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

Discussion Paper No. 70-22 December 1, 1970

APPROXIMATIONS TO THE DISTRIBUTION FUNCTIONS OF THEIL'S K - CLASS ESTIMATORS IN THE CASE OF TWO INCLUDED ENDOGENOUS VARIABLES

By

ROBERTO S. MARIANO

APPROXIMATIONS TO THE DISTRIBUTION FUNCTIONS OF THEIL'S K - CLASS ESTIMATORS IN THE CASE OF TWO INCLUDED ENDOGENOUS VARIABLES

ROBERTO S. MARIANO*

1. SUMMARY

Assuming that the sample size N is fixed, we present in this paper large μ asymptotic approximations to the distribution functions of the k-class estimators. The result, which holds for k non-stochastic and the case of two included endogenous variables, is obtained by slightly modifying the reduction to canonical form for the k-class estimator as given in Mariano (2) and then applying the same technique used in arriving at approximations to the two-stage least squares and ordinary least squares distributions functions. For example, such technique was used in Mariano (1) or Mariano (3). Since the 2SLS and OLS estimators belong to the set of k-class estimators (k non-stochastic), the result given here is a generalization of the approximations to the 2SLS and OLS distribution functions.

2. THE K-CLASS ESTIMATORS

The underlying model is a simultaneous system of G linear stochastic equations relating G endogenous and K predetermined variables. The equation being estimated may be written as

---

*The author would like to acknowledge the support of the Rockefeller Foundation for the research presented in this paper.

$$\underset{N\times 1}{\underset{\sim}{y}} = \underset{N\times G_1}{Y_1} \underset{G_1\times 1}{\underset{\sim}{\beta}} + \underset{N\times K_1}{Z_1} \underset{N\times K_1}{\underset{\sim}{\gamma_1}} + \underset{N\times 1}{\underset{\sim}{u}} \tag{2.1}$$

and the reduced form equations for the G_1+1 endogenous variables included in (2.1) may be written as

$$\underset{Nx(G_1+1)}{Y} = \underset{NxK}{Z} \underset{Kx(G_1+1)}{\Pi'} + \underset{Nx(G_1+1)}{V} . \tag{2.2}$$

In the above equations, the symbols are defined as follows:

$Y = (\underset{\sim}{y} \;\; Y_1)$ is the $Nx(G_1+1)$ matrix of included endogenous variables,

Z_1 the NxK_1 matrix of included predetermined variables,

Z_2 the NxK_2 matrix of excluded predetermined variables,

$Z = (Z_1 \;\; Z_2)$ the NxK matrix of pretermined variables,

V the $Nx(G_1+1)$ matrix of reduced form disturbance terms,

Π the $(G_1+1)xK$ matrix of reduced form coefficients,

later to be partitioned as $(\underset{K_1}{\Pi_1} \;\; \underset{K_2}{\Pi_2})$.

In this model, we assume that all predetermined variables are exogeneous; the equation being estimated is identified by zero restrictions on the structural coefficients; the sample size is greater than or equal to $G + K$; Z is a matrix of constants and is of full rank; and finally, the rows of V are mutually independent and identically distributed as normal random vectors with zero mean vector and positive definite $(G_1+1) \times (G_1+1)$ covariance matrix Σ.

In this set-up, the k-class estimator of β may be expressed as

$$\hat{\beta}_{(k)} = C_{22}^{-1} \, c_{21} \, , \tag{2.3}$$

where

$$C = kW + (1-k)A, \tag{2.4}$$

$$A = Y' \left[I - Z_1 (Z_1'Z_1)^{-1} Z_1'\right] Y \sim W_{G_1+1} (N-K_1, \Sigma; M), \tag{2.5}$$

$$W = Y' \left[Z(Z'Z)^{-1}Z' - Z_1(Z_1'Z_1)^{-1}Z_1'\right] Y \sim W_{G_1+1}(K_2, \Sigma; M), \tag{2.6}$$

$$M = \Pi_2 Z_2' \, P_1 Z_2 \Pi_2'. \tag{2.7}$$

It can be easily shown that for $S = A-W$, S is independent of W,

$$S \sim W_{G_1+1} \; (N-K, \Sigma; 0), \tag{2.8}$$

and in terms of S and W,

$$C = W + (1-k) \, S. \tag{2.9}$$

From the reduction given in Mariano (2), it follows then that

$$\hat{\beta}_{(k)} = \beta + \omega\theta' \, C^*{}_{22}^{-1} \, c^*{}_{21} \, , \tag{2.10}$$

where

$$C^* = W^* + (1-k)\, S^*, \tag{2.11}$$

$$W^* \sim W_{G_1+1}(K_2,\ \Sigma^*;\ M^*), \tag{2.12}$$

$$S^* \sim W_{G_1+1}(N-K,\ \Sigma^*;\ \underset{\sim}{0}), \tag{2.13}$$

$$\Sigma^* = \begin{pmatrix} 1 & \underset{\sim}{\rho}' \\ \underset{\sim}{\rho} & I \end{pmatrix} \begin{matrix} 1 \\ G_1 \end{matrix}, \quad \text{columns: } 1 \quad G_1 \tag{2.14}$$

$$M^* = \begin{pmatrix} 0 & \underset{\sim}{0}' \\ \underset{\sim}{0} & D \end{pmatrix} \begin{matrix} 1 \\ G_1 \end{matrix}, \quad \text{columns: } 1 \quad G_1 \tag{2.15}$$

$$\omega^2 = \sigma_{21} - 2\underset{\sim}{\beta}'\Sigma_{21} + \underset{\sim}{\beta}'\Sigma_{22}\beta\ , \tag{2.16}$$

and Θ is a $G_1 x G_1$ non-singular matrix such that

$$\Theta\, \Sigma_{22}\, \Theta' = I \tag{2.17}$$

$$\Theta\, M_{22}\, \Theta' = D \tag{2.18}$$

where in (2.15) and (2.18), D is a $G_1 x G_1$ diagonal matrix whose main-diagonal elements are the characteristics roots of $\Sigma_{22}^{-1} M_{22}$ arranged in increasing order. Also, in (2.14),

$$\underset{\sim}{\rho} = \frac{\Theta}{\omega}\, (\Sigma_{21} - \Sigma_{22}\, \underset{\sim}{\beta}). \tag{2.19}$$

3. THE CASE OF TWO INCLUDED ENDOGENOUS VARIABLES

In this case, there is only one coefficient being estimated and this is denoted by β. The k-class estimator in canonical forms reduces to

$$\hat{\beta}_{(k)} = \beta + \frac{\omega}{\sqrt{\sigma_{22}}} \hat{\beta}^*_{(k)} , \tag{3.1}$$

where

$$\hat{\beta}^*_{(k)} = \frac{c^*_{21}}{c^*_{22}} . \tag{3.2}$$

By virtue of (2.11), (2.12), and (2.13), we can further write

$$\hat{\beta}^*_{(k)} = \frac{\sum_{i=1}^{m} x_i^* y_i^* + (1-k) \sum_{j=1}^{r} u_j^* v_j^*}{\sum_{i=1}^{m} y_i^{*2} + (1-k) \sum_{j=1}^{r} v_j^{*2}} , \tag{3.3}$$

where

$$m = K_2 , \qquad r = N-K, \tag{3.4}$$

and the $(x_i^* \ y_i^*)$'s and $(u_j^* \ v_j^*)$'s are mutually independent bivariate normal with common covariance matrix

$$\Sigma^* = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \tag{3.5}$$

and with zero mean vectors except for $(x_m^* \; y_m^*)$ which has mean $(o \; \mu)$. Here,

$$\rho = \frac{\sigma_{21} - \beta \sigma_{22}}{\omega \sqrt{\sigma_{22}}} \tag{3.6}$$

and

$$\mu^2 = \frac{1}{\sigma_{22}} \left\{ \Pi_{22} Z_2' \left[I - Z_1 (Z_1' Z_1)^{-1} Z_1' \right] Z_2 \Pi_{22}' \right\} \tag{3.7}$$

where $\underset{\sim}{\Pi}_{22}$ is the second row of Π_2. μ^2 has been referred to in the literature as the concentration parameter.

Let the variables x_i's, y_i's, u_j's, and v_j's be such that

$$x_i^* = \sqrt{1-\rho^2}\; x_i + \rho y_i \quad , \quad i = 1, 2, \ldots, m \tag{3.8a}$$

$$y_i^* = \begin{cases} y_i & , \quad i = 1, 2, \ldots, m-1 \\ y_i + \mu & , \quad i = m \end{cases} \tag{3.8b}$$

$$u_j^* = \sqrt{1-\rho^2}\; u_j + \rho v_j \quad , \quad j = 1, 2, \ldots, r \tag{3.8c}$$

$$v_j^* = v_j \quad , \quad j = 1, 2, \ldots, r. \tag{3.8d}$$

Then the x_i's, y_i's, u_j's, and v_j's are mutually independent standard normal variables and in terms of these variables,

$$\hat{\beta}^*_{(k)} = \frac{F^2}{\mu^2}\left\{\sum_{i=i}^{m} y_i\,(\sqrt{1-\rho^2}\,x_i+\rho y_i) + \mu(\sqrt{1-\rho^2}\,x_m + \rho y_m)\right.$$

$$\left. + (1-k)\sum_{j=1}^{r} v_j(\sqrt{1-\rho^2}\,u_j + \rho v_j)\right\} \qquad (3.9)$$

$$= \rho + \frac{F^2}{\mu^2}\left\{\sqrt{1-\rho^2}\left[\sum_{i=1}^{m} x_i y_i + \mu x_m + (1-k)\sum_{j=1}^{r} u_j v_j\right] - \rho\mu(y_m+\mu)\right\} \qquad (3.10)$$

$$= \rho + \frac{\sqrt{1-\rho^2}\,z\,F}{\mu} - \frac{\rho\,(y_m + \mu)F^2}{\mu}\,, \qquad (3.11)$$

where

$$F^2 = \frac{\mu^2}{\sum_{i=1}^{m=1} y_i^2 \quad (y_m+\mu)^2 + (1-k)\sum_{j=1}^{r} v_j^{\,2}}\,, \qquad (3.12)$$

and z is a standard normal random variable independent of y_1, y_2, ..., y_m, v_1, v_2, ..., v_r.

From (3.9) and (3.12), it follows that as N is kept fixed and $\mu \to \infty$, $\text{plim}\ (\mu\,\hat{\beta}^*_{(k)} - \sqrt{1-\rho^2}\,x_m - \rho y_m) = 0$ so that the following theorem holds:

<u>Theorem 3.1.</u> <u>For fixed</u> N, <u>the limiting distribution of</u> $\frac{\mu\sqrt{\sigma_{22}}}{\omega}(\hat{\beta}_{(k)} - \beta)$ <u>is a standard normal distribution as</u> $\mu \to \infty$.

4. APPROXIMATIONS

Under the assumption that the sample size N is fixed, we present in this section a large μ asymptotic approximation to the distribution function of the k-class estimator. The procedure we apply is exactly the same as that used in Mariano (3) to approximate the distribution for of the two-stage least squares and ordinary least squares estimators. For this reason, we shall simply outline the procedure and refer the reader to Mariano (3) for a more detailed justification of each step in the procedure.

For b an arbitrary real number, let

$$H = \frac{F}{\sqrt{1-\rho^2}}\left\{\frac{1}{\mu F^2}\left(\frac{b}{\mu} - \rho\right) + \rho(y_m+\mu)\right\}, \tag{4.1}$$

$$\hat{H} = \frac{1}{\sqrt{1-\rho^2}}\left\{-\frac{\rho}{\mu}\left[\sum_{i=1}^{m-1} y^2 + (1-k)\sum_{j=1}^{r} v_j^2\right] + \left(\frac{b}{\mu} - \rho\right)y_m + b\right\} \tag{4.2}$$

$$\xi^{-2} = 1 - \frac{2b\rho}{\mu} + \frac{b^2}{\mu^2} . \tag{4.3}$$

Then, the following equalities hold (assuming that N is fixed):

$$\Pr(\mu\hat{\beta}^*_{(k)} \leq b) = \Pr(z \leq H) \tag{4.4}$$

$$= \Pr(z \leq \hat{H}) + 0\left(\frac{1}{\mu^2}\right) \text{ as } \mu \to \infty \tag{4.5}$$

$$= \Pr\left\{z' \leq b\xi - \frac{\rho\xi}{\mu}\left[\sum_{i=1}^{m-1} y_i^2 + (1-k)\sum_{j=1}^{r} v_j^2\right]\right\} + 0\left(\frac{1}{\mu^2}\right) \text{ as } \mu \to \infty \tag{4.6}$$

$$= E\,\Phi\left[b\xi - \frac{\rho\xi}{\mu}\left(\sum_{i=1}^{m-1} y_i^2 + (1-k)\sum_{j=1}^{r} v_j^2\right)\right]$$

$$+ 0\left(\frac{1}{\mu^2}\right) \quad \text{as} \quad \mu \to \infty \tag{4.7}$$

$$= \Phi(b) + \frac{\rho}{\mu}\phi(b)\left[b^2 - m + 1 + (1-k)r\right] + 0\left(\frac{1}{\mu^2}\right) \text{ as } \mu \to \infty\,. \tag{4.8}$$

In the above, $\phi(\cdot)$ and $\Phi(\cdot)$ denote the standard normal density and distribution functions.

(4.4) is an immediate consequence of (3.11), (3.12), and (4.1). In (4.6), z' is a standard normal random variable independent of $y_1, y_2, \ldots, y_m, v_1, v_2, \ldots, v$. (4.6) is obtained from (4.5) by a straightforward manipulation of the inequality $z \leq H$. (4.7) follows immediately from (4.6). For detailed proofs of (4.5) and (4.8), see the appendix of Mariano (3).

The following now holds by virtue of (3.1), (3.4), and (4.8):

Theorem 4.1. In the case of two included endogenous variables in the equation being estimated, let the sample size N be fixed. Then an approximation to the distribution function of the k-class estimator is given by

$$\Pr\left\{(\hat{\beta}_{(k)}-\beta) \leq \frac{b\omega}{\mu\sqrt{\sigma_{22}}}\right\} = \Phi(b) + \frac{\rho}{\mu}\phi(b)\left[b^2-K_2+1+(1-k)(N-K)\right] + 0\left(\frac{1}{\mu^2}\right) \quad \text{as} \quad \mu \to \infty ,$$

__where__ b __is an arbitrary real number__,

$$\rho = \frac{1}{\omega\sqrt{\sigma_{22}}}(\sigma_{21} - \beta\,\sigma_{22}) ,$$

$$\omega^2 = \sigma_{11} - 2\beta\sigma_{21} + \beta^2\sigma_{22} ,$$

__and__

$$\mu^2 = \frac{1}{\sigma_{22}}\left\{\underset{\sim}{\Pi}_{22}Z_2'\left[1-Z_1(Z_1'Z_1)^{-1}Z'\right]Z_2\underset{\sim}{\Pi}_{22}'\right\} .$$

REFERENCES

(1) Mariano, R. (1969). On distributions and moments of single-equation estimators in a set of simultaneous linear stochastic equations. Unpublished doctoral dissertation, Department of Statistics and Technical Report No. 27, Institute for Mathematical Studies in the Social Sciences, Stanford University.

(2) Mariano, R. (1970). On the reduction to canonical form and the existence of moments of the ordinary least squares and two -stage least squares estimators. Unpublished Discussion Paper No. 70-7, School of Economics, University of the Philippines.

(3) Mariano, R. (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. Unpublished Discussion Paper No. 70-10, School of Economics, University of the Philippines.