-K-3)$$

$$q = \frac{1}{2}(K_2 - 3) \quad . \qquad (4.5)$$

The actual evaluation of $\omega(i,j)$ is given in Lemma A.4 of the Appendix.

Term by term integration of (4.2) with respect to $e_{11}$ leads to the following joint probability density function of $(e_{21}, e_{22}, r)$ for $e_{21} \geq 0$, $-\infty < e_{22} < \infty$, $-\infty < r < \infty$:

$$4C_1 e^{-\frac{1}{2}\tau^2} e^{-\frac{1}{2}(e_{21}^2 + e_{22}^2)} |e_{22} + re_{21}|^{N-K_1-2}$$

$$\times \sum_{i,j=o}^{\infty} \gamma(i,j) 2^{\frac{1}{2}(N-K_1)+i-1} \Gamma(\tfrac{1}{2}(N-K_1)+i) \frac{e_{21}^{2j}}{(1+r^2)^{\frac{1}{2}(N-K_1)+i}} . \quad (4.6)$$

For simplicity, let us now assume that $N-K_1$ is an even number greater than or equal to two.[5]

Then

$$|e_{22} + re_{21}|^{N-K_1-2} = \sum_{k=0}^{N-K_1-2} \binom{N-K_1-2}{k} e_{22}^{k} (re_{21})^{N-K_1-2-k} \quad (4.7)$$

and hence, the joint probability density function of $(e_{21}, e_{22}, r)$ as given by (4.6) simplifies to

$$4C_1 e^{-\frac{1}{2}\tau^2} e^{-\frac{1}{2}(e_{21}^2 + e_{22}^2)} \sum_{k=0}^{N-K_1-2} \sum_{i,j=0}^{\infty} \gamma(i,j) \binom{N-K_1-2}{k} \Gamma(\tfrac{1}{2}(N-K\ )+i)$$

$$2^{\frac{1}{2}(N-K_1)+i-1} e_{22}^{k} e_{21}^{2j+N-K_1-2-k} \frac{(r^2)^{\frac{1}{2}(N-K_1-k)-1}}{(1+r^2)^{\frac{1}{2}(N-K_1)+i}} . \quad (4.8)$$

To derive the marginal probability density function of $r$, we integrate (4.8) with respect to $e_{21}$ and $e_{22}$. Term by term integration is permissible and hence we obtain the following expression for the density of $r$, for $-\infty < r < \infty$:

$$e^{-\frac{1}{2}\tau^2} \sum_{h=0}^{\frac{1}{2}(N-K_1-2)} \sum_{i,j=0}^{\infty} \tau^{2(i+j)} \delta(i,j,h) \frac{(r^2)^{\frac{1}{2}(N-K_1)-h-1}}{(1+r^2)^{\frac{1}{2}(N-K_1)+i}} , \quad (4.9)$$

where

$$\delta(i,j,h) = 2^{N-K_1-i-j} C_1 \binom{N-K_1-2}{2h} \omega(i,j)$$

$$\frac{\Gamma(h+\frac{1}{2})\ \Gamma(\frac{1}{2}(N-K_1)+i)\Gamma(\frac{1}{2}(N-K_1-1)+j-h)}{\Gamma(i)\ \Gamma(j)\ \Gamma(\frac{1}{2}K_2+i+j)} , \quad (4.10)$$

and $\omega(i,j)$ is the double integral given in (4.4). Finally, by making use of (4.1), we obtain the following probability density function of $\hat{\beta}$, for $-\infty < \hat{\beta} < \infty$:

$$e^{-\frac{1}{2}\tau^2} |\Sigma|^{\frac{1}{2}} \sum_{h=0}^{\frac{1}{2}(N-K_1-2)} \sum_{i,j=0}^{\infty} \{ \delta(i,j,h)\tau^{2(i+j)}$$

$$\frac{\left[(\phi^{11}\hat{\beta}-\phi^{12})^2\right]^{\frac{1}{2}(N-K_1)-h-1} \left[(\phi^{21}\hat{\beta}-\phi^{22})^2\right]^{h+i}}{\left[(\phi^{11}\hat{\beta}-\phi^{12})^2 + (\phi^{21}\hat{\beta}-\phi^{22})^2\right]^{\frac{1}{2}(N-K_1)+i}} \}, (4.11)$$

with $\delta(i,j,h)$ as given in (4.10).

If $\beta = \frac{\sigma_{12}}{\sigma_{11}}$, note from (3.16) that $\phi^{21} = 0$, and thus, it can be seen from (4.11) that in this special case, $\hat{\beta}$ is symmetrically distributed around the true parameter $\beta$. In this respect, the LIML estimator shares a common characteristic with the ordinary and two-stage least squares estimators (see Sawa (11) ).

In the general case where assumption (2.4) concerning the orthogonality of the columns of Z is <u>not</u>

necessarily satisfied, the means sigma matrix for $G^*$ is given by

$$T^{**} = \mu^2 \begin{pmatrix} 1 & \beta \\ \beta & \beta^2 \end{pmatrix} , \qquad (4.12)$$

where $\mu^2 = \Pi_{21}' Z_2'(I-Z_1(Z_1'Z_1)^{-1}Z_1')Z_2\Pi_{21}$. Comparing the above expression with (3.7), we deduce that the probability density function of $\hat{\beta}$ in this case is also given by (4.11) with $\tau^2$ replaced by $\lambda^2$, the positive characteristic root of $\Sigma^{-1}T^{**}$.

Using (3.16), we can further express the density of $\beta$ in terms of the original parameters of our model. We state the final result in the following theorem:

<u>Theorem 4.1.</u> <u>For the case where the structural equation being estimated contains two endogenous variables and</u> $K_1$ <u>exogenous variables such that</u> $N-K_1$ <u>is an even number greater than one, the probability density function of the LIML estimator</u> $\hat{\beta}$ <u>is</u>

$$e^{-\frac{1}{2}\lambda^2} \sum_{h=o}^{\frac{1}{2}(N-K_1-2)} \sum_{i,j=o}^{\infty} \{\delta(i,j,h)\lambda^{2(i+j)} \frac{(\sigma_{11}\sigma_{22}-\sigma_{12}^2)^{\frac{1}{2}(N-K_1-1)-h}}{(\sigma_{22}-2\beta\sigma_{12}+\beta^2\sigma_{11})^{\frac{1}{2}(N-K_1)+i-1}}$$

$$\frac{\left[(\hat{\beta}-\beta)^2\right]^{\frac{1}{2}(N-K_1)-h-1}\left[\sigma_{22}-\sigma_{12}(\beta+\hat{\beta})+\sigma_{11}\beta\hat{\beta}\right]^{2(h+i)}}{(\sigma_{22}-2\sigma_{12}\hat{\beta}+\sigma_{11}\hat{\beta}^2)^{\frac{1}{2}(N-K_1)+i}}\} ,$$

for $-\infty < \hat{\beta} < \infty$ ; where

$$\lambda^2 = \frac{\mu^2(\sigma_{22}-2\beta\sigma_{12}+\beta^2\sigma_{11})}{\sigma_{11}\sigma_{22}-\sigma_{12}^2},$$

$$\mu^2 = \Pi_{21}'Z_2'(I-Z_1(Z_1'Z_1)^{-1}Z_1')Z_2\Pi_{21},$$

and $\delta(i,j,h)$, as given in (4.10), depends on the parameters $N, K_1, K_2$.

## 5. NON-EXISTENCE OF MOMENTS

The mathematical structure of the density function just derived is so complicated that it is difficult to deduce any more definite conclusions about the form of the density function of the LIML estimator. However, it reveals the important fact that the LIML estimator does not possess even the first-order moment, regardless of the sample size or the number of the identifiability restrictions. This is an immediate consequence of Theorem 4.1 since by elementary calculus, the integral

$$\int_{-\infty}^{\infty} \frac{x^r}{(ax^2 + bx + c)^m} dx$$

converges if and only if $2m > r + 1$, where $a \neq 0$.

For other single-equation estimators, Sawa (11) has shown that the ordinary and the two-stage least squares estimators of $\beta$ possess finite moments up to order $N-2$ and $K_2-1$ respectively.

However, the existence or non-existence of moments alone can hardly be used as a basis for determining the merits of the various estimators. For instance, one may be comparing a Cauchy distribution which is very concentrated and a normal distribution with an extremely large standard deviation. Finally, the main result in this paper indicates that in comparing the small-sample properties of the LIML method with other estimation methods, moments should not be used as a criterion of the goodness of an estimator. Such a criterion will clearly give results unfavorable to the LIML estimator, since it does not possess moments of any order.

APPENDIX

Suppose the p-componet vectors $X_1, \ldots, X_N$ $(N > p)$ are mutually independent such that $X\alpha$ has the distribution $N(\mu_\alpha, \Sigma)$ where the mean vectors $\mu_1, \mu_2, \ldots, \mu_N$ need not be all identical. Let

$$A = \sum_{\alpha=1}^{N} (X_\alpha - \overline{X})(X_\alpha - \overline{X})'$$

$$T = \sum_{\alpha=1}^{N} (\mu_\alpha - \overline{\mu})(\mu_\alpha - \overline{\mu})'$$

where

$$\overline{X} = \frac{1}{N} \sum_{\alpha=1}^{N} X_\alpha, \quad \overline{\mu} = \frac{1}{N} \sum_{\alpha=1}^{N} \mu_\alpha,$$

and let the rank of $T$ be $t$. $A$ is said to have a Wishart distribution of order $p$, with $N-1$ degrees of freedom, covariance matrix $\Sigma$ and means sigma matrix $T$. This distribution is denoted by $W'(\Sigma, T, t, N)$. If $t = o$, the distribution is said to be central and is denoted by $W(\Sigma, N)$.

Lemma A.1. For $t = 1$, the joint probability density function of A(see Anderson (4) and Anderson and Girshick (5)) is

$$K \, |A|^{\frac{1}{2}(N-p-2)} \exp\left(-\tfrac{1}{2}\mathrm{tr}\, \Sigma^{-1}A\right)$$

$$\times \left(\mathrm{tr}\, A\Sigma^{-1}T\Sigma^{-1}\right)^{-\frac{1}{4}(N-3)} \exp\left(-\tfrac{1}{2}\mathrm{tr}\Sigma^{-1}T\right)$$

$$\times I_{\frac{1}{2}(N-3)} (\sqrt{\mathrm{tr}\, A\Sigma^{-1}T\Sigma^{-1}}),$$

where

$$K^{-1} = |\Sigma|^{\frac{1}{2}(N-1)} \, 2^{\frac{1}{2}p(N-1)} \, \pi^{\frac{1}{4}p(p-1)} \prod_{i=1}^{p-1} \Gamma(\tfrac{1}{2}(N-1-i))$$

and

$$I_n(x) = \sum_{\alpha=0}^{\infty} \frac{(\frac{x}{2})^{n+2\alpha}}{\alpha!\,\Gamma(n+\alpha+1)} .$$

Lemma A.2. Suppose the p-component vectors $X_1,\ldots,X_N$ $(N>p)$ are mutually independent such that $X_\alpha$ has the distribution $N(BZ_\alpha,\Sigma)$ where the $Z_\alpha$'s are non-stochastic K-component vectors and $B$ is a pxK matrix of unknown constants. Let $\hat{B} = \sum_{\alpha=1}^{N} X_\alpha Z_\alpha' C^{-1}$ where $C = \sum_{\alpha=1}^{N} z_\alpha z_\alpha'$ is assumed to be non-singular. Then $\sum_{\alpha=1}^{N} X_\alpha X_\alpha' - \hat{B}C\hat{B}'$ is distributed according to $W(\Sigma, N-K+1)$ and is independent of $\hat{B}$.

For the proof of this lemma, see Anderson (3), pp. 83 - 84 and p. 183.

Lemma A.3. Suppose $A$ is a $p$ by $p$ matrix and is distributed as $W'(\Sigma, T, t, N)$. Then $\psi' A \psi$ is distributed as $W'(\psi'\Sigma\psi, \psi' T \psi, t, N)$, where $\psi$ is an arbitrary $p$ by $p$ nonsingular matrix.

Lemma A.4. Let p, q and $\omega(i,j)$ be as given in (4.4) - (4.5). Then, for $i+j+K_2>2$,

$$\omega(i,j) = \frac{B(2p+3,\ j+q+1)}{(p+1)(p+2)}\ {}_3F_2\ (p+1,\ 2p+3,\ -i-q;2p+j+q+4,p+3;\ 1).$$

Proof. Denote the complete and incomplete Beta functions (for $r, s>0$) and the generalized hypergeometric series by the following, respectively:

$$B(r,s) = \int_0^1 x^{r-1}(1-x)^{s-1}dx \tag{A.1}$$

$$B_t(r,s) = \int_0^t x^{r-1}(1-x)^{s-1}dx,\ 0\leq t<1. \tag{A.2}$$

$$_mF_n(a_1,\ldots,a\ ;\ b\ ,\ldots,b\ ;\ x) = \sum_{h=0}^{\infty} \frac{(a_1)_h\cdot\ldots\cdot(a_m)_h}{(b_1)_h\cdot\ldots\cdot(b_n)_h}\frac{x^h}{h!} \tag{A.3}$$

where $(a)_0=1$ and $(a)_h= a(a+1)\ldots(a+h-1)$ for $h=1,2,\ldots$ .

Note that

$$B_t(r,s) = \frac{t^r}{r}\ {}_2F_1(r,\ 1-s;\ r+1;\ t) \tag{A.4}$$

(for example, see Erdelyi (8), p. 373). We also refer to

$$\frac{1}{\Gamma(z)}\int_0^1 f(y)(1-y)^{z-1}dy,\ z>0$$

as the Riemann-Liouville integral of $f(y)$ of order $z$, (see Erdelyi (9), p. 181). Now, for $\omega(i,j)$, we have to evaluate integrals of the form

$$J(r,s,u,v) = \int_0^1 \int_0^y x^{r-1}(1-x)^{s-1}y^{u-1}dxdy$$

$$= \int_0^1 B_y(r,s)y^{u-1}\ (1-y)^{v-1}\ dy \qquad (A.5)$$

by (A.2). For $0<c<1$, by the dominated convergence theorem, we have

$$J(r,s,u,v) = \lim_{c\to 1} \int_0^1 B_{cy}(r,s)y^{u-1}(1-y)^{v-1}dy$$

$$= \lim_{c\to 1} J_c(r,s,u,v), \text{ say.} \qquad (A.6)$$

Furthermore, by (A.4), $J_c(r,s,u,v)$ is the Riemann-Liouville integral of $\left\{\frac{c^r}{r}\,{}_2F_1(r,1-s;\ r+1;\ cy)\ y^{r+u-1}\right\}$ of order $v$, and hence, from Erdelyi (9), p. 200, we get

$$J_c(r,s,u,v) = \frac{c^r}{r}\ B(r+u,v)\ {}_3F_2(r+u,r,1-s;\ r+u+v,\ r+1;\ c). \qquad (A.7)$$

From (4.4) in section 4,

$$\omega(i,j) = J(p+1,i+q+1,p+2,j+q+1) - J(p+2,i+q+1,p+1,j+q+1)$$

$$= B(2p+3,j+q+1) \lim_{c\to 1} \left\{\frac{1}{p+1}\ {}_3F_2(2p+3,p+1,-i-q;\ 2p+j+q+4,\ p+2;\ c) - \frac{1}{p+2}\ {}_3F_2(2p+3,p+2,-i-q;\ 2p+j+q+4;\ p+3;\ c)\right\} \qquad (A.8)$$

$$= \frac{B(2p+3,j+q+1)}{(p+1)(p+2)} \lim_{c\to 1} {}_3F_2(p+1,2p+3,-i-q;\ 2p+j+q+4,p+3;\ c). \tag{A.9}$$

(A.8) follows from (A.7) and (A.9) is a simplification of (A.8).

Now, consider the hypergeometric series in (A.9). In absolute value, each term in the expansion of this series is less than or equal to the corresponding term of ${}_2F_1(2p+3, -i-q;\ 2p+j+q+4;\ c)$ which in turn absolutely converges for $|c| = 1$ if $i+j+2q+1>0$, (see Abramowitz and Stegun (1), p. 556). Hence the hypergeometric series in (A.9) uniformly converges for $|c|\le 1$ and

$$\omega(i,j) = \frac{B(2p+3,j+q+1)}{(p+1)(p+2)}\ {}_3F_2(p+1,2p+3,-i-q;\ 2p+j+q+4,$$

$p+3;\ 1)$ for $i+j+2q+1>0$ or equivalently, for $i+j+K_2>2$, by (4.5) of section 4.

## FOOTNOTES

[1]The results in this paper were arrived at independently by the two authors. The authors are grateful for the helpful comments of Professor T.W. Anderson in the preparation of this paper.

[2]Part of a Ph.D. dissertation submitted to the Department of Statistics of Stanford University. The author is sincerely grateful for the suggestions and guidance given by his adviser, Professor T.W. Anderson, in the preparation of his dissertation. The author also acknowledges the support of the Rockefeller Foundation and the National Science Foundation (GS 2635) for the research presented here.

[3]Sawa's original paper was printed as the discussion paper No. 020, Kyoto Institute of Economic Research, Kyoto University.

[4]Consider the case where the equation being estimated contains no exogenous variables so that $\gamma_1 = 0$ in (2.1). If we redefine the matrices $A^*$, $G^*$, $W^*$ and $T^*$ as follows:

$$A^* = Y'Y$$

$$G^* = Y'(ZZ')Y \qquad = N\hat{\Pi}'\hat{\Pi}$$

$$W^* = Y'(I - ZZ')Y \quad = Y'Y - N\hat{\Pi}'\hat{\Pi}$$

$$T^* = N\Pi'\Pi$$

where $Z = (Z_1 \; Z_2)$ and $\Pi' = (\Pi_1' \; \Pi_2')$, then $\hat{\beta}$ is as given in (3.10) and $W^*$ and $G^*$ are independently distributed as $W(\Sigma, N - K + 1)$ and $W'(\Sigma, T^*, 1, K + 1)$, respectively.

Thus, the additional assumption that there are no exogenous variables in the equation being estimated leads only to changes in the degrees of freedom and the means sigma matrix for the non-central Wishart matrix $G^*$, the rank of the means sigma matrix being the same.

It can be shown that the above remarks also hold for the general case where there is an arbitrary number of endogenous variables in the equation being estimated.

[5]To derive the density of $r$, we assumed that $N_1 - K_1$ is even so that we can use the binomial expansion given by (4.7).

For the general case of arbitrary $N-K_1$ (not necessarily even), we can proceed from (4.2), let $y = \frac{e_{22}}{re_{21}}$ be a transformation from $e_{22}$ to $y$, derive the joint pdf of $(e_{11}, e_{21}, r, y)$ and integrate this expression with respect to $e_{11}$, $e_{21}$ and $y$. The resulting marginal density of $r$ is:

$$4C_1 e^{-\frac{1}{2}\tau^2} \sum_{i,j=0}^{\infty} \gamma(i,j) \, 2^{N-K_1-2+i+j} \, \Gamma\left(\frac{N-K_1}{2}+i\right) \Gamma\left(\frac{N-K_1}{2}+j\right)$$

$$\frac{(r^2)^{\frac{1}{2}(N-K_1-1)}}{(1+r^2)^{\frac{1}{2}(N-K_1)+i}} \int_{-\infty}^{\infty} \frac{|1+y|^{N-K_1-2}}{(1+r^2y^2)^{\frac{1}{2}(N-K_1)+j}} \, dy$$

for $-\infty < r < \infty$. It can be verified that the above expression simplifies to (4.8) if $N-K_1$ is assumed to be an even number.

## REFERENCES


(1) Abramowitz, M. and I. Stegum (eds.)(1965). Handbook of Mathematical Functions, Dover Publications, New York.

(2) Anderson, T.W. (1946). The non-central Wishart distribution and certain problems of multivariate statistics, Ann. Math. Stat., 17, 409-431.

(3) Anderson, T.W. (1958). An Introduction to Multivariate Statistical Analysis, J. Wiley and Sons, New York.

(4) Anderson, T.W. (1964). Corrections to "Some extensions of the Wishart distribution" and "The non-central Wishart distribution and certain problems of multivariate statistics", Ann. Math. Stat., 35, 923-924.

(5) Anderson, T.W. and M.A. Girshick (1944). Some extensions of the Wishart distribution, Ann. Math. Stat., 15, 345-357.

(6) Anderson, T.W. and H. Rubin (1949). Estimation of the parameters of a single equation in a complete system of stochastic equations, Ann. Math. Stat., 20, 46-63.

(7) 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, Ann. Math. Stat., 21, 570-582.

(8) Erdelyi, A. (ed.) (1954). Tables of Integral Transforms, Vol. 1, McGraw-Hill, New York.

(9) Erdelyi, A. (ed.) (1954). Tables of Integral Transforms, Vol. II, McGraw-Hill, New York.

(10) Richardson, D.H. (1968). The exact distribution of a structural coefficient estimator, J. Amer. Stat. Assoc., 63, 1214-1226.

(11) Sawa, T. (1969). The exact sampling distribution of ordinary least squares and two-stage least squares estimators, J. Amer. Stat. Assoc., 64, 923-937.