数理解析研究所講究録
第 507 巻 1983 年 148-172

PROPERTIES OF SIMULTANEOUS-EQUATION ESTIMATORS IN THE ECONOMETRIC MODEL

by

Kimio Morimune

2. The Model and Symbols

A single structural equation such as (1.2) or (1.3) is denoted as

$$y_1 = y_2\beta + Z_1\gamma + u \tag{2.1}$$

where y_1 and y_2 are T-component (column) vectors of T observations on two endogenous variables, Z_1 is a $T \times K_1$ matrix of observations on K_1 exogenous variables, β is a scalar parameter, γ_1 is a K_1-component vector of parameters, and u is a T-component vector of (unobservable) disturbances. The whole system consists of many structural equations such as (1.1). The whole system may be written as $YB+Z\Gamma=U$ where B and Γ are coefficient matrices of endogenous and exogenous variables, respectively. Post-multiplying the system of equations by B^{-1}, it follows the reduced form equations of y_1 and y_2 which may be denoted as

$$Y \equiv (y_1\ y_2) = Z\Pi + V \tag{2.2}$$

where Z is a T×K matrix of exogenous variables of rank K, Π is K×2 and consists of two columns of $-B\Gamma^{-1}$, and $V=(v_1\ v_2)$ is a T×2 matrix of (unobservable) disturbances consisting of two columns of UB^{-1}. The rows of V are independently and normally distributed, each row having mean 0 and (nonsingular) covariance matrix $\Omega=(\omega_{ij})$, i,j=1,2. (The reduced form equation represents the stochastic structure of the endogenous variables: they are determined by all exogenous variables which may be taken as input into the system, reduced form coefficients, and error term V. However, the reduced form equation does not include any economic informations. On the contrary, specification of each structural equation is determined by a priori or economic informations.) To relate the structural equation (2.1) to the reduced form equation we partition Π into K_1 and $K_2(=K-K_1)$ rows and into two columns: $\Pi = (\pi_{ij})$, i,j=1,2. If we post-multiply the reduced form equation by $(1, -\beta)'$, the structural equation follows, where $\gamma_1 = \pi_{11} - \pi_{12}\beta$ and $u = v_1 - v_2\beta$ In order that the structural equation be properly written with Z_2 omitted, it must satisfy that

$$\pi_{21} = \pi_{22}\beta \tag{2.3}$$

which brings on restrictions on the likelihood function and termed <u>over-identifiability condition.</u> In order that (2.3) be appropriate as K_2 linear equations the matrix $(\pi_{21} \;\; \pi_{22})$ must be of rank one, and π_{22} must have at least one non-zero component. The components of u are assumed to be independently and normally distributed with means 0 and variances σ^2, which is defined to be $\omega_{11} - 2\beta\omega_{12} + \beta^2\omega_{22}$.

Symbols used throughout this paper are the number of restrictions on β included in (2.3) termed <u>degree of over-identifiability:</u>

$$L = K_2 - 1, \tag{2.4}$$

<u>the noncentrality parameter:</u>

$$\delta^2 = \frac{1}{\omega_{22}} \pi_{22}' Z'(I - Z_1(Z_1' Z_1)^{-1} Z_1')Z \pi_{22} , \tag{2.5}$$

defining $\alpha = (\omega_{22}\beta - \omega_{12})/\sqrt{|\Omega|}$ <u>the coefficient of simultaneity:</u>

$$\rho \equiv \mathrm{cor}(y_{2t}, u_t) = -\frac{\alpha}{\sqrt{1+\alpha^2}}, \tag{2.6}$$

and the total degree of freedom in estimation: T-K. (ρ and α are interchangeably used because they are uniquely related. The term of simultaneity arises from the fact that the correlation between y_2 and u in (2.1) is essential for the simultaneous equation method. The symbol μ^2 is defined to be $(1+\alpha^2)\delta^2$, which is also termed noncentrality parameter in some papers. See Morimune (1983).) It is assumed throughout that

$$\lim_{T\to\infty}(1/T)\delta^2 = \text{constant}, \quad \lim_{T\to\infty}(1/(T-K))\delta^2 = \nu^2 \tag{2.7}$$

The first condition is reasonable because δ^2 includes a moment matrix among exogenous variables which grows in value as the sample size increases. The second condition is necessary to distinguish T and T-K in our study. For simplicity we standardize estimators in the next form:

$$\hat{e}_i = \frac{\sqrt{\omega_{22}}}{\sigma}\delta(\hat{\beta}_i - \beta) \tag{2.8}$$

where i is LIML, TSLS, or OLS. The (traditional) large sample asymptotic distribution of the BAN estimators such as the LIML and TSLS estimators standardized as $\hat{e}$ is the standard normal distribution:

$$\hat{e}_i \sim n(0,1). \qquad i=\text{LIML, TSLS} \tag{2.9}$$

The k-class estimator of β is defined to be

$$\hat{\beta}_k = \frac{g_{12} + (1-k)c_{12}}{g_{22} + (1-k)c_{22}} \tag{2.10}$$

where g_{ij} and c_{ij} are the (k, l)th elements of $G=(g_{kl})=Y'Z'(I-Z_1(Z_1'Z_1)^{-1}Z_1')ZY$, and $C=(c_{kl})=Y'(I-Z(Z'Z)^{-1}Z')Y$, k,l=1,2; the OLS and TSLS estimators are the cases when k=0 and 1, respectively, and the LIML estimator is obtained when $k=1+\lambda_1$, where λ_1 is the smallest root of the equation $|G-\lambda C| = 0$.

Under our model, the estimators standardized as $\hat{e}$ includes only four parameters which are T-K, L, δ^2, and α or alternatively ρ. (See Anderson, Morimune, and Sawa (1983) which estimate values of α and δ^2 for various models.)

3. Empirical Distributions

We compare the LIML, TSLS, and OLS estimators of β by their empirical distributions of 50,000 observations.[3/] In experiments values of four key parameters ρ , T-K, L, and δ^2 are chosen as follows.

Table 1: Values of Parameters in Four Models

	ρ^2	δ^2	T-K	L
Model A	0.5	25	25	5
Model B	0.5	25	20	10
Model C	0.5	25	15	15
Model D	0.5	25	10	20

The value of L+(T-K) is kept to thirty in four models. Since $T=[L+(T-K)]+(1+K_1)$ where $1+K_1$ is the number of coefficients in the structural equation (2.1), the sample size T is thirty five if $1+K_1$ is five. It is easy to perform further experiments by choosing various values of (L+T-K), δ^2, and ρ, but I want to confine this study to bring out effects of greater values of L compared with T-K keeping their sum fixed. (Figures of empirical density functions under various values of ρ are given in the Appendix.) There are some empirical reasons to choose the value of δ^2 to be 25. See Anderson, Morimune, and Sawa (1983).

Figure 1 gives four empirical density functions of $\hat{e}_{LIML}$ of Models A to D. The density in Model A is quite close to the standard normal density function even though it is skewed. If the relative value of L is greater, the density function is flatter and more spread out while the mode almost stays still. It is seen that the LIML density function is deviated far from the standard normal density function in cases C and D.

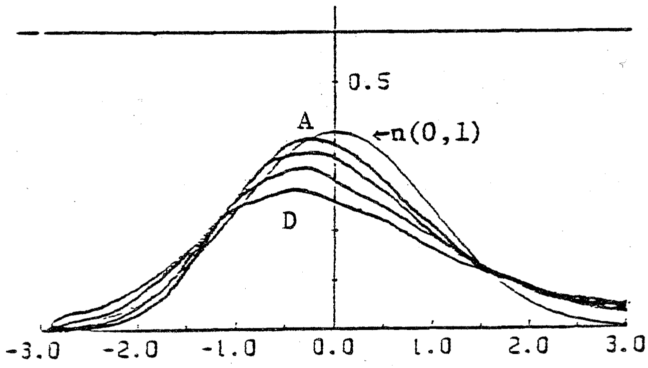

FIG 1: Empirical LIML Densities for Models A, B, C, and D.

The effect of greater value of L is clearly found if we look at the empirical distribution function. Fig 1-2 shows that the distribution function "rotates" about the origin as the value of L gets larger. This figure implies that,

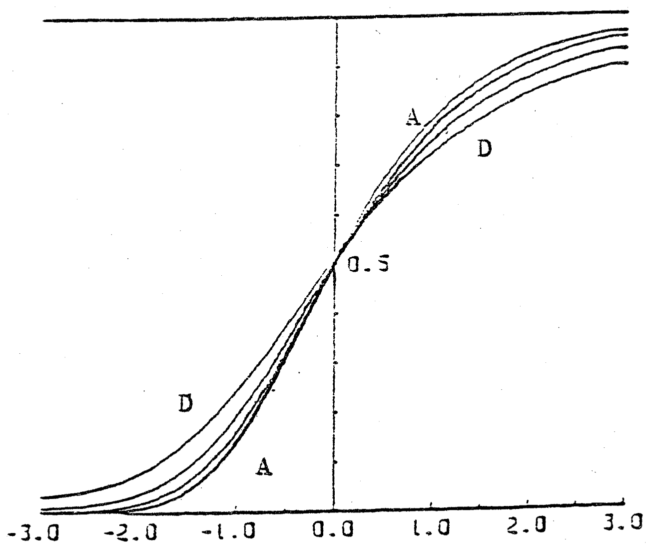

FIG 1-2: Empirical LIML Distribution Functions for Models A, B, C, and D.

taking $\hat{e}_{LIML}$ as the function of degrees of over-identifiability, $P\{ | \hat{e}_{LIML}(L_1) | \geq \xi \} \geq P\{ |\hat{e}_{LIML}(L_2)| \geq \xi \}$ if $L_1 \geq L_2$: The distribution function has thicker tails if the value of L is greater.

Figure 2 gives four empirical density functions of $\hat{e}_{TS}$ of Models A to D. The TSLS density function is quite deviated away from the normal density even in Model A. It is deviated more when the value of L is greater. The TSLS density is concentrated more about the mode than the LIML density, but the mode of the TSLS density is far from the origin. It seems improper to represent the TSLS distribution by the standard normal distribution. This implies that the BAN criterion is misleading to understand properties of the TSLS estimator in realistic situations.

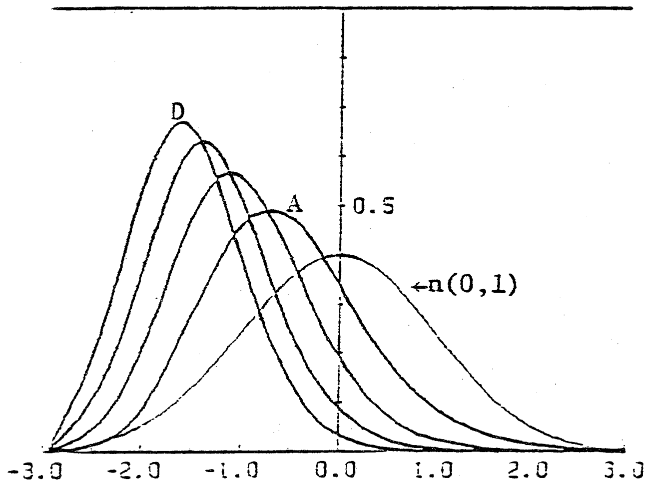

FIG 2: Empirical TSLS Densities for Models A, B, C, and D.

Figure 3 gives an empirical density function of $\hat{e}_{OLS}$. The OLS density function is invariant in Models A to D because it is dependent only upon L+(T-K) which is constant in four models and equal to $T-1-K_1$: the sample size

the number of unknown coefficients. This figure clearly depicts how the

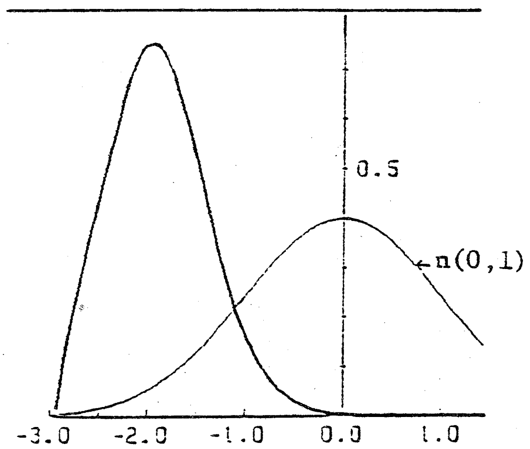

FIG 3: Empirical OLS Densities for Models A, B, C, and D.

inconsistent estimator is distributed and concentrated about a wrong value. The probability that $\hat{e}_{OLS}$ is positive is almost zero in this case.

In sum it is found out from the empirical density functions that the LIML and TSLS estimators are deviated from the normal density in spite of their BAN properties. In particular the TSLS density is not close to that of normal even when L is five or ten. The value of L which is ten is not unrealistic but often found in empirical studies. Therefore it seems improper to use the TSLS estimators in real models relying on its BAN property.

4. Asymptotic Expansions I

Properties of estimators are reviewed from the view point of the conventional large sample asymptotic expansions in this Section. It may be new to compare the estimators by nonparametric measures such as the asymptotic mode and percentiles.

Anderson and Sawa (1973) derived an asymptotic expansion of the distribution of the TSLS estimator of β to $O(T^{-3/2})$. Their result to $O(T^{-1})$ is, denoting $\Phi(.)$ and $\phi(.)$ as the standard normal cdf and its density function,

$$P\{\hat{e}_{TSLS} \leq \xi\} = \Phi(\xi) + \{\frac{\rho}{\delta}(\xi^2 - L) - \frac{\xi}{2\delta^2}[(1-\rho^2)(-L+\xi^2) + \rho^2(L^2 - 2(L+1)\xi^2 + \xi^4)]\}\phi(\xi). \tag{4.1}$$

Anderson (1974) derived an approximation to the distribution of the LIML estimator. His result to $O(T^{-1})$ is

$$P\{\hat{e}_{LIML} \leq \xi\} = \Phi(\xi) + \{\frac{\rho}{\delta}\xi^2 - \frac{\xi}{2\delta^2}[(1 + \rho^2)(L + \xi^2)$$

$$+ \rho^2(-2\xi^2 + \xi^4)]\}\phi(\xi). \qquad (4.2)$$

Anderson (1974) made some comparative statements about these two expansions. The probability in every symmetric interval around β is greater for the LIML estimator if $\rho^2 > 2/(L+2)$. It is expected that the LIML estimator is concentrated more about a true value than the TSLS estimator if the degree of over-identifiability and the simultaneity are large. He also made some comparisons using the first two moments of asymptotic expansions 4/. Instead of moments, some nonparametric characteristics of asymptotic expansions are compared below. First the asymptotic mode of $\hat{e}_{TSLS}$ and $\hat{e}_{LIML}$ are derived by setting the second derivative of (4.1) and (4.2) to be zero:

$$AMODE_{TS} = (L + 2)\rho/\delta + o(T^{-1}), \qquad (4.3)$$

$$AMODE_{LI} = 2\rho/\delta + o(T^{-1}). \qquad (4.4)$$

The AMODE is not derived as an approximation to the exact mode. However, this is expected to remain true under usual regularity conditions because the exact mode may exist on the distributions. (Similar remark is necessary for the asymptotic percentiles discussed below.) It is found from (4.3) and (4.4) that the centers of two estimators have the same sign as ρ ; $\hat{e}_{TSLS}$ is biased more than $\hat{e}_{LIML}$, in particular, when L is large. Next, asymptotic percentiles of the $\hat{e}_{TSLS}$ and $\hat{e}_{LIML}$ are calculated from (4.1) and (4.2) in terms of the percentiles of n(0,1) : denoting x_α the "α-percentile" of n(0,1), the same TSLS percentile is given by

$$x_{TS}^{\alpha} = x_{\alpha} - \frac{\rho}{\delta}(x_{\alpha}^{2} - L) + x_{\alpha}\{\frac{\rho^{2}}{\delta^{2}}(x_{\alpha}^{2} - 2L) + \frac{1-\rho^{2}}{2\delta^{2}}(x_{\alpha}^{2} - L)\} + o(T^{-1}); \qquad (4.5)$$

similarly for the LIML estimator,

$$x_{LI}^{\alpha} = x_{\alpha} - \frac{\rho}{\delta}x_{\alpha}^{2} + x_{\alpha}[\frac{\rho^{2}}{\delta^{2}}x^{2} + \frac{1-\rho^{2}}{2\delta^{2}}(x_{\alpha}^{2} + L)] + o(T^{-1}). \qquad (4.6)$$

By substituting zero for x_{α}, it is found that the asymptotic median of $\hat{e}_{TSLS}$ is $(L\rho/\delta)$, but it is zero for $\hat{e}_{LIML}$. The LIML estimator is asymptotically median unbiased, but the TSLS estimator is asymptotically median biased, and it is worse as L and ρ are greater in value. It is also possible to calculate the asymptotic inter-quantile range (AIQR) between the $(100 - \alpha)$ percentile and the α percentile. Then, by (4.5) and (4.6), the AIQR for the TSLS and LIML are

$$AIQR_{TS}^{\alpha} = 2x_{\alpha}\{1 + [\frac{\rho^{2}}{\delta^{2}}(x_{\alpha}^{2} - 2L) + \frac{1-\rho^{2}}{2\delta^{2}}(x_{\alpha}^{2} - L)]\} + o(T^{-1}), \qquad (4.7)$$

$$AIQR_{LI}^{\alpha} = 2x_{\alpha}\{1 + [\frac{\rho^{2}}{\delta^{2}}x_{\alpha}^{2} + \frac{1-\rho^{2}}{2\delta^{2}}(x_{\alpha}^{2} + L)]\} + o(T^{-1}). \qquad (4.8)$$

It is found that $AIQR_{TS}^{\alpha} < AIQR_{LI}^{\alpha}$ for all values of x_{α}. This implies that $\hat{e}_{TSLS}$ is concentrated more about the center of its distribution, i.e., mode or median not about the origin, than $\hat{e}_{LIML}$ is. It also follows that

$$\partial AIQR_{TS}^{\alpha} / \partial L < 0, \qquad (4.9)$$

$$\partial AIQR_{TS}^{\alpha} / \partial\rho^{2} < 0 \quad \text{if} \quad x_{\alpha}^{2} < 3L. \qquad (4.10)$$

Then $\hat{e}_{TSLS}$ is concentrated more about its center if the values of L and ρ^{2} are greater. Since the first term $2x_{\alpha}$ on the right side of (4.7) is the IQR of

$n(0, 1)$, $\hat{e}_{TSLS}$ is concentrated more about the center than the standard normal distribution as long as $x_\alpha^2 \leq L$. (If L is 3, x_α is the 92 percentile so that the equality holds.) On the other hand the mode of $\hat{e}_{LI}$ is less biased than that of $\hat{e}_{TSLS}$, and $\hat{e}_{LIML}$ is asymptotically median unbiased. The distribution of $\hat{e}_{LIML}$ has fatter tails than $\hat{e}_{TSLS}$ and $n(0,1)$ because $AIQR^\alpha_{LI} > AIQR^\alpha_{TS}$ and $AIQR^\alpha_{LI} > 2x_\alpha$ uniformly. Further

$$\partial AIQR^\alpha_{LI} / \partial L > 0 \tag{4.11}$$

which is opposite to (4.9), and

$$\partial AIQR^\alpha_{LI} / \partial \rho^2 < 0 \quad \text{if } x_\alpha^2 < L. \tag{4.12}$$

The $\hat{e}_{LIML}$ is less concentrated if the value of L is greater and more concentrated if the value of ρ^2 is greater about its center of distribution.

These theoretical properties derived by asymptotic expansions (4.1) and (4.2) reassures characters of estimators found by examining empirical distributions in the Section 3: Effects of L and ρ^2 on expansions summarized by (4.9) to (4.12) have been realized by empirical distributions. (The effect of ρ on empirical densities are summarized in the Appendix.) It may be reasonable to say that properties of estimators observed from empirical distributions hold more generally than only to the few models examined in the Section 3.

5. Asymptotic Expansions II

Anderson and Sawa (1979) performed various numerical evaluations of the exact TSLS distributions. One result by Anderson and Sawa (1979) is that the asymptotic expansion given by (4.1) provides poor approximations to the exact

TSLS distributions even when L is five, and α is two ($\rho^2=0.8$).

This inaccuracy of approximation may be explained, at least partly, by terms of $(L/\delta)^i$ i=1,2,3 in the expansion. (The expansion is originally to $O(\delta^{-3})$, and i=3 appears in terms of $O(\delta^{-3})$.) To avoid this difficulty a new condition is introduced recently: $L=O(T)$ or $L=O(\sqrt{T})$. This seems the only way to incorporate into the theoretical framework such an intuitive situation phrased as "the value of is not negligible compared with the sample size, and L/δ is even greater than one". For simplicity effects of the condition $L=O(T)$ on asymptotic expansions are summarized in this section. (See Morimune (1983) for more comprehensive analyses.) Together with the assumption (2.2), $L=O(T)$ implies that[5/]

$$\lim_{T\to\infty} \delta^2/L = \tau^2, \tag{5.1}$$

where τ^2 is a newly defined constant. Once a new condition is added to the analyses, the probability limit of $\hat{e}_{TSLS}$ is given by

$$\operatorname*{plim}_{T\to\infty}(1/\sqrt{T})\hat{e}_{TSLS} = \lim_{T\to\infty}(\delta/\sqrt{T})(1 - (1/\psi))\rho \tag{5.2}$$

where $\psi=1+\zeta$ defining $\zeta=1/\tau^2$. The eq.(5.2) implies that the TSLS estimator is inconsistent (p.608 of Theil (1971); Kunitomo (1980)). The eq.(5.2) which is a location of $\hat{e}_{TSLS}$ is a increasing function of ρ and L. Defining $\varepsilon=[1+(1-\rho^2-(2\rho^2/\psi^2))\zeta]$, the asymptotic variance of $\hat{e}_{TSLS}$ is ε/ψ^2 which is less than one. This variance is a decreasing funciton of ρ and L. These relationships between the two asymptotic moment and L or ρ are in agreement with the properties found from empirical distributions. The asymptotic expansion to $O(T^{-1})$ is given as follows.

$$P\{\frac{\psi}{\sqrt{\varepsilon}}(\hat{e}_{TSLS}-(1-\frac{1}{\psi})\delta\rho)\leq\xi\} = \Phi(\xi) - \frac{1}{\delta}\Big\{\{\kappa_1 + \frac{\kappa_3}{6}(\xi^2-1)\}$$
$$+\frac{\xi}{\delta}\{\frac{\kappa_4}{24}(\xi^2-3) + \frac{\kappa_3^2}{72}(\xi^4-10\xi^2+15) + \frac{1}{2}(\kappa_2+\kappa_1^2)\}\Big\}\phi(\xi) + o(T^{-1}), \quad (5.3)$$

where $\kappa_1 = \rho(A+1)/(\psi\sqrt{\varepsilon})$, $\kappa_2 = (1/\varepsilon)[1-\rho^2+(2\rho^2/\psi^2)]+(1/\psi^2)[(5\rho^2/\varepsilon)A^2+12+6\zeta]$ $-(2/(\psi\varepsilon))[2(1-\rho^2)-(\rho^2/\psi)(A+8B)]-(2/\psi)[1+(2C/\varepsilon)]$, $\kappa_3/6 = (\rho/\sqrt{\varepsilon})[(A/\psi)-(1-\rho^2)(E/\varepsilon)]$ $\kappa_4/24 = [(1-\rho^2)/\varepsilon^2]\{(1/2)[1+(1-\rho^2)(0.5-(4/\psi^2))\zeta]-(\rho^2/\psi)E-2(\rho^2/\psi)[\zeta-(D/(1-\rho^2))]\}$ $+(1/\psi^2)\{(7\rho^2/(2\varepsilon))A^2+2+\zeta\}-(\rho^2/(\psi\varepsilon))\{(1-\rho^2)(4/\varepsilon)A\cdot E+1-(1/\psi)(A+4B)\}+(2\rho^4/(9\psi^4\varepsilon^2))\zeta$ $-C/(\psi\varepsilon)$, defining $A=-2/\psi$, $B=(1/\psi)\{\psi-2-(2/3)\zeta\}$, $C=1+[1-\rho^2-(8\rho^2/(3\psi^2))]\zeta$, $D=1+[1-\rho^2-(28\rho^2/(9\psi^2))]\zeta$, and $E=(1/\psi)[\zeta-(C/(1-\rho^2))]$. This final result was given by Morimune (1983) in the large L expansion of the general fixed k-class estimator.[6/] It is also proved by Morimune (1983) that the conventional large sample asymptotic expansion of $\hat{e}_{OLS}$ to $O(T^{-1})$ coincides with (5.3) if we change the definition of ζ to the following:

$$\zeta = (1/\tau^2) + (1/\nu^2) = \lim_{T\to\infty}((L+T-K)/\delta^2). \quad (5.4)$$

The $\hat{e}_{TSLS}$ expansion in this section is different from the conventional $\hat{e}_{OLS}$ expansion by one element. The new ζ has effects on the location (5.2), the dispersion, and the expansion itself. However it is naturally expected that the distributions of $\hat{e}_{OLS}$ and $\hat{e}_{TSLS}$ are similar. Figure 4 clearly depicts the TSLS empirical density in Model A to D and the OLS empirical density. The TSLS density in Model D is very close to the OLS density in shape and location. It is also seen that the OLS density is

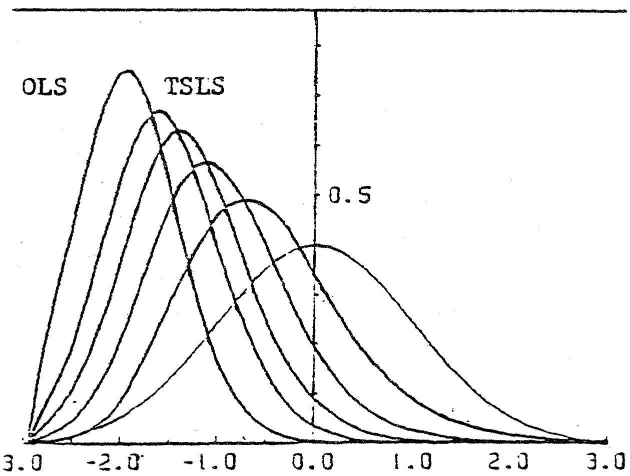

FIG 4: Empirical TSLS and OLS Densities. The leftest density is of the OLS

a sort of limit of the TSLS density in the sequence where L increases.

The asymptotic expansion of the distribution of $\hat{e}_{LIML}$ under the additional condition (5.1) is given below:

$$P\{\frac{1}{\sqrt{\eta}}\hat{e}_{LIML} \le \xi\} = \Phi(\xi) + \frac{\rho}{\delta}\sqrt{\eta}\,\xi^2\,\phi(\xi)$$

$$- \frac{\xi}{2\delta^2}\{\rho^2\eta(\xi^4 - 2\xi^2) + \frac{1-\rho^2}{2\eta^2}(1 + \lambda_0 - \lambda_0^2)(\xi^2 - 3)$$

$$+ 2(1-\rho^2)(\eta + \lambda_0)\xi^2 + \frac{1-\rho^2}{2\eta}[(\lambda_0^2 - 5\lambda_0 - 3)\xi^2 + \lambda_0^2 + 7\lambda_0 + 3]\}\,\phi(\xi), \quad (5.5)$$

to $O(T^{-1})$ defining $\lambda_0 = \upsilon^2/\tau^2$, and $\eta = 1+((1-\rho^2)/\tau^2)(1+\lambda_0)$.

The mean and variance of the asymptotic distribution of $\hat{e}_{LIML}$ under a new condition are zero and η. Contrary to $\hat{e}_{TSLS}$, the location of $\hat{e}_{LIML}$ remains at the origin. However its dispersion η is greater than unity which is the asymptotic variance of $\hat{e}_{LIML}$ in the conventional large sample theory. It seems that η characterizes the well known "fat" tails of the LIML distribution better than the traditional analyses. The definition of η inplies that the asymptotic variance is a increasing function of L but a decreasing function of ρ^2. These relationships agree with properties of $\hat{e}_{LI}$ summarized from empirical distributions.

6. Accuracy of Approximations

Now I examine accuracy of the two kinds of expansions for each of $\hat{e}_{LIML}$ and $\hat{e}_{TSLS}$ by Figures. Since the exact density has never been calculated for the LIML estimator the empirical density is taken as the "true" density in accuracy checkings. Models are the same as those used in Section 3. I denote A, B, C, or D in each Table to make clear which model it is about. The thick solid line is evaluation of the density of the conventional large sample expansion. The thin solid line gives the empirical density without smoothing. Asterisks* are evaluation of the expansion given is Section 5. Asterisks are dotted at every two decimals on the horizontal axis to avoid confusion with the empirical density.

First Figures 5A to 5D depict the empirical density and numerical evaluations of derivative of two expansions (4.1) and (5.3) for the TSLS estimator. It is obvious from the four Figures that accuracy of the conventional expansion is rather miserable; accuracy is worse if the value of L is greater but the expansion is useless even when L is ten. On the other hand, the expansion given by (5.3) is extremely accurate. It is accurate when L is as large as twenty in Model D, but it is as accurate as the large sample expansion when L is five.

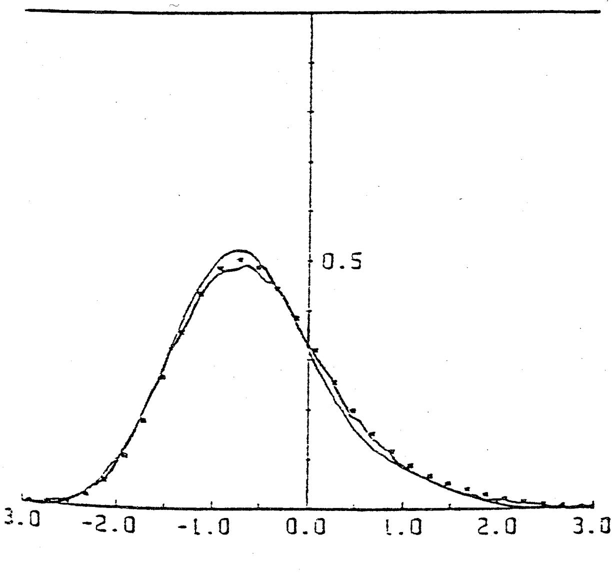

FIG 5A: Empirical TSLS Density and Asymptotic Expansions for Model A.

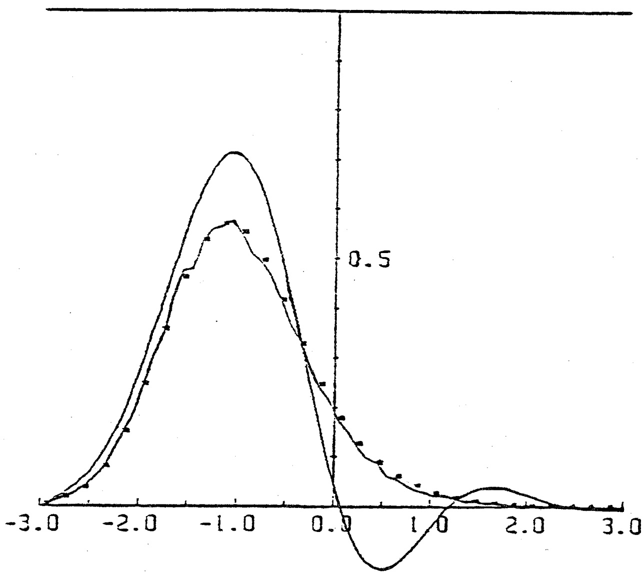

FIG 5B: Empirical TSLS Density and Asymptotic Expansions for Model B.

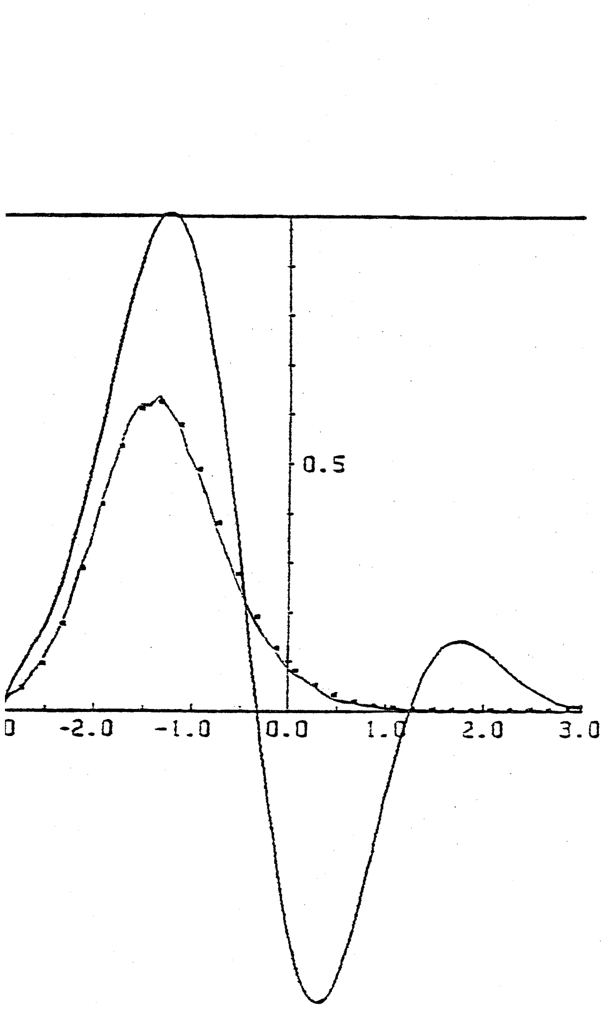

FIG 5C: Empirical TSLS Density and Asymptotic Expansions for Model C.

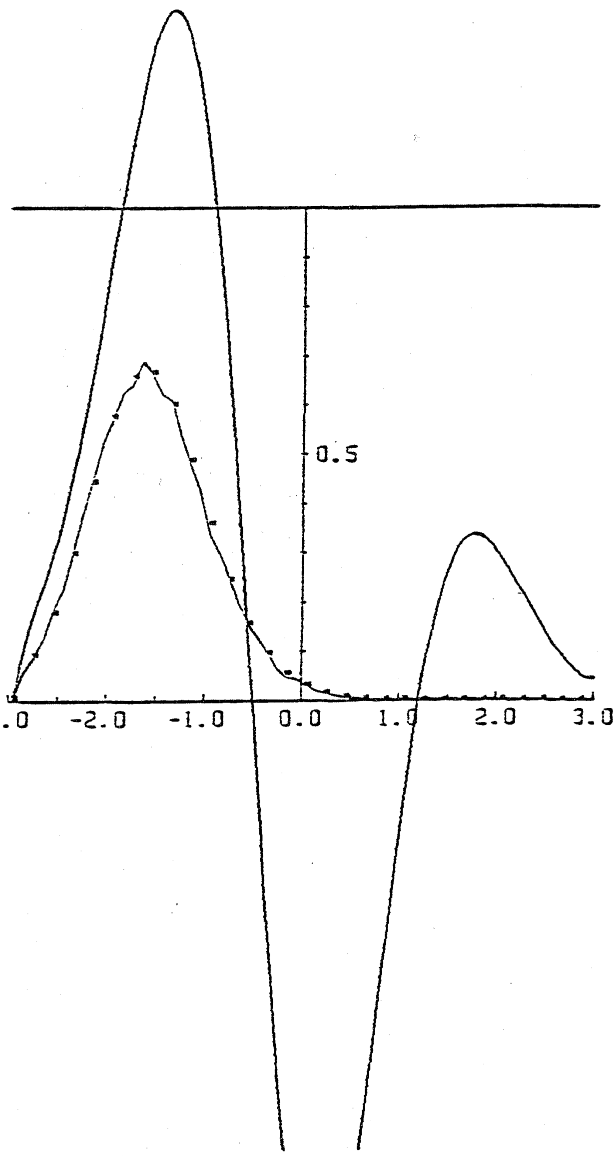

FIG 5D: Empirical TSLS Density and Asymptotic Expansions for Model D.

Table 2 below tabulates the mean and standard deviation of the empirical distribution of $\hat{e}_{TSLS}$. Table 2 also tabulates the probability limit and the asymptotic standard deviation of two asymptotic expansions for $\hat{e}_{TSLS}$. They are zero and 1 in the conventional expansion, and they are (5.2)* and $\sqrt{\epsilon}/\psi$ in the large-L expansion. It is surprising to see how closely the first two empirical moments are approximated by the large-L expansion.

Table 2: Mean and Standard Deviation of $\hat{e}_{TSLS}$

	Model A			Model B			Model C			Model D		
	E.D.	large -T	large -L	E.D.	large -T	large -L	E.D.	large -T	large -L	E.D.	large -T	large -L
mean	-0.60	0	-0.59	-0.99	0	-1.01	-1.32	0	-1.32	-1.57	0	-1.57
stand. dev.	0.89	1	0.82	0.74	1	0.71	0.67	1	0.65	0.62	1	0.60

* $\hat{e}_{TSLS} \sim n(0,1)$ in the large-T sequence, but it is $n(\delta(1-(1/\psi))\rho\ ,\ \epsilon/\psi^2)$ in the large-L sequence.

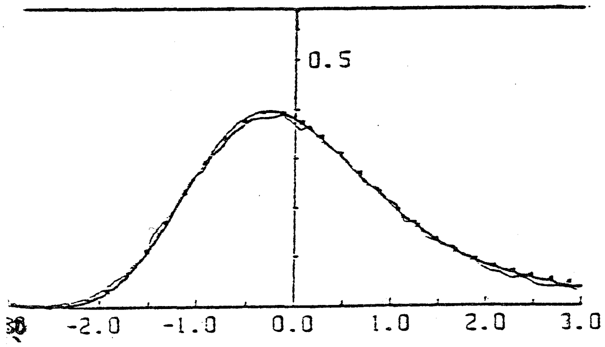

FIG 6A: The Empirical Density and Asymptotic Expansions for Model A.

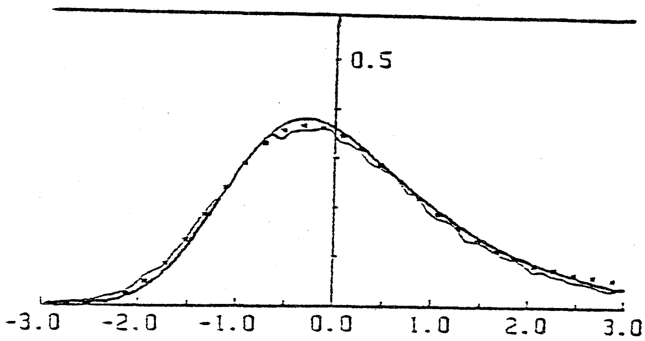

FIG 6B: The Empirical Density and Asymptotic Expansions for Model B.

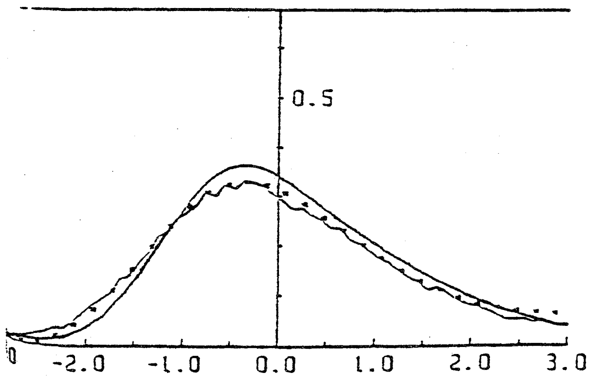

FIG 6C: The Empirical Density and Asymptotic Expansions for Model C.

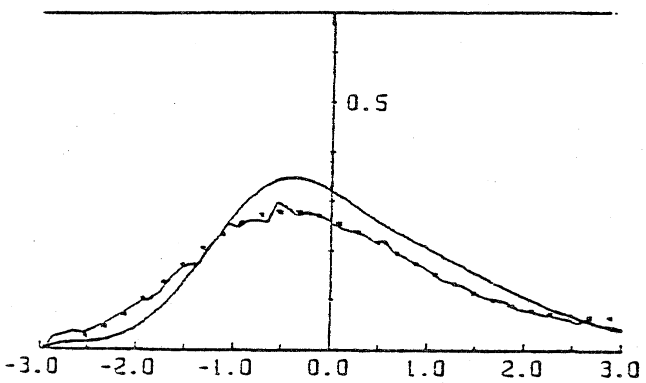

FIG 6D: The Empirical Density and Asymptotic Expansions for Model D.

Figures 6A to 6D depict the empirical density and derivative of two expansions (4.2) and (5.5) for the LIML estimator. It is again obvious from these figures that the expansion given by (5.5) is generally more accurate than the large sample expansion, and particularly when L is large. It is found again that the (5.5) expansion is not worse than the (4.2) expansion even when L is five. Accuracy of asymptotic moments are not studied for the LIML estimator because there are not proper estimators of exact LIML moments which "do not exist."

7. Improvement of the LIML Estimator

There have been some efforts to improve estimators. Probably the most important improvement from a practical point of view was proposed by Fuller (1977). He modified the LIML estimator by subtracting a small constant from the smallest characteristic root: λ is replaced by $(\lambda - c/(T-K))$ where c is any constant.[7/] For simplicity the modified LIML estimator by Fuller will be referred to as the F estimator. The exact moments exist for the F estimator. The traditional asymptotic distribution of the F estimator is the same as those of the LIML and TSLS estimators. The F estimator is BAN. Fuller also derived the first two moments of the large sample asymptotic expansion and proved that the F estimator is asymptotically unbiased to $O(T^{-1})$ if c is set to unity.

Takeuchi and Morimune (1983) has arrived at the F estimator from a different view point. The first theorem in their paper is the extension of the recently developed "third order efficiency" theories that are mostly about i.i.d. observations to the multivariate linear regression problem. The theorem is then applied to the simultaneous equation model, in particular, to the subsystem model. Results on the single equation model is a specific case of the subsystem model. If we confine to the estimation of β in a single structural equation, their result is summarized as follows: let $\hat{\beta}_{BAN}$ be a regular BAN estimator of β, and $\hat{\beta}_{ALI}$ be the adjusted LIML estimator such that $AM_T(\hat{e}_{ALI})=AM_T(\hat{e}_{BAN})$ where $AM_T(.)$ is the mean operator dependent upon the derivative of the asymptotic expansion of distribution to $O(T^{-1})$ such as (4.1). Then, for any interval which includes the origin, it follows that

$$\lim_{T\to\infty} T\left\{P\{\xi_1 \leq \hat{e}_{ALI} \leq \xi_2\} - P\{\xi_1 \leq \hat{e}_{BAN} \leq \xi_2\}\right\} \geq 0. \tag{7.1}$$

This result itself does not yeild a practical solution. However if the comparison is limited to the BAN estimators which do not include higher order biases, namely that $AM_T(\hat{e}_{BAN})=0$, then the F estimator with c=1 is also the ALI estimator. The F estimator is the best in the sense of (7.1) among all bias removed BAN estimators.

It seems that the F estimator is the best for practice because its exact moments exist, it is BAN, third order efficient, and easy to compute. It is possible to add another advantage to this list. Morimune (1983) derived the asymptotic expasion of the F distribution under the additional condition of (5.1). It is found that, for any symmetric interval about the origin,

$$P\{|\hat{e}_{LIML}| \leq \xi\} \leq P\{|\hat{e}_F| \leq \xi\} + o(T^{-1}), \tag{7.2}$$

where the inequality is in terms of the large-L asymptotic expansions to $O(T^{-1})$.

Since the F estimator has not been numerically analyzed much, I first give the empirical density function of $\hat{e}_F$ in Figure 7 for Models A to D. If Figure 7 is compared with Figure 1, it is found that shapes of the empirical density function of $\hat{e}_F$ is very similar to that of $\hat{e}_{LIML}$. The large-L asymptotic expansion of this estimator is given by Morimune (1983). Here I study accuracy of expansions only by comparing empirical moments and moments of asymptotic expansions as they are given in Table 3. Accuracy of the

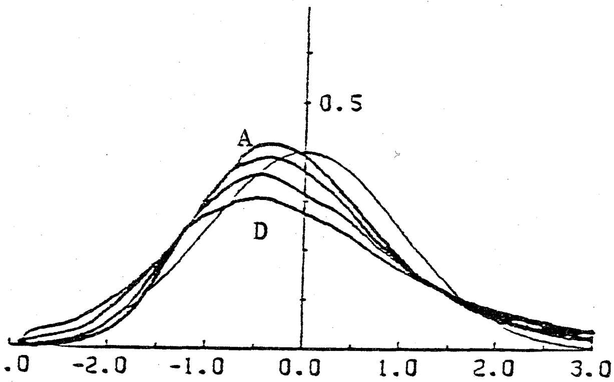

FIG 7: The Empirical F Densities for Models A, B, C, and D.

Table 3: Mean and Standard Deviation of $\hat{e}_F$

	Model A			Model B			Model C			Model D		
	E.D.	large -T	large -L	E.D.	large -T	large -L	E.D.	large -T	large -L	E.D.	large -T	large -L
mean	0.00	0.0	0.0	0.02	0.0	0.0	0.02	0.0	0.0	-0.04	0.0	0.0
stand.dev.	1.11	1.0	1.06	1.26	1.0	1.14	1.48	1.0	1.26	1.94	1.0	1.48

* $\hat{e}_F \sim n(0,1)$ in the large-T sequence, but it is $n(0,\eta)$ in the large L sequence.

asymptotic standard deviation in the large-L sequence is not so accurate as it has been observed in Table 2 with respect to $\hat{e}_{TSLS}$. However it gives better approximation to the empirical standard deviation than the conventional large sample sequence.

Finally I calculate the probability of concentration about the origin for $\hat{e}_{LIML}$ and $\hat{e}_F$. It is not sure whether $P\{-\xi \leq \hat{e}_F \leq \xi\}$ is greater than $P\{-\xi \leq \hat{e}_{LIML} \leq \xi\}$ in small samples even though this inequality holds in the large-L sequence. Figures 8A to 8D clearly confirm that the inequality holds in small samples as far as Models A to D are concerned.

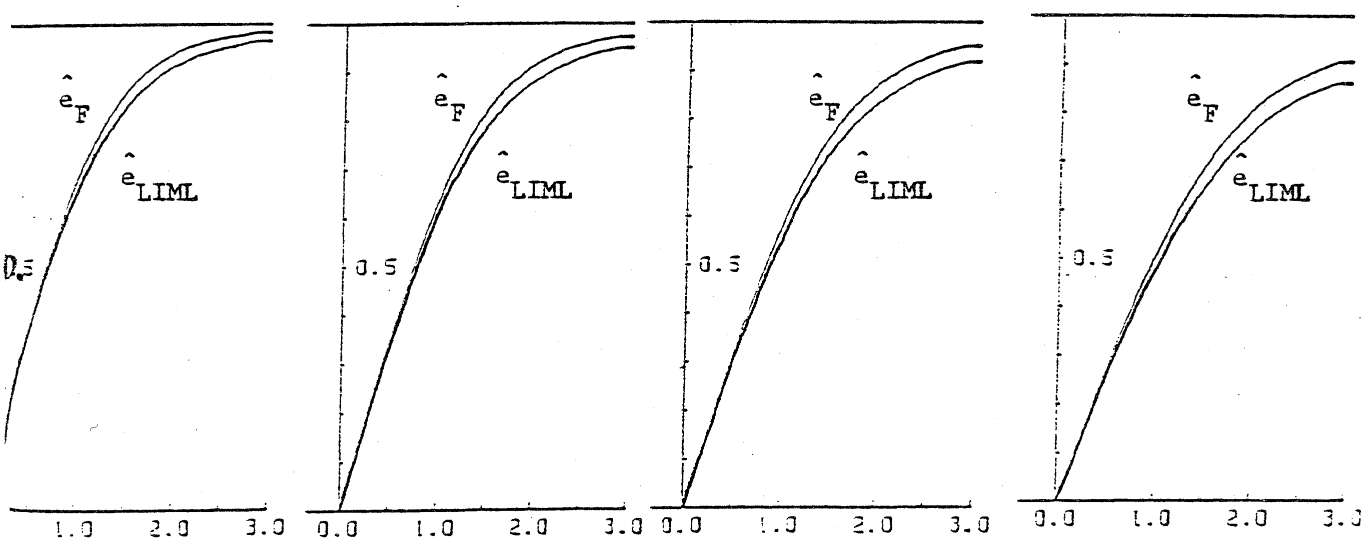

FIG. 8A to 8 D: $P\{-\xi \leq e \leq \xi\}$ for positive ξ is figured with respect to $\hat{e}_{LIML}$ and $\hat{e}_F$

8. Conclusion

There have been two objectives in this article to study properties of estimators: first the partial effect of the degree of over-identifiability and then the coefficient of simultaneity on the density functions of estimators.

For this purpose I have set the value of key parameters as presented by Table 1 and Table 4 in the Appendix A, and performed Monte Carlo experiments to calculate empirical densities. This analysis has brought out the fact that the BAN property of the TSLS estimator is misleading and not trustworthy in small samples. This estimator is biased, and it is biased as much as the OLS estimator in small samples when the degree of over-identifiability is large. As for the LIML estimator, it stays still about the origin in small samples under various sets of parameters even though the density function has a thick tail when the degree of over-identifiability is large. It seems, on the whole, that the LIML estimator is more reliable than the TSLS estimator because the former is distributed about the true value any way, but the latter is distributed about a wrong value in small samples. Numerical analysis also confirms that the OLS estimator is biased in small samples: it is correct to assess the OLS estimator in small samples by its inconsistency in the large sample theory.

Secondly the accuracy of the conventional large sample asymptotic expansions and the new kind of asymptotic expansions for the estimators is studied by comparing expansions with empirical densities. It was found that the new kind of expansions are more accurate than the conventional expansions for the LIML estimator and particularly for the TSLS estimator when the degree of over-identifiability is large. The new expansions are as accurate as the conventional expansions even when the degree of over-identifiability is small. This implies that the small sample properties of the LIML and TSLS estimators are represented better by the new expansions than by the conventional expansions.

This may be alternatively summarized as that the properties of the esimators derived from the empirical distributions hold more generally than only to a few models for which experiments are performed: the TSLS estimator is biased and concentrated about the bias; the LIML estimator is not biased much but has thicker tails than expected by the conventional theory.

It is known that the LIML and TSLS estimators are identical when the degree of over-identifiability is zero. It has been also proved that the LIML, TSLS, and OLS estimators are identical when T-K is zero, i.e., the degree of over-identifiability is equal the sample size minus the number of coefficients. The conventional expansion and the new expansion describe properties of estimators between these two extremities. The conventional asymptotic theory classify the LIML and TSLS estimator into the class of BAN (good) estimators but the OLS estimator into the class of inconsistent (bad) estimators. The new expansion classify the LIML estimator into the class of the best estimators (Kunitomo (1982)) but the TSLS and OLS estimators into the class of inconsistent estimators. The numerical analyses assure that the small-sample properties of estimators are better assessed by the new expansion.

Finally I gave some numerical and theoretical analyses on the modified LIML estimator by Fuller. In short, this modified estimator has similar distributions as the LIML estimator, but it is more efficient than the LIML estimator asymptotically in the new sequence and has exact moments.

Footnote

1/ Other method includes the full-system method: to estimate all coefficients included in a system jointly, and the sub-system method: to estimate all coefficients included in a subset of equations in a system.

2/ The exact LIML distribution was derived by Mariano and Sawa (1972) but has never been numerically evaluated. The exact TSLS distribution was derived by Sawa (1969) and numerically evaluated by Anderson and Sawa (1979), but it uses an Edgeworth expansion of the doubly noncentral F distribution in "exact" calculations.

3/ In short $\hat{e}_i$ is a function of the four key parameters and independently distributed standard normal random variables. See Morimune and Tsukuda (1983) for details.

4/ These moments are defined to be the expected values of $\hat{e}$ and $\hat{e}^2$ where the integral is taken with respect to the derivative of (4.1) and (4.2). However there are some ambiguity left in comparing moments of asymptotic expansions because they are not necessarily approximations to the exact moment-paricularly when they do not exist. Further, the exact moments do not exist for the LIML estimator. See Mariano and Sawa (1972).

5/ If it is possible to derive a valid expansion such as by Durbin (1980) or by Kariya and Maekawa (1982), it may be unnecessary to assume that the value of L increases to infinity nor to assume that the sample size T increases to infinity. See Taylor (1983) for criticisms against the large-L sequence.

6/ The equation (5.3) should be identical to the large sample asymptotic expansion of the least squares estimator in the functional relationship model. This latter expansion was derived by Kunitomo (1980) but the final equation is not given.

7/ It is known that the exact moments do not exist for the LIML estimator because $(\underset{\sim}{G}_{22}-\lambda\underset{\sim}{C}_{22})$ is positive semidefinite, but not positive definite. Note that $\underset{\sim}{G}_{22}$ and $\underset{\sim}{C}_{22}$ are diagonal sub-blocks of $\underset{\sim}{G}$ and $\underset{\sim}{C}$ corresponding to $\underset{\sim}{Y}_2$, and a matrix inversion of $(\underset{\sim}{G}_{22} -\lambda\underset{\sim}{C}_{22})$ is necessary in the LIML estimation. If ψ is the smallest root of an equation $|\underset{\sim}{G}_{22}-\psi\underset{\sim}{C}_{22}|=0$, then $\psi \leq \lambda$, and there is a positive probability that the equality holds.

References

Anderson, T. W. (1974), "An Asymptotic Expansion of the Limited Information Maximum Likelihood Estimate of a Coefficient in a Simultaneous Equation System," Journal of the American Statistical Association, 69, 565-573,

Anderson, T. W., Morimune, K., and Sawa, T. (1983), "The Numerical Values of Some Key Parameters In Econometric Models," Journal of Econometrics, 21, 229-243.

Anderson, T. W. and Rubin, Herman (1949), "Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations," Annals of Mathematical Statistics, 20, 46-63.

Anderson, T. W. and Sawa, T. (1973), "Distribution of Estimates of Coefficients of a Single Equation in a Simultaneous System and Their Asymptotic Expansions," Econometrica, 41, 683-714.

Anderson, T. W. and Sawa, T. (1979), "Evaluation of the Distribution Function of the Two-stage Least Squares Estimate," Econometrica, 47, 163-183.

Basmann, R. L. (1957), "A Generalized Classical Method of Linear Estimation of Coefficients in a Structural Equation," Econometrica, 25, 77-83.

Durbin, J. (1980), "Approximations for Densities of Sufficient Estimates," Biometrika, 67, 311-333.

Fuller, Wayne (1977), "Some Properties of a Modification of the Limited Information Estimator," Econometrica, 45, 939-953.

Kariya, T., and Maekawa, K. (1982), "A Method for Approximations to the PDF's and CDF's of GLSE's and Its Application to the Seemingly Unrelated Regression Model," Annals of the Institute of Statistical Mathematics, 34, 281-297.

Klein, Laurence (1950), Economic Fluctuations in the United States, 1921-1941. New York: John Wiley and Sons.

Kunitomo, N. (1980), "Asymptotic Expansions of the Distributions of Estimators in a Linear Functional Relationship and Simultaneous Equations," Journal of the American Statistical Association, 75, 693-700.

________. (1982), "Asymptotic Efficiency and Higher Order Efficiency of the LIML Estimator When the Number of Exogenous Variables Is Large," mimeographed.

Mariano, R. S., and Sawa, T. (1972), "The Exact Finite-Sample Distribution of the Limited-Information Maximum Likelihood Estimator in the Case of Two Included Endogenous Variables," Journal of the American Statistical Association, 67, 159-163.

Morimune, K. (1983), "Approximate Distributions of k-Class Estimators When the Degree of Overidentifiability is Large Compared With the Sample Size," Econometrica, 51, 821-842.

Morimune, K. and Tsukuda, Y. (1981), "Asymptotic Expansions of Distributions of the Structural Variance Estimators in a Simultaneous Equation System," ~~submitted to the journal~~.

Taylor, W. E. (1983), "On the Relevance of Finite Sample Distribution Theory," Econometric Reviews, in print.

Takeuchi, K. and Morimune, K. (1982), "Third-Order Efficiency of the Extended Maximum Likelihood Estimators In a Simultaneous Equation System," Econometrica, tentatively accepted.

Theil, Henri (1958), Economic Forecast and Policy, Amsterdam; North Holland.

__________, (1971), Principles of Econometrics, New York: John Wiley.