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

Discussion Paper No. 70-10 April 23, 1970

APPROXIMATIONS TO THE DISTRIBUTION FUNCTIONS OF THE ORDINARY LEAST SQUARES AND TWO-STAGE LEAST SQUARES ESTIMATORS IN THE CASE OF TWO INCLUDED ENDOGENOUS VARIABLES

By

ROBERTO S. MARIANO, 1944-

SUMMARY

This paper deals with single-equation estimators in a simultaneous system of linear stochastic equations. Under the assumptions that all predetermined variables in the model are exogenous and that the equation being estimated contains two endogenous variables, the distribution function of the two-stage least squares (2SLS) estimator is approximated up to terms whose order of magnitude is $\frac{1}{\sqrt{N}}$, where N is the sample size.

The 2SLS estimator is expressed in terms of mutually independent bivariate normal random vectors. By applying Taylor series expansions to this expression, it is shown that the 2SLS distribution function is equal to the sum of three terms, the first being the limiting standard normal distribution function, the second a correction term which is $O(\frac{1}{\sqrt{N}})$ and the third a remainder term which is $O(\frac{1}{N})$ as $N \to \infty$. It is shown that the correction term is $O(\frac{1}{\mu})$ and the remainder term $O(\frac{1}{\mu^2})$ as $\mu \to \infty$, where μ^2 is what is referred to in the literature as the <u>concentration parameter</u>.

For fixed N, an approximation to the OLS distribution function is also obtained up to terms whose order of magnitude is $O(\frac{1}{\mu})$ as $\mu \to \infty$.

APPROXIMATIONS TO THE DISTRIBUTION FUNCTIONS OF THE ORDINARY LEAST SQUARES AND TWO-STAGE LEAST SQUARES ESTIMATORS IN THE CASE OF TWO INCLUDED ENDOGENOUS VARIABLES[1]

ROBERTO S. MARIANO[2]
UNIVERSITY OF THE PHILIPPINES

1. INTRODUCTION

Two well-known single-equation estimators in a simultaneous system of linear stochastic equations are the limited-information maximum likelihood (LIML) and the two-stage least squares (2SLS) estimators. These two estimators are known to be asymptotically normal and equivalent, under appropriate conditions, as the sample size increases, (see Anderson and Rubin (3) and Basmann (4).)

Under the assumptions that the predetermined variables in the model are exogenous and the equation being estimated is identified and contains two endogenous variables, with both the number of exogenous variables excluded and the number of equations in the

---

[1]Based on a dissertation submitted to the Department of Statistics, Stanford University for a Ph.D. degree in Statistics. The author is sincerely grateful for the suggestions and guidance given by his adviser, Professor T.W. Anderson, in the preparation of his dissertation.

[2]The author acknowledges the support of the Rockefeller Foundation and the National Science Foundation (GS-2635) for the research presented here.

model being arbitrary, Richardson (11) and Sawa (12) have derived the exact probability distribution of the 2SLS estimator. For more special cases for which the exact 2SLS distribution is obtained, see Basmann (5,6), Bergstrom (7) and Kabe (8,9). More recently, the exact probability distribution of the LIML estimator in the case discussed by Sawa and Richardson has also been derived, (see Mariano and Sawa (10)).

Unfortunately, the expressions derived for the exact distributions of the 2SLS and LIML estimators are too complicated to provide any basis for comparing the two estimators. Hopefully, approximations to the distribution functions of these two estimators will provide more tractable expressions for comparison. By virtue of the large-sample asymptotic equivalence of the two estimators, the approximations must be better than that given by the limiting normal distribution.

For the case considered by Sawa and Richardson, this paper presents an approximation to the distribution function of the 2SLS estimator up to terms whose order of magnitude is $\frac{1}{\sqrt{N}}$, where N is the sample size.

In section 2, the 2SLS estimator is reduced to canonical form in terms of mutually independent bivariate normal random vectors and in section 3, Taylor series expansions are used to show that the 2SLS distribution function is equal to the sum of three terms, the first being the limiting standard normal distribution function, the second a correction term which is $O(\frac{1}{\sqrt{N}})$ and the third, a remainder term which is $O(\frac{1}{N})$ as $N \to \infty$. It is also shown in section 3 that in the approximation obtained, the correction term is $O(\frac{1}{\mu})$ and the remainder term $O(\frac{1}{\mu^2})$ as $\mu \to \infty$, where μ^2 is what has been referred to in the literature as the <u>concentration parameter</u>, (for example, see Richardson (11)). As a corollary to the methods developed in section 3, an approximation to the OLS distribution function may also be obtained for fixed N, up to terms whose order of magnitude is $O(\frac{1}{\mu})$ as $\mu \to \infty$. This approximation is given in section 4.

This paper makes use of certain results concerning the Wishart distribution. If $W = Z'Z$, where Z is a $n\times p$ random matrix $(n \geq p)$ whose rows are mutually independent normal random vectors with a common covariance matrix Σ, then W is said to have a Wishart distribution of order p, with n degrees of freedom, covariance matrix Σ and means sigma matrix $(EZ)'(EZ)$. The distribution of W is said to be central if $(EZ)'(EZ) = 0$.

Furthermore, if C is an arbitrary $p\times p$ nonsingular matrix and Q a symmetric idempotent $n\times n$ matrix of rank q, then CWC' has a Wishart distribution of order p, with n

degrees of freedom, covariance matrix $C\Sigma C'$ and means sigma matrix $C(EZ)'(EZ)C'$; and $Z'QZ$ has a Wishart distribution of order p, with q degrees of freedom, covariance matrix Σ and means sigma matrix $(EZ)'Q(EZ)$.

Finally, if the means sigma matrix $(EZ)'(EZ)$ is diagonal with main diagonal elements $\mu_1^2, \mu_2^2, \ldots, \mu_p^2$, say, then there exists a nxp random matrix S such that $W = S'S$ and

$$ES = \begin{pmatrix} 0 & 0 \\ 0 & D \end{pmatrix},$$

where D is a pxp diagonal matrix with $\mu_1, \mu_2, \ldots, \mu_p$ as its main diagonal elements. For a proof of this last statement, see Anderson and Girshick (2).

2. THE 2SLS ESTIMATOR IN CANONICAL FORM

In the case of two included endogenous variables, we may write the equation to be estimated as

$$\underset{\sim}{y}_1 = \underset{\sim}{y}_2\beta + Z_1\underset{\sim}{\gamma} + \underset{\sim}{u} \tag{2.1}$$

where $Y = (\underset{\sim}{y}_1\ \underset{\sim}{y}_2)$ is the Nx2 matrix of included endogenous variables, Z_1 is the NxK_1 matrix of included predetermined variables, $\underset{\sim}{u}$ is the Nx1 vector of residuals, and β and $\underset{\sim}{\gamma}$ are unknown parameters.

We assume that (2.1) is the first equation in a simultaneous system of $G(\geq 2)$ linear stochastic equations relating G endogenous and K predetermined variables. The $N \times K$ matrix of predetermined variables is denoted by Z, which is further partitioned as $(Z_1 \; Z_2)$, where Z_2 is the $N \times K_2$ matrix of predetermined variables excluded from (2.1).

The reduced form equations for the two endogenous variables included in (2.1) are

$$\begin{aligned} Y &= Z\Pi' + V \\ &= Z_1\Pi_1' + Z_2\Pi_2' + V \end{aligned} \tag{2.2}$$

where the $2 \times K$ matrix Π of reduced form coefficients is partitioned as $(\Pi_1 \; \Pi_2)$, Π_1 being $2 \times K_1$ and Π_2 being $2 \times K_2$. V is the $N \times 2$ matrix of reduced form residuals.

We make the following assumptions about the model:

(1) All predetermined variables are exogenous.

(2) The equation to be estimated is identified by zero-restrictions on the structural coefficients in the model.

(3) The sample size N is greater than or equal to $G + K$.

(4) The $N \times K$ matrix Z of exogenous variables is a matrix of constants and is of full rank.

(5) $\frac{Z'Z}{N}$ tends to a finite positive definite matrix as $N \to \infty$.

(6) The rows of V are mutually independent and identically dis-

tributed as bivariate normal random vectors with zero mean vector and positive definite covariance matrix Σ.

It is well known that assumption (2) is equivalent to the assumption that Π_2 is of rank 1, which in turn implies that $K_2 \geq 1$. Also, assumption (5) implies that in (2.1), $E\underset{\sim}{u} = \underset{\sim}{0}$.

By writing the normal equations for the 2SLS estimator $\hat{\beta}$ of β, it can be verified that

$$\hat{\beta} = \frac{\underset{\sim}{y}_2' \, P \, \underset{\sim}{y}_1}{\underset{\sim}{y}_2' \, P \, \underset{\sim}{y}_2} \tag{2.3}$$

where

$$P = Z(Z'Z)^{-1}Z' - Z_1(Z_1'Z_1)^{-1}Z_1' \, . \tag{2.4}$$

Note that $\hat{\beta}$ may also be written as the ratio between the (2,1)th and (2,2)th elements of the 2x2 matrix $Y'PY$.

Since by assumption (4) Z is of full rank, there exists a $K\times K$ non-singular upper triangular matrix Δ such that $\Delta'Z'Z\Delta = I$. Partition Δ as follows:

$$\Delta = \begin{pmatrix} \Delta_{11} & \Delta_{12} \\ 0 & \Delta_{22} \end{pmatrix}$$

where Δ_{11} is $K_1 \times K_1$ and Δ_{22} is $K_2 \times K_2$, and let

$$\begin{aligned} X &= Z\,\Delta \\ &= (Z_1\Delta_{11} \quad Z_1\Delta_{12} + Z_2\,\Delta_{22}) && (2.5) \\ &= (X_1 \qquad X_2\,). && (2.6) \end{aligned}$$

Note that $X'X = I$, Δ_{11} and Δ_{22} are upper triangular, non-singular and

$$P = X_2X_2' \;, \qquad (2.7)$$

$$X_2'Z_1 = 0 \;. \qquad (2.8)$$

It can also be verified by using (2.1), (2.7) and (2.8) that

$$M = (EY)'P\,(EY) = \tau^2 \begin{pmatrix} \beta^2 & \beta \\ \beta & 1 \end{pmatrix}, \qquad (2.9)$$

where

$$\begin{aligned} \tau^2 &= (E\,\underset{\sim}{y}_2)'\;P\;(E\,\underset{\sim}{y}_2) \\ &= \Pi_{22}Z_2'\left[\,I-Z_1(Z_1'Z_1)^{-1}Z_1'\,\right]Z_2\Pi_{22}' && (2.10) \end{aligned}$$

and Π_{22} is the second row of Π_2 in (2.2).

Now, by (2.2) and assumption (6), the N rows of Y are mutually independent bivariate normal random vectors with common covariance matrix Σ. Hence $Y'P\,Y$ is a Wishart matrix of order 2, with covariance matrix Σ, means sigma matrix M, and with $m = K_2$ degrees of freedom.

Let

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{21} \\ \sigma_{21} & \sigma_2^2 \end{bmatrix} , \qquad (2.11)$$

$$\omega^{-2} = \sigma_1^2 - 2\beta\sigma_{21} + \beta^2\sigma_2^2 , \qquad (2.12)$$

and

$$\psi = \begin{bmatrix} \omega & -\beta\omega \\ 0 & \frac{1}{\sigma_2} \end{bmatrix} . \qquad (2.13)$$

Then

$$\psi\Sigma\psi' = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \qquad (2.14)$$

and

$$\psi M\psi' = \begin{bmatrix} 0 & 0 \\ 0 & \frac{\tau^2}{\sigma_2^2} \end{bmatrix} \qquad (2.15)$$

where

$$\rho = \frac{\omega}{\sigma_2}(\sigma_{21} - \beta\sigma_2^2). \qquad (2.16)$$

Note that $|\rho| < 1$ since $|\psi\Sigma\psi'| = 1-\rho^2 > o$.

If we now let

$$A = \psi\ (Y'P\ Y)\ \psi' = (a_{ij}), \tag{2.17}$$

then A is a non-central Wishart matrix of order 2, with K_2 degrees of freedom and with covariance and means sigma matrices given by (2.14) and (2.15) respectively, so that A can also be expressed as (see Anderson and Girschik (2)):

$$A = \sum_{i=1}^{m} \begin{bmatrix} x_i^* \\ y_i^* \end{bmatrix} (\ x_i^*\ y_i^*\ ) \tag{2.18}$$

where the bivariate vectors $(x_i^*\ y_i^*)$ are mutually independent normal with common covariance matrix given by (2.14) and with means

$$E\ (\ x_i^*\ y_i^*\ ) = \begin{cases} (o \quad o),\ i=1,\ldots,m-1 \\ (o \quad \mu),\ \ i = m \end{cases} \tag{2.19}$$

where $m = K_2$ and

$$\mu = \frac{\tau}{\sigma_2} \tag{2.20}$$

Using (2.3), (2.17) and (2.18), we can express the 2SLS estimator of β in terms of A as

$$\hat{\hat{\beta}} = \beta + \frac{1}{\omega\sigma_2}\hat{\beta}^* , \tag{2.21}$$

where

$$\hat{\beta}^* = \frac{a_{21}}{a_{22}} = \frac{\sum_{i=1}^{m} x_i^* y_i^*}{\sum_{i=1}^{m} y_i^{*2}} . \tag{2.22}$$

3. APPROXIMATIONS TO THE 2SLS DISTRIBUTION FUNCTIONS

Let $\phi(x)$ and $\Phi(x)$ denote the standard normal density and distribution functions evaluated at x. Let the variables x_i's and y_i's be such that

$$x_i^* = \sqrt{1-\rho^2}\, x_i + \rho\, y_i , \quad i=1, 2, \ldots, m \tag{3.1}$$

$$y_i^* = \begin{cases} y_i & , \; i=1, 2, \ldots, m-1 \\ y_i + \mu & , \; i= m . \end{cases} \tag{3.2}$$

The x_i's and y_i's are jointly mutually independent standard normal variables and in terms of these variables,

$$\hat{\beta}^* = \frac{\sum_{i=1}^{m} y_i(\sqrt{1-\rho^2}\ x_i + \rho y_i) + \mu(\sqrt{1-\rho^2}\ x_m + \rho y_m)}{\sum_{i=1}^{m-1} y_i^2 + (y_m+\mu)^2} \qquad (3.3)$$

$$= \rho + \frac{\sqrt{1-\rho^2}\left[\sum_{i=1}^{m-1} x_i y_i + x_m(y_m+\mu)\right] - \rho\mu(y_m+\mu)}{\sum_{i=1}^{m-1} y_i^2 + (y_m+\mu)^2}. \qquad (3.4)$$

By (2.10), (2.20), and assumption (5), it follows that $\frac{\mu^2}{N}$ tends to a finite non-zero constant as $N \to \infty$. Note that by (2.21) and (3.3), the 2SLS distribution depends on N only through μ. Also, it follows from (3.3) that

$$\operatorname*{plim}_{N\to\infty} \left(\mu\hat{\beta}^* - \sqrt{1-\rho^2}\ x_m - \rho y_m \right) = 0.$$

Since $(\sqrt{1-\rho^2}\ x_m + \rho y_m)$ is a standard normal random variate, this implies that $\mu\hat{\beta}^*$ has a limiting standard normal distribution. Thus, $\mu\hat{\beta}^*$ is the normalized function of $\hat{\beta}$ whose distribution function we should try to approximate.

Given $y_1, y_2, \ldots, y_m$, $\dfrac{\sum_{i=1}^{m-1} x_i y_i + x_m(y_m+\mu)}{\sqrt{\sum_{i=1}^{m-1} y_i^2 + (y_m+\mu)^2}}$ is conditionally distributed as a standard normal random variable. It follows from (3.4) that

$$\hat{\beta}^* = \rho + \frac{\sqrt{1-\rho^2}\, z}{\sqrt{\sum_{i=1}^{m-1} y_i^2 + (y_m+\mu)^2}} - \frac{\rho\mu(y_m+\mu)}{\sum_{i=1}^{m-1} y_i^2 + (y_m+\mu)^2}, \qquad (3.5)$$

where z is a standard normal random variable independent of $y_1, y_2, \ldots, y_m$. Define the following for $m \geq 1$:

$$\underset{\sim}{y} = (y_1, y_2, \ldots, y_m), \qquad (3.6)$$

$$f(\underset{\sim}{y}) = \frac{1}{\sqrt{\sum_{i=1}^{m-1} \left(\frac{y_i}{\mu}\right)^2 + \left(1+\frac{y_m}{\mu}\right)^2}} \qquad (3.7)$$

$$h(\underset{\sim}{y};b) = \frac{f(\underset{\sim}{y})}{\sqrt{1-\rho^2}} \left\{ \frac{1}{\mu} \left(\frac{b}{\mu} - \rho\right) \left[\sum_{i=1}^{m-1} y_i^2 + (y_m+\mu)^2 \right] + \rho(y_m+\mu) \right\}, \qquad (3.8)$$

$$\hat{h}(\underset{\sim}{y};b) = \frac{1}{\sqrt{1-\rho^2}} \left\{ -\frac{\rho}{\mu} \sum_{i=1}^{m-1} y_i^2 + \left(\frac{b}{\mu} - \rho\right) y_m + b \right\}. \qquad (3.9)$$

Note that for the just-identified case $(K_2=1)$, $\hat{\beta}^* = \dfrac{x^*}{y^*}$, where $(x^* \; y^*)$ is normal with mean $(o \; \mu)$ and covariance matrix given by (2.14). Thus, the counterparts of (3.3) and (3.4) as expressions for $\hat{\beta}^*$ in the just-identified case are obtained by deleting all terms which involve $x_1, x_2, \ldots, x_{m-1}, y_1, y_2, \ldots, y_{m-1}$ and writing x, y in place of x_m, y_m. The corresponding expressions for (3.6)-(3.9) are obtained in the same way.

For b an arbitrary real number, let

$$\xi^{-\frac{1}{2}} = 1 - \frac{2b\rho}{\mu} + \frac{b^2}{\mu^2} . \tag{3.10}$$

Then the following equalities hold:

$$\Pr(\mu\hat{\beta}^* \leq b) \doteq \Pr(z \leq h(\underset{\sim}{y};b)) \tag{3.11}$$

$$= \Pr(z \leq \hat{h}(\underset{\sim}{y};b)) + O(\frac{1}{\mu^2}) \text{ as } \mu \to \infty \tag{3.12}$$

$$= \Pr(z' \leq b\xi - \frac{\rho\xi}{\mu} \sum_{i=1}^{m-1} y_i^2) + O(\frac{1}{\mu^2}) \text{ as } \mu \to \infty \tag{3.13}$$

$$= E\,\Phi(b\xi - \frac{\rho\xi}{\mu} \sum_{i=1}^{m-1} y_i^2) + O(\frac{1}{\mu^2}) \text{ as } \mu \to \infty \tag{3.14}$$

$$= \Phi(b) + \frac{\rho}{\mu}\,\phi(b)\,(b^2-m+1) + O(\frac{1}{\mu^2}) \text{ as } \mu \to \infty . \tag{3.15}$$

(3.11) follows immediately from (3.5) and (3.8). For (3.12), the function $\hat{h}(\underset{\sim}{y};b)$ may be obtained from $h(\underset{\sim}{y};b)$ in succeeding steps. First delete the term with factor $\frac{1}{\mu^2}$ from the factor of $h(\underset{\sim}{y};b)$ in braces in (3.8). Then approximate $f(\underset{\sim}{y})$ by using Taylor's theorem, which is applicable if we restrict our approximation to the region $\{\underset{\sim}{y}: |y_m| \leq (1-\varepsilon)\mu\}$ for some fixed ε such that $0 < \varepsilon < 1$. By multipying the expressions obtained in the first and second steps and again deleting terms with factor $\frac{1}{\mu^2}$, we finally get $\hat{h}(\underset{\sim}{y};b)$. In (3.13), z' is a standard normal variable independent of $y_1, y_2, \ldots, y_{m-1}$. (3.13) follows from (3.12) by a straightforward manipulation of the inequality $z \leq \hat{h}(\underset{\sim}{y};b)$. (3.14) is an immediate consequence of (3.13) and (3.15) is obtained by using Taylor's theorem to expand the leading term in (3.14) about b as well as to expand the expression for ξ about unity. Detailed proofs of (3.12) and (3.15) are given in the appendix.

The following theorem now holds by virtue of (2.21) and (3.15):

<u>Theorem 3.1</u>. <u>In the case of two included endogenous variables in the equation being estimated, an approximation to the 2SLS distribution function is given by</u>

$$\Pr\left\{(\hat{\beta} - \beta) \leq \frac{b}{\mu\omega\sigma_2}\right\} = \Phi(b) + \frac{\rho}{\mu}\,\phi(b)(b^2-K_2+1) + O\left(\frac{1}{\mu^2}\right) \text{ as } \mu \to \infty$$

where b is an arbitrary real number,

$$\rho = \frac{\omega}{\sigma_2} (\sigma_{21} - \beta\sigma_2^2),$$

$$\omega^{-2} = \sigma_1^2 - 2\beta\sigma_{21} + \beta^2\sigma_2^2 ,$$

and

$$\mu^2 = \frac{1}{\sigma_2^2} \left\{ \Pi_{22}' Z_2' \left[I - Z_1(Z_1'Z_1)^{-1}Z_1' \right] Z_2 \Pi_{22}' \right\}$$

$$= O(N).$$

It can be seen from the proof given in the appendix that the remainder term in the above approximation depends on the sample size N only through μ. Since $\mu^2 = O(N)$, it follows then that the remainder term is also $O(\frac{1}{N})$ and the second term in the approximation, which we may call the correction term, is $O(\frac{1}{\sqrt{N}})$. Thus, Theorem 3.1 gives both a large N and a large μ asymptotic approximation to the 2SLS distribution.

For the just-identified case $(K_2 = 1)$, we can improve on Theorem 3.1 to be more precise concerning the remainder term in the approximation to the 2SLS distribution. We make use of the fact indicated previously, that for $K_2 = 1$, β^* is the ratio of two correlated normal random variables, with the mean of the numerator being equal to zero and that of the denominator being equal to μ.

If $K_2 = 1$, we get from (3.4)

$$\hat{\beta}^* = \rho + \frac{\sqrt{1-\rho^2}\ x - \rho\mu}{y+\mu}. \tag{3.16}$$

where x and y are independent standard normal random variables. Hence

$$\Pr(\mu\hat{\beta}^* \leq b) = \Pr\left\{\frac{\sqrt{1-\rho^2}\ x - \rho\mu}{y+\mu} \leq \frac{b}{\mu} - \rho\right\}$$

$$= \Pr\left\{\sqrt{1-\rho^2}\ x - \left(\frac{b}{\mu} - \rho\right)y \leq b \text{ and } y+\mu > 0\right\}$$

$$+ \Pr\left\{\sqrt{1-\rho^2}\ x - \left(\frac{b}{\mu} - \rho\right)y \geq b \text{ and } y+\mu < 0\right\}$$

$$= \Pr\left\{\sqrt{1-\rho^2}\ x - \left(\frac{b}{\mu} - \rho\right)y \leq b\right\} + R, \tag{3.17}$$

where

$$R = \Pr\left\{\sqrt{1-\rho^2}\ x-(\tfrac{b}{\mu} - \rho)y \geq b \quad \text{and} \quad y+\mu < 0\right\}$$

$$-\Pr\left\{\sqrt{1-\rho^2}\ x-(\tfrac{b}{\mu} - \rho)y \leq b \quad \text{and} \quad y+\mu < 0\right\}.$$

Since both terms in the above expression for R are less than or equal to

$$\Pr(y+\mu < 0) = \Phi(-\mu) \leq \frac{\phi(\mu)}{\mu} = \frac{1}{\sqrt{2\pi}\ \mu e^{\frac{1}{2}\mu^2}},$$

it follows that

$$|R| \leq \frac{1}{\sqrt{2\pi}\ \mu e^{\frac{1}{2}\mu^2}}, \tag{3.18}$$

Since $\Pr\left\{\sqrt{1-\rho^2}\, x - (\frac{b}{\mu} - \rho)y \le b\right\} = \Phi(b\xi)$, the following theorem then follows from (3.17) and (3.18).

<u>Theorem 3.2. If the equation being estimated contains two endogenous variables, and is just-identified, then</u>

$$\Pr\left\{(\hat{\beta} - \beta) \le \frac{b}{\mu\omega\sigma_2}\right\} = \Phi(b\xi) + R,$$

<u>where</u> μ, ω <u>and</u> ξ <u>are as given in Theorem 3.1, and</u>

$$|R| \le \frac{1}{\sqrt{2\pi}\;\mu e^{\frac{1}{2}\mu^2}} .$$

4. <u>APPROXIMATION TO THE OLS DISTRIBUTION FUNCTION</u>

Within the same framework used in this paper, it can be shown that the OLS estimator $\tilde{\beta}$ of β may be expressed as the ratio between the (2,1)th and (2,2)th elements of $Y'[I-Z_1(Z_1'Z_1)^{-1}Z_1']Y$ a 2×2 Wishart matrix which has the same covariance and means sigma matrices, namely Σ and M, as $Y'PY$, but has $N-K_1$ degrees of freedom instead of K_2.

Thus, except for the difference in the degrees of freedom, the OLS estimator of β has exactly the same reduction to canonical form as the 2SLS estimator. More specifically, (2.21) and (2.22) also hold for the OLS estimator with $m = N-K_1$. Also, if we assume that the sample size N is fixed, we may use the procedure in section 3 to obtain a large μ asymptotic approximation to the OLS distribution function. The result is as given in Theorem 3.1 with K_2 replaced by $N-K_1$. That is,

Theorem 4.1. In the case of two endogenous variables present in the equation to be estimated, let $\tilde{\beta}$ be the OLS estimator of β and let the sample size N be fixed. Then

$$\Pr\left\{ (\tilde{\beta} - \beta) \leq \frac{b}{\mu\omega\sigma_2} \right\} = \Phi(b) + \frac{\rho}{\mu}\phi(b)\,(b^2 - N + K_1 + 1) + O\left(\frac{1}{\mu^2}\right) \text{ as } \mu \to \infty,$$

where b is an arbitrary real number and ρ, μ and ω are as given in Theorem 3.1.

APPENDIX

For a fixed ε such that $0 < \varepsilon < 1$, let

$$Q_\varepsilon = \left\{ \underset{\sim}{y}: \ |y_m| \le (1-\varepsilon)\mu \right\}.$$

To prove (3.12) and (3.15), we need the following lemma:

<u>Lemma A.1.</u> <u>For</u> $\mu > 0$ <u>and</u> $\underset{\sim}{y} \in Q_\varepsilon$,

$$\left| f(\underset{\sim}{y}) - (1 - \frac{y_m}{\mu}) \right| \le \frac{1}{2\mu^2} \sum_{i=1}^{m} y_i^2 + \frac{3}{8\varepsilon^5} \left(2\frac{y_m}{\mu} + \frac{1}{\mu^2} \sum_{i=1}^{m} y_i^2 \right)^2.$$

<u>Proof.</u> For $\mu > 0$ and $\underset{\sim}{y} \in Q_\varepsilon$,

$$1 + \frac{2y_m}{\mu} + \frac{1}{\mu^2} \sum_{i=1}^{m} y_i^2 \ge \varepsilon^2 > 0 \tag{A.1}$$

and hence, we can apply Taylor's theorem to expand $f(\underset{\sim}{y})$ around zero where the argument in the expansion is $\frac{2y_m}{\mu} + \frac{1}{\mu^2} \sum_{i=1}^{m} y_i^2$:

$$f(\underset{\sim}{y}) = \frac{1}{\sqrt{1 + \frac{2y_m}{\mu} + \frac{1}{\mu^2} \sum_{i=1}^{m} y_i^2}}$$

$$= 1 - \frac{y_m}{\mu} - \frac{1}{2\mu^2} \sum_{i=1}^{m} y_i^2 + \frac{3}{8} \left(1 + y^*\right)^{-\frac{5}{2}} \left(\frac{2y_m}{\mu} + \frac{1}{\mu^2} \sum_{i=1}^{m} y_i^2\right)^2, \tag{A.2}$$

for some y^* between 0 and $\frac{2y_m}{\mu} + \frac{1}{\mu^2} \sum_{i=1}^{m} y_i^2$. By (A.1), $\left(1+y^*\right)^{-\frac{5}{2}} \leq \varepsilon^{-5}$ and thus the lemma follows from (A.2). Q.E.D.

Proof of (3.12). It follows from the above lemma and the expressions for $h(\underset{\sim}{y};b)$ and $\hat{h}(\underset{\sim}{y};b)$ as given by (3.8) and (3.9) that for $\underset{\sim}{y} \in Q_\varepsilon$ and $\mu>0$,

$$\left| h(\underset{\sim}{y};b) - \hat{h}(\underset{\sim}{y};b) \right| \leq p(\underset{\sim}{y};\mu) \tag{A.3}$$

where $p(\underset{\sim}{y};\mu)$, an expression not depending on N except through μ, is a non-negative polynomial of finite degree in $|y_m|$ and $\sum_{i=1}^{m} y_i^2$ such that

$$E\, p(\underset{\sim}{y};\mu) = O(\frac{1}{\mu^2}) \text{ as } \mu \to \infty. \tag{A.4}$$

By the Mean-Value Theorem,

$$\left| \Phi\left[h(\underset{\sim}{y};b)\right] - \Phi\left[\hat{h}(\underset{\sim}{y};b)\right] \right| \leq \phi(o) \left| h(\underset{\sim}{y};b) - \hat{h}(\underset{\sim}{y};b) \right|, \tag{A.5}$$

which implies by (A.3) that

$$\int_{Q_\varepsilon} \left|\Phi[h(\underset{\sim}{y};b)] - \Phi[\hat{h}(\underset{\sim}{y};b)]\right| \prod_{i=1}^{m} d\Phi(y_i)$$

$$\leq \phi(0) \int_{Q_\varepsilon} p(y;\mu) \prod_{i=1}^{m} d\Phi(y_i) \tag{A.6}$$

$$\leq \phi(0)\ E\ p(y;\mu)$$

$$= O\left(\frac{1}{\mu^2}\right) \quad \text{as} \quad \mu \to \infty. \tag{A.7}$$

Furthermore, for Q_ε^c the complement of Q_ε in m-space,

$$\int_{Q_\varepsilon^c} |\Phi[h(\underset{\sim}{y};b)] - \Phi[\hat{h}(\underset{\sim}{y};b)]| \prod_{i=1}^{m} d\Phi(y_i)$$

$$\leq P(Q_\varepsilon^c) = 2\Phi[-(1-\varepsilon)\mu]$$

$$\leq \frac{2\phi[(1-\varepsilon)\mu]}{(1-\varepsilon)\mu} = \frac{1}{(1-\varepsilon)\sqrt{2\pi}\ \mu\varepsilon^{\frac{(1-\varepsilon)^2}{2}\mu^2}}. \tag{A.8}$$

Therefore,

$$|\Pr\{z \leq h(\underset{\sim}{y};b)\} - \Pr\{z \leq \hat{h}(\underset{\sim}{y};b)\}|$$

$$= \left| E\left\{\Phi[h(\underset{\sim}{y};b)] - \Phi[\hat{h}(\underset{\sim}{y};b)]\right\}\right|$$

$$\leq E\left|\Phi[h(\underset{\sim}{y};b)] - \Phi[\hat{h}(\underset{\sim}{y};b)]\right|$$

$$\leq O\left(\frac{1}{\mu^2}\right) \quad \text{as} \quad \mu \to \infty,$$

by (A.7) and (A.8). <u>Q.E.D.</u>

Proof of (3.15). By applying Taylor's theorem to expand $\phi[b\xi - \frac{\rho\xi}{\mu}\sum_{i=1}^{m-1} y_i^2]$ about b, we get

$$\phi\left[b\xi - \frac{\rho\xi}{\mu}\sum_{i=1}^{m-1} y_i^2\right] = \phi(b) + \left[b(\xi-1) - \frac{\rho\xi}{\mu}\sum_{i=1}^{m-1} y_i^2\right]\phi(b)$$

$$+ \frac{\phi'(b^*)}{2}\left[b(\xi-1) - \frac{\rho\xi}{\mu}\sum_{i=1}^{m-1} y_i^2\right]^2 , \qquad \text{(A.9)}$$

where b^* is some value between b and $b\xi - \frac{\rho\xi}{\mu}\sum_{i=1}^{m-1} y_i^2$.

Now, take a constant ∂ such that $0 < \partial < 1$. Since $|\rho|<1$, it follows that

$$1 - \frac{2b\rho}{\mu} + \frac{b^2}{\mu^2} > \partial(1-\rho^2) + (\rho - \frac{b}{\mu})^2 > 0.$$

Thus, for all values of b and ρ, Taylor's theorem may be used to get

$$\xi = 1 + \frac{b\rho}{\mu} - \frac{b^2}{2\mu^2} + \frac{3}{8}\left(\frac{2b\rho}{\mu} - \frac{b^2}{\mu^2}\right)^2 (1-b^{**})^{-\frac{5}{2}} , \qquad \text{(A.10)}$$

where b^{**} is some value between 0 and $\frac{2b\rho}{\mu} - \frac{b^2}{\mu^2}$. Hence

$$\xi = 1 + \frac{b\rho}{\mu} + O\left(\frac{1}{\mu^2}\right) \quad \text{as} \quad \mu \to \infty . \tag{A.11}$$

Since $\max |\phi'(x)| = \phi(1)$, it follows that the contribution of the third term in (A.9) to $\Pr\left\{z \leq \hat{h}(\underset{\sim}{y};b)\right\}$ is $O(\frac{1}{\mu^2})$ as $\mu \to \infty$. Thus, (3.15) follows from (3.14), (A.9) and (A.10). Q.E.D.

REFERENCES

(1) 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.

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

(3) 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.

(4) Basmann, R.L. (1957). A generalized classical method of linear estimation of coefficients in a structural equation. Econometrica, 25, 77-83.

(5) Basmann, R.L. (1961). A note on the exact finite sample frequency functions of generalized classical linear estimators in two leading over-identified cases, J. Amer. Stat. Assoc., 56, 619-636.

(6) Basmann, R.L. (1963). A note on the exact finite sample frequency functions of generalized classical linear estimators in a leading three equation case, J. Amer. Stat. Assoc., 58, 161-171.

(7) 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.

(8) Kabe, D.G. (1963). A note on the exact distributions of the GCL estimators in two leading over-identified cases, J. Amer. Stat. Assoc., 58, 535-537.

(9) Kabe, D.G. (1964). A note on the exact distribution of the GCL estimators in a leading three equation case, J. Amer. Stat. Assoc., 59, 881-894.

REFERENCES (Cont.)

(10) Mariano, R. and T. Sawa (1970). Exact finite-sample distribution of the limited-information maximum likelihood estimator in the case of two included endogenous variables. Unpublished paper.

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

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