

Discussion Paper No. 502
ON A THIRD ORDER OPTIMUM PROPERTY
OF THE LIML ESTIMATOR WHEN THE
SAMPLE SIZE IS LARGE

by

Naoto Kunitomo*

October 1981

* Department of Economics-Northwestern University

This paper is based on Chapter 4 of the author's Ph.D. dissertation submitted to Stanford University. The author is grateful to his principal advisor, Professor Theodore W. Anderson, for his encouragement and comments in the development of the thesis.

Abstract

We make some comparison of single-equation methods in the simultaneous equations system. We show that the Limited Information Maximum Likelihood (LIML) estimator and the modification by Fuller (F estimator) have an optimum property among a class of almost median-unbiased and almost mean-unbiased estimators. We also investigate the relationships among the modified estimators by several econometricians when the sample size increases and the effect of exogenous variables increases along with the sample size.

1. Introduction

For estimating the coefficients of a single equation in the complete system of simultaneous structural equations, the Two-Stage Least Squares (TSLS), the Limited Information Maximum Likelihood (LIML), the Ordinary Least Squares (OLS) methods have been commonly used in practice. Under appropriate general conditions, the first two methods yield consistent estimators; the two sets of estimators normalized by the square root of the sample size or alternatively the noncentrality parameter have the same limiting joint normal distributions with the covariance of the standardized Fisher information matrix. Therefore, both are asymptotically efficient and are called the best asymptotically normal (BAN) estimator.

Since there exist two BAN estimators, several modifications have been proposed in order to improve the LIML and TSLS estimators. Sawa [1973a] modified TSLS by combining it with OLS and removing its asymptotic bias (S estimator). By the same token, Morimune [1978], and Morimune and Kunitomo [1980] modified LIML by combining it with TSLS and removing its asymptotic bias (M estimator). Further, they showed that the M estimator has smaller Asymptotic Mean Squared Error (AMSE) than the LIML estimator does. These improvements of S and M estimators were done by use of the small disturbance asymptotics initiated by Kadane [1971].

On the other hand, Fuller [1977] modified the LIML estimation method by perturbing the smallest root of the characteristic equation (F estimator). Takeuchi [1978] proposed another modification of LIML (T estimator) and suggested

that the proposed estimator has a third order asymptotic efficiency, whose concept has been developed in last few years in the asymptotic theory. (See Akahira and Takeuchi [1979], for example.)

In this article we shall give two theorems (Theorem 1 and Theorem 2) stating that the F and LIML estimators have optimum properties among a class of almost mean-unbiased and median-unbiased estimators, respectively, with respect to any bounded bowl-shaped loss function when the sample size increases and the effect of the exogenous variables increases along with the sample size (i.e. the large sample asymptotic theory). We also show that the T estimator is asymptotically equivalent to the F estimator and so they share the same asymptotic optimum property to terms of $T^{-3/2}$ where T is the sample size. (Corollary 1.) Further, it turns out that the S estimator is asymptotically equivalent to a k-class estimator proposed by Nagar [1959] and Kadane [1971] (N estimator). This finding together with Theorem 1 leads to the conclusion that both the F estimator and the T estimator dominate the M estimator, the S estimator and the N estimator, uniformly to terms of T^{-1} . (Theorem 3.) These results may shed some light on the comparison of single-equation estimation methods in the simultaneous equation system.

2. Model and Estimators

We consider a single structural equation represented by

$$(2.1) \quad \underline{y}_1 = \underline{Y}_2 \underline{\beta} + \underline{Z}_1 \underline{\gamma} + \underline{u} \quad ,$$

where \underline{y}_1 and \underline{Y}_2 are $T \times 1$ and $T \times G_1$ matrices of T observations on the endogenous variables, \underline{Z}_1 is a $T \times K_1$ matrix of T observations on the K_1 exogenous variables, $\underline{\beta}$ and $\underline{\gamma}$ are column vectors of G_1 and K_1 parameters, and \underline{u} is a $T \times 1$ column vector of unobservable disturbances. The reduced form of the system of structural equations is defined as

$$(2.2) \quad \underline{Y} = \underline{Z} \underline{\Pi} + \underline{V} \quad ,$$

where $\underline{Y} = (\underline{y}_1 \underline{Y}_2)$, $\underline{Z} = (\underline{Z}_1 \underline{Z}_2)$ is a $T \times K$ ($K = K_1 + K_2$) matrix of exogenous variables, $\underline{\Pi}$ is a $K \times (1 + G_1)$ matrix of the reduced form coefficients, and $\underline{V} = (\underline{v}_1 \underline{V}_2)$ is a $T \times (1 + G_1)$ matrix of unobservable disturbances. We make the following conventional assumptions.

Assumption 1: The rows of \underline{V} are independently normally distributed, each row having mean 0 and nonsingular covariance matrix

$$(2.3) \quad \underline{\Omega} = \begin{pmatrix} \omega_{11} & \omega_{12} \\ \omega_{21} & \omega_{22} \end{pmatrix} .$$

Assumption 2: The matrix Z is of rank K and $q = T - K > 0$.

In order to relate (2.1) and (2.2), postmultiplying (2.2) by

$(1 - \beta')$, we obtain $\underline{u} = \underline{v}_1 - \underline{V}_2\beta$, $\underline{\gamma} = \underline{\pi}_{11} - \underline{\Pi}_{12}\beta$ and

$$(2.4) \quad \underline{\pi}_{21} = \underline{\Pi}_{22}\beta \quad ,$$

where

$$(2.5) \quad \underline{\Pi} = \begin{pmatrix} \underline{\pi}_1 & \underline{\Pi}_2 \\ \underline{\pi}_{21} & \underline{\Pi}_{22} \end{pmatrix} = \begin{pmatrix} \underline{\pi}_{11} & \underline{\Pi}_{12} \\ \underline{\pi}_{21} & \underline{\Pi}_{22} \end{pmatrix}$$

is partitioned into K_1 and K_2 rows, and into 1 and G_1 columns, respectively.

Assumption 3: The submatrix $\begin{pmatrix} \underline{\pi}_{21} & \underline{\Pi}_{22} \end{pmatrix}$ is of rank G_1 and $\underline{\Pi}_{22}$ is also of rank G_1 .

This assumption implies that

$$(2.6) \quad L = K_2 - G_1 \geq 0 \quad ,$$

where L is the degree of overidentification of the structural equation. The components of \underline{u} are independently normally distributed with mean 0 and variances

$$(2.7) \quad \sigma^2 = \omega_{11} - 2\beta' \omega_{21} + \beta' \Omega_{22} \beta .$$

Let \underline{p}_{21} and \underline{p}_{22} be the least squares estimator for π_{21} and π_{22} , and

$$(2.8) \quad \underline{G} = \underline{Y}' (\underline{\bar{P}}_{\underline{Z}_1} - \underline{\bar{P}}_{\underline{Z}}) \underline{Y} = \begin{pmatrix} \underline{p}'_{21} \\ \underline{p}'_{22} \end{pmatrix} \underline{A}_{22.1} (\underline{p}_{21} \underline{p}_{22})$$

$$= \begin{pmatrix} \underline{g}_{11} & \underline{g}_{12} \\ \underline{g}_{21} & \underline{g}_{22} \end{pmatrix} ,$$

where $\underline{A}_{22.1} = \underline{Z}'_2 \underline{Z}_2 - \underline{Z}'_2 \underline{Z}_1 (\underline{Z}'_1 \underline{Z}_1)^{-1} \underline{Z}'_1 \underline{Z}_2$. And also let

$$(2.9) \quad \underline{C} = \underline{Y}' \underline{\bar{P}}_{\underline{Z}} \underline{Y} = \begin{pmatrix} \underline{c}_{11} & \underline{c}_{12} \\ \underline{c}_{21} & \underline{c}_{22} \end{pmatrix} ,$$

where, for any matrix \underline{S} , $\underline{P}_{\underline{S}} = \underline{I} - \underline{S}(\underline{S}'\underline{S})^{-1}\underline{S}'$ is the projection onto the space orthogonal to the column vectors of \underline{S} . The limited information maximum likelihood (LIML) estimator of $\underline{\beta}$ is

$$(2.10) \quad \hat{\underline{\beta}}_{LI} = - \frac{\underline{b}_2}{\underline{b}_1} ,$$

where $\underline{b}' = (\underline{b}_1 \underline{b}_2')$ is the characteristic vector corresponding to the smallest characteristic root, λ_{\min} , of

$$(2.11) \quad |G - \lambda C| = 0 \quad .$$

For convenience we define the R-class estimator which includes the (fixed) k-class estimator and the LIML estimator:

$$(2.12) \quad \hat{\beta}_R = (G_{22} - \lambda^* C_{22})^{-1} (g_{21} - \lambda^* c_{21}) \quad ,$$

and

$$(2.13) \quad \lambda^* = a \lambda_{\min} + b \quad ,$$

where a and b are constants. Then it is clear that this estimator is identical to the LIML estimator for $a = 1$ and $b = 0$, and is identical to the (fixed) k-class estimator for $a = 0$ and $b = k - 1$. In particular, it is equal to the TSLS estimator for $a = b = 0$, and is equal to the OLS estimator for $a = 0$ and $b = -.1$. It is also identical to a modification of the LIML estimator by Fuller [1977] for $a = 1$ and $b = -c/(T - K)$ where c is some constant.

The estimator of the coefficients of included exogenous variables $\underline{\gamma}$ in any method discussed here is

$$(2.14) \quad \hat{\underline{\gamma}} = (Z_1' Z_1)^{-1} Z_1' (y_1 - Y_2 \hat{\beta})$$

where $\hat{\beta}$ is an estimator of β .

The matrix C has a central Wishart distribution with q degrees of freedom and covariance matrix Ω . And the matrix G has a non-central Wishart distribution with K_2 degrees of freedom, covariance

matrix $\underline{\Omega}$, and the noncentrality matrix:

$$(2.15) \quad \underline{\theta} = \begin{pmatrix} \underline{\pi}'_{21} \\ \underline{\pi}'_{22} \end{pmatrix} A_{22.1} (\underline{\pi}_{21} \underline{\Pi}_{22}) = \begin{pmatrix} \underline{\beta}' \\ \underline{I}_{G_1} \end{pmatrix} \underline{\Pi}'_{22} A_{22.1} \underline{\Pi}_{22} (\underline{\beta} \underline{I}_{G_1}) .$$

The exact distributions and asymptotic expansions of distributions of estimators follow only from the distributions of two independent random matrices \underline{C} and \underline{G} . In large sample asymptotic theory we make the following assumption.

Assumption 4: $\underline{\theta}_{22} = T^{-1} \underline{\Pi}'_{22} A_{22.1} \underline{\Pi}_{22}$ is bounded, i.e.,

$$\underline{\theta} = O(T).$$

We define the parameters which appear in the approximate distributions of the R-estimator:

$$(2.16) \quad \underline{F} = \underline{\theta}^{-1/2} \underline{\Omega}^{1/2} \underline{\Gamma}^{-1/2} ,$$

and

$$(2.17) \quad \underline{f} = -(1/\sqrt{c}) \underline{\theta}^{-1/2} \underline{\Omega}^{1/2} \underline{\alpha} ,$$

where

$$(2.18) \quad \underline{\alpha} = (1/\sqrt{\omega_{11.2}}) \underline{\Omega}^{-1/2} (\underline{\Omega}_{22} \underline{\beta} - \omega_{21}) ,$$

$$(2.19) \quad \omega_{11.2} = \omega_{11} - \omega_{12} \underline{\Omega}_{22}^{-1} \omega_{21} ,$$

$$(2.20) \quad c = 1 + \underline{\alpha}' \underline{\alpha} = \sigma^2 / \omega_{11.2} ,$$

and

$$(2.21) \quad \underline{\Gamma} = \underline{I}_{G_1} + c \underline{\alpha} \underline{\alpha}' .$$

See the discussion of the above parameters in Anderson [1974] and Fujikoski et al. [1979]. Some authors expand the distributions of estimators in terms of the following noncentrality parameter increasing:

$$(2.22) \quad \mu^2 = \text{tr} \Omega_{22}^{-1/2} (I + \alpha\alpha') \Pi_{22}' A_{22.1} \Pi_{22} \Omega_{22}^{-1/2},$$

which is the sum of nonzero characteristic roots of the population equation $|\Theta - \lambda\Omega| = 0$. It is seen that $\mu^2 = O(T)$ under Assumption 4.

Finally, it is convenient to derive some properties of estimators in terms of the standardized estimator:

$$(2.23) \quad \hat{\underline{e}} = \begin{pmatrix} \hat{\underline{e}}_{\beta} \\ \hat{\underline{e}}_{\gamma} \end{pmatrix} = R \begin{pmatrix} \hat{\underline{\beta}} - \underline{\beta} \\ \hat{\underline{\gamma}} - \underline{\gamma} \end{pmatrix},$$

where

$$(2.24) \quad R = \frac{1}{\sigma} \begin{pmatrix} (\Pi_{22}' A_{22.1} \Pi_{22})^{1/2} & 0' \\ (Z_1' Z_1)^{-1} Z_1' Z_2 \Pi_2 & (Z_1' Z_1)^{1/2} \end{pmatrix},$$

and $\hat{\underline{e}}$ is divided into the first G_1 and the last K_1 elements. We shall denote $\hat{\underline{e}}$ with the LIML estimator, for example, as $\hat{\underline{e}}_{LI}$.

3. Statement of Results

An approximate distribution and the first two moments of the standardized R-estimator in the large sample asymptotics are given in the following.

Lemma 1: An asymptotic expansion of the density function of $\hat{e}_{\underline{R}}$ is given by

$$\begin{aligned}
 (3.1) \quad f_{\underline{R}}(\underline{\xi}) = & \phi(\underline{\xi}) \left(\left[1 + \frac{1}{\sqrt{T}} (\underline{f}' \underline{\xi}_1) (G_1 + 1 + K_1 + m - \underline{\xi}' \underline{\xi}) \right. \right. \\
 & + \frac{1}{2T} \{ (2m - L) \text{tr} \underline{F} \underline{F}' + [2m - 2L(1 - a)^2 - m^2] \underline{f}' \underline{f} \\
 & - \underline{\xi}' \underline{F} \underline{F}' \underline{\xi}_1 (G_1 + K_1 + 2 - L + 2m - \underline{\xi}' \underline{\xi}) \\
 & + (\underline{f}' \underline{\xi}_1)^2 [(G_1 + 1 + K_1 + m - \underline{\xi}' \underline{\xi})^2 \\
 & \left. \left. + G_1 + 1 + K_1 + 2L(1 - a)^2 - 3\underline{\xi}' \underline{\xi}] \right\} \right) + o(T^{-3/2}),
 \end{aligned}$$

where $\phi(\underline{\xi})$ is the $G_1 + K_1$ dimensional standard normal density,

$\underline{\xi} = (\underline{\xi}' \underline{\xi})'$ is a $1 \times (G_1 + K_1)$ vector and $m = (1 - a)L + b$.

Then the moments of $\hat{e}_{\underline{R}}$ based on the density given (3.1) above to terms of order T^{-1} are

$$(3.2) \quad AM(\hat{e}_{\underline{R}}) = \frac{1}{\sqrt{T}} (m - 1) \begin{pmatrix} \underline{f} \\ 0 \end{pmatrix},$$

and

$$(3.3) \quad AM(\hat{\underline{e}}_{\underline{R}} \quad \hat{\underline{e}}'_{\underline{R}}) = \underline{I}_{\underline{G}_1} + \underline{K}_1 + \frac{1}{T} [(3 - 2m)\underline{f}'\underline{f} + \text{tr}\underline{FF}']\underline{I}_{\underline{G}_1} + \underline{K}_1 \\ + \frac{1}{T} \begin{pmatrix} [6 - 6m + m^2 + 2L(1 - a)^2]\underline{ff}' + [L + 2 - 2m]\underline{FF}' & \underline{0} \\ \underline{0} & \underline{0} \end{pmatrix},$$

where $AM(\cdot)$ stands for expectation with respect to (3.1).

The proof will be given in the Appendix. We note that the results of this Lemma are identical to the expressions due to Nagar [1959], Kadane [1971], and Fujikoshi et al. [1979] in the case of LIML and TSLS estimators. From (3.2), the asymptotic mean-unbiasedness of the estimator up to order T^{-1} requires $m = 1$, i.e., $(1 - a)L + b = 1$. We define a class of estimators D_1 which satisfies this asymptotic consistency condition in $(\hat{\underline{\beta}}_{\underline{R}} \quad \hat{\underline{\gamma}}_{\underline{R}})$. Then we have the following theorem.

Theorem 1:

(i) For any symmetric convex set S about origin,

$$(3.4) \quad \lim_{T \rightarrow \infty} T[\Pr\{\hat{e}_F \in S\} - \Pr\{\hat{e} \in S\}] \geq 0,$$

where \hat{e} is any estimator in the class D_1 .

(ii) Let $L(\cdot)$ be bounded negative-unimodal loss function, which is symmetric around origin. Then,

$$(3.5) \quad \lim_{T \rightarrow \infty} T[E\{L(\hat{e}_F)\} - E\{L(\hat{e})\}] \leq 0.$$

The proof will be given in the Appendix. It is important to note that Rothenberg [1978] suggested that among the class D_1 the F estimator minimizes the Asymptotic Mean Squared Error (AMSE), which is defined by the Mean Squared Error based on the asymptotic expansion of distribution. This proposition can be confirmed easily from (3.3). The first part of the theorem means that the F estimator is optimal among the class D_1 in the sense of probability concentration. It includes the sets $S_1 = \{\max_i |\hat{e}_i| \leq w, i=1, \dots, G_1+K_1\}$ and $S_2 = \{|\hat{e}_i| \leq d_i, i=1, \dots, G_1+K_1\}$, where \hat{e}_i is the i -th component of \hat{e} and d_i 's are some constants.

Now, turning to the asymptotic median-unbiasedness, we notice from Lemma 1 that

$$(3.6) \quad \Pr\{\hat{e}_i \leq \xi_i\} = \Phi(\xi_i) + \frac{f_i}{\sqrt{T}} \{m + \frac{2}{i}\} \phi(\xi_i) + O(T^{-1}),$$

where ξ_i and f_i are the i -th component of the vectors

ξ and f , respectively. Therefore, $m = 0$ is required for the coordinate-wise median-unbiasedness to the term of $O(T^{-1})$. We define a class of estimators D_2 which satisfies this median-unbiasedness condition. Then we have the next theorem.

Theorem 2:

(i) For any symmetric convex set S about origin,

$$(3.7) \quad \lim_{T \rightarrow \infty} T[\Pr\{\hat{e}_{\text{LIML}} \in S\} - \Pr\{\hat{e} \in S\}] \geq 0,$$

where \hat{e} is any estimator in the class D_2 .

(ii) Let $L(\cdot)$ be the loss function satisfying the conditions in the second part of Theorem 1. Then,

$$(3.8) \quad \lim_{T \rightarrow \infty} T[E\{L(\hat{e}_{\text{LIML}})\} - E\{L(\hat{e})\}] \leq 0.$$

The proof of Theorem 2 is similar to that of Theorem 1, as so we omit it. Theorem 2 implies that the LIML estimator has a third-order optimum property in the sense of median-unbiasedness. We also note that the LIML estimator minimizes the AMSE among the class D_2 , which can be easily shown from (3.3). It is important to note that the results in Theorem 1 and Theorem 2 do not contradict the results by Fuller [1979]. Further discussions on these points are given in Takeuchi and Morimune [1979].

Recently several modifications of single-equation methods have been proposed by several econometricians. Among them Takeuchi [1978], and Takeuchi and Morimune [1979] proposed a modified LIML estimator (the T estimator):

$$(3.9) \quad \hat{\beta}_T = \hat{\beta}_{LI} - \frac{1}{T} G_{22}^{-1} [C_{22} \hat{\beta}_{LI} - C_{21}] ,$$

Then we can prove the next lemma and a corollary of Theorem 1.

Lemma 2: The T estimator is asymptotically equivalent to the F estimator up to order $T^{-3/2}$.

Corollary 1: Both the F estimator and the T estimator are third order efficient in the class D_1 .

Another class of modification of single-equation estimation method proposed is the combined estimator: the linear combination of the OLS, TSLS, and LIML estimators, for instance, the M estimator and the S estimator. We define the generalized combined estimator by

$$(3.10) \quad \hat{\beta}_c = c_1 \hat{\beta}_{LIML} + c_2 \hat{\beta}_{k_1} + c_3 \hat{\beta}_{k_2} ,$$

where $c_1 = 1 - c_2 - c_3$, c_2 and c_3 are real numbers, $\hat{\beta}_{k_1}$ and $\hat{\beta}_{k_2}$ are k-class estimators with fixed k_1 and k_2 . It is identical to the S estimator for $c_2 = 1 + (L-1)/q$, $c_3 = 1(L-1)/q$, $k_1 = 1$, and $k_2 = 0$. Also it is identical to the M estimator for $c_2 = 1/L$, $c_3 = 0$, and $k_1 = 1$. Obviously, it is reduced to the (fixed) k-class estimator when $c_2 = 1$, $c_3 = 0$, and $k_1 = k$.

An approximate distribution and the first two moments of some combined estimators in the large sample asymptotics are given in the following. (The proof is similar to that of Lemma 1 and so it is omitted.)

Lemma 3: An asymptotic expansion of the density of the combined estimator is given by

$$(3.11) \quad f(\underline{\xi}) = \phi(\underline{\xi}) \left\{ \left[1 + \frac{1}{\sqrt{T}} \left(\underline{f}' \underline{\xi}_1 \right) \left(G_1 + K_1 + \frac{5-d_1}{2} \right) - \underline{\xi}'_1 \underline{\xi}_1 \right] \right. \\ \left. + \frac{1}{2T} \{ (2-d_3) \text{tr} \underline{F} \underline{F}' + (2d_1-d_2) \underline{f}' \underline{f} \right. \\ \left. - \underline{\xi}'_1 \underline{F} \underline{F}' \underline{\xi}_1 \left(G_1 + K_1 + 4-d_3 - \underline{\xi}'_1 \underline{\xi}_1 \right) \right. \\ \left. + (\underline{f}' \underline{\xi}_1)^2 \left[\left(G_1 + K_1 + 1 - \underline{\xi}'_1 \underline{\xi}_1 \right)^2 + (4-d_1) \left(G_1 + K_1 + 1 - \underline{\xi}'_1 \underline{\xi}_1 \right) \right. \right. \\ \left. \left. - 2 \underline{\xi}'_1 \underline{\xi}_1 - 3(d_1-1) + d_2 \right] \right\} + o(T^{-3/2}),$$

where $\phi(\underline{\xi})$ is the $G_1 + K_1$ dimensional standard normal density, $\underline{\xi}' = (\underline{\xi}'_1 \underline{\xi}'_2)$ is a $1 \times (G_1 + K_1)$ vector and $d_1=1$, $d_3=L$ (for the M, N, and S estimators), $d_2=1+2/L$ (for the M estimator) and $d_2 = 1+2L$ (for the N and S estimators). Then the first two moments based on (3.11) to the term T^{-1} are given by

$$(3.12) \quad AM(\hat{\underline{e}}_{\underline{c}}) = \frac{1}{\sqrt{T}} (d_1-1) \begin{pmatrix} \underline{f} \\ \underline{0} \end{pmatrix},$$

$$(3.13) \quad AM(\hat{\underline{e}}_{\underline{c}} \hat{\underline{e}}'_{\underline{c}}) = I_{G_1+K_1} + \frac{1}{T} [d_1 \underline{f}' \underline{f} + \text{tr} \underline{F} \underline{F}'] I_{G_1+K_1} \\ + \frac{1}{T} \begin{pmatrix} d_2 \underline{f}' \underline{f} + d_3 \underline{F} \underline{F}' \underline{0} \\ \underline{0} \quad \underline{0} \end{pmatrix},$$

where $AM(\cdot)$ stands for the expectation based on the approximate density (3.11).

Then we have some interesting relationships among several modified estimators proposed so far, which are summarized in the following.

Theorem 3:

- (i) The S estimator is asymptotically equivalent to the N estimator up to order $T^{-3/2}$.
- (ii) The F estimator dominates the M, N, S estimators uniformly up to order T^{-1} , provided $L \geq 1$. Further, the M estimator dominates the N and S estimators uniformly up to order T^{-1} , provided $L \geq 1$.

We note that the above uniform domination of the F estimator over the M, N and S estimators to order T^{-1} also holds in terms of the AMSE. The uniform domination of the F estimator over the N and S estimators can be regarded as a special case of Theorem 1.

4. Some Remarks on Asymptotic Theories

The comparison of estimators in this article has been made in terms of the approximate distributions based on the asymptotic expansion of distribution in large sample theory. In this asymptotic theory the sample size increases and the effect of exogenous variables increases along with the sample size (Assumption 4). Besides this asymptotics, several asymptotic theories have been proposed by some econometricians. Kadane [1971] initiated the small disturbance approach: the parameter sequence in which the variance of disturbances goes to zero while the sample size is fixed. Anderson [1977] clarified the meaning of Kadane's method; he showed that the small disturbance asymptotics is equivalent to the parameter sequence when the noncentrality parameter (2.22) increases and the sample size is constant. It is important to note that the uniform domination such as Theorems 1-3 in Section 3 is not obtainable in this situation. (It will be explored in another article.)

An asymptotic theory in large econometric models, called the large- K_2 asymptotic theory, was introduced by Kunitomo [1980]. In his asymptotics, the number of excluded exogenous variables K_2 increases along with the sample size. The higher order efficiency of the LIML estimator when the model is large is investigated by Kunitomo [1981].

Another important approach in econometrics in the past is exact finite sample theory. For example, Sawa [1973b] compared the exact bias and the MSE of the OLS, TSLS, and S estimators when $G_1=1$ (the case of two endogenous variables.) However, the resulting analytical expressions are too complicated to permit drawing any meaningful conclusion. In this respect, this approach may have some limitations for practical purposes, although the numerical evaluation of the moments and distribution of estimator gives some valuable information.

Appendix

Proof of Lemma 1: The method of our derivation is a slight modification of that in Fujikoshi et al.[1979]. Thus, we will refer to their results extensively.

Using (5.2) and (5.3) in Fujikoshi et al.[1979], we have

$$(A.1) \quad \begin{pmatrix} \underline{f} \\ \underline{F} \end{pmatrix} [\underline{G}^* - \lambda_0 \underline{C}^*] \begin{pmatrix} \underline{f} \\ \underline{F} \end{pmatrix}' \hat{\underline{e}}_{\underline{\beta}} = \sqrt{T} \begin{pmatrix} \underline{f} \\ \underline{F} \end{pmatrix} (\underline{G}^* - \lambda_0 \underline{C}^*) \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

Then, equating the terms of the same order on both sides, we find

$$(A.2) \quad \hat{\underline{e}}_{\underline{\beta}} = \underline{e}_{\underline{\beta}}^{(0)} + \frac{1}{\sqrt{T}} \underline{e}_{\underline{\beta}}^{(1)} + \frac{1}{T} \underline{e}_{\underline{\beta}}^{(2)} + o_p(T^{-3/2}),$$

where

$$(A.3) \quad \underline{e}_{\underline{\beta}}^{(0)} = \underline{x}_{\underline{1}},$$

$$(A.4) \quad \underline{e}_{\underline{\beta}}^{(1)} = (\underline{f}\underline{x}'_2 + \underline{F}\underline{X}'_2)\underline{x}_2 - (\underline{f}\underline{x}'_1 + \underline{F}\underline{X}'_1)\underline{x}_1 - (\underline{a}\underline{x}'_2 - b)\underline{f},$$

$$\begin{aligned}
(A.5) \quad \underline{e}_{\beta}^{(2)} &= (\underline{S}'_1{}^2 - \underline{S}_2 \underline{S}'_2) \underline{x}_1 - (\underline{S}_1 \underline{S}_2 + \underline{S}'_1 \underline{S}'_2) \underline{x}_2 \\
&+ 2a \underline{f} \underline{x}'_1 \underline{F} \underline{X}'_2 \underline{x}_2 + \underline{f} \underline{x}'_1 \underline{x}_2 \underline{w}_{11} \\
&+ (a \underline{x}'_2 \underline{x}_2 - b) \{ (\underline{f} \underline{f}' + \underline{F} \underline{F}') \underline{x}_1 - \underline{f} \underline{w}_{11} - \underline{F} \underline{w}_{21} + (\underline{S}_1 + \underline{S}'_1) \underline{f} \} ,
\end{aligned}$$

and

$$(A.6) \quad \underline{S}_1 = \underline{f} \underline{x}'_1 + \underline{F} \underline{X}'_1, \quad \underline{S}_2 = \underline{f} \underline{x}'_2 + \underline{F} \underline{X}'_2 .$$

Then we note that the characteristic function of $\hat{\underline{e}}_R$ can be expressed as (4.9) in Fujikoshi et al. [1979] with the following conditional expectations:

$$(A.7) \quad \underline{\&}(e^{(1)} | e^{(0)}) = \begin{pmatrix} -(\underline{f}' \underline{x}_1) \underline{x}_1 + \underline{m} \underline{f} \\ -(\underline{f}' \underline{x}_1) \underline{u}^* \end{pmatrix},$$

$$(A.8) \quad \underline{\&}(e^{(2)} | e^{(0)}) = \begin{pmatrix} (\underline{f}' \underline{x}_1)^2 \underline{x}_1 + (1 - \underline{m}) \underline{F} \underline{F}' \underline{x}_1 - 2\underline{m} (\underline{f}' \underline{x}_1) \underline{f} - \underline{m} (\underline{f}' \underline{f}) \underline{x}_1 \\ -[\underline{m} (\underline{f}' \underline{f}) - (\underline{f}' \underline{x}_1)^2] \underline{u}^* \end{pmatrix},$$

$$(A.9) \quad \underline{\&}(e_{\beta}^{(1)} e_{\beta}^{(1)} | e^{(0)}) = \underline{L} \underline{F} \underline{F}' + (\underline{f}' \underline{x}_1)^2 \underline{x}_1 \underline{x}'_1 + (\underline{x}'_1 \underline{F} \underline{F}' \underline{x}_1) \underline{I}_{G_1}$$

$$\begin{aligned}
& + [m^2 + 2L(1 - a)^2] \underline{ff}' - m(\underline{f}'\underline{x}_1)(\underline{fx}'_1 + \underline{x}_1\underline{f}') \quad , \\
(A.10) \quad & \& (e_{\underline{r}}^{(1)}, e_{\underline{\beta}}^{(1)} | e_{\underline{r}}^{(0)}) = -m(\underline{f}'\underline{x}_1)\underline{fu}^* + (\underline{f}'\underline{x}_1)^2 \underline{u}^* \underline{x}'_1 \quad .
\end{aligned}$$

The Fourier inverse transform of each term can be calculated just as in the Appendix in Fujikoshi et al. [1979]. After some simplification, we have (3.1). Then the first and the second moments of $\hat{e}_{\underline{R}}$ based on the asymptotic expansions for its distribution are easily computed by direct calculation or by using (A.2) ~ (A.6). We note that the results (A.7) and (A.10) are identical to the expressions in Fujikoshi et al. [1979] when $m = 0$. The validity of formal expansions is discussed in Fujikoski et al. [1979], Phillips [1977] and Sargan [1975]. (QED)

Proof of Theorem 1:

Putting $m = 1$ in (3.3), we find

$$\begin{aligned}
(A.11) \quad \text{AM} (\hat{e}_{\underline{R}}, \hat{e}'_{\underline{R}}) &= \underline{I}_{G_1} + \underline{K}_1 + \frac{1}{T} (\underline{f}'\underline{f} + \text{tr}\underline{FF}') \underline{I}_{G_1} + \underline{K}_1 \\
&+ \frac{1}{T} \begin{pmatrix} [1 + 2L(1 - a)^2] \underline{ff}' + \underline{LFF}' & \underline{0} \\ \underline{0} & \underline{0} \end{pmatrix} .
\end{aligned}$$

Thus it is seen that the AMSE is minimized at $a = 1$. From (3.1) we have

$$(A.12) \quad f_{\underline{a}}(\underline{\xi}) - f_{\underline{a}'}(\underline{\xi}) = \frac{L}{T} \phi(\underline{\xi}) [(1-a)^2 - (1-a')^2] [(f'_{\underline{1}})^2 - f'_{\underline{1}}] \\ + o_p(T^{-3/2}) .$$

Then

$$(A.13) \quad \Pr\{\hat{e}_{\underline{a}} \in S\} - \Pr\{\hat{e}_{\underline{a}'} \in S\} \\ = \int_S \dots \int [f_{\underline{a}}(\underline{\xi}) - f_{\underline{a}'}(\underline{\xi})] d\underline{\xi} \\ = \frac{L}{T} [(1-a)^2 - (1-a')^2] f'_{\underline{1}} \left[\int_S \dots \int (\underline{\xi}'_{\underline{1}} - \underline{1}) \phi(\underline{\xi}) \right]_{\underline{1}} + o(T^{-3/2}),$$

where S is any convex set, which is symmetric around origin. Since the integration on the right hand side of (A.13) is non-positive definite (See Lemma 5.8 in Pfanzagl and Wefelmeyer [1978] for instance), the theorem is established. (QED)

REFERENCES

- Akahira, Masafumi and Kei Takeuchi. [1979]. The Concept of Asymptotic Efficiency and Higher Order Asymptotic Efficiency in Statistical Estimation Theory, Lecture Notes.
- Anderson, T.W. [1974]. "An Asymptotic Expansion of the Distribution of the Limited Information Maximum Likelihood Estimate of a Coefficient in a Simultaneous Equation System," Journal of the American Statistical Association, 69, 565-572.
- Anderson, T.W. [1977]. "Asymptotic Expansions of the Distributions of Estimators in Simultaneous Equations for Alternative Parameter Sequences," Econometrica, 45, 509-518.
- Fujikoshi, Y., K. Morimune, N. Kunitomo and M. Taniguchi. [1979]. "Asymptotic Expansions of the Distributions of the Estimates of Coefficients in a Simultaneous Equation System," Technical Report No. 287, Economics Series, Institute for Mathematical Studies in the Social Sciences (IMSSS), Stanford University.
- Fuller, W.A. [1977]. "Some Properties of a Modification of the Limited Information Estimator," Econometrica, 45, 939-953.
- Kadane, Joseph B. [1971]. "Comparison of k-class Estimators When the Disturbances Are Small," Econometrica, 65, 723-737.
- Kunitomo, Naoto. [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.
- Kunitomo, Naoto. [1981]. "Asymptotic Efficiency and Higher Order Efficiency of the Limited Information Maximum Likelihood Estimator in Large Econometric Models," Unpublished Ph.D. Dissertation, Stanford University.
- Morimune, Kimio. [1978]. "Improving the Limited Information Maximum Likelihood Estimator When the Disturbances Are Small," Journal of the American Statistical Association, 73, 867-871.
- Morimune, Kimio and Naoto Kunitomo. [1980]. "Improving the Maximum Likelihood Estimate in Linear Functional Relationships for Alternative Parameter Sequences," Journal of the American Statistical Association, 75, 230-237.

- Nagar, A.L. [1959]. "The Bias and Moment Matrix of the General of the General k-Class Estimators of the Parameters in Simultaneous Equations," Econometrica, 27, 575-595.
- Pfanzagl, J. and W. Wefelmeyer. [1978]. "A Third Order Optimum Property of the Maximum Likelihood Estimator," Journal of Multivariate Analysis, 8, 1-29.
- Phillips, P.C.B. [1977], "A General Theorem in the Theory of Asymptotic Expansions as Approximations to Finite Sample Distributions of Econometric Estimators," Econometrica, 45, 1517-1534.
- Rothenberg, T. [1978]. "Second Order Efficiency of Estimators and Tests in Simultaneous Equations Models," Notes for Joint Berkeley/Stanford Seminar, November 30.
- Sargan, John D. [1975], "Gram-Charlier Approximation Applied t-ratios of k-Class Estimators," Econometrica, 43, 327-346.
- Sawa, Takamitsu. [1973a]. "Almost Unbiased Estimators in Simultaneous Equations System," International Economic Review, 14, 97-106.
- Sawa, Takamitsu. [1973b], "The Mean Square Error for a Combined Estimator and Numerical Comparison with the TSLS Estimator," Journal of Econometrics, 1, 115-132.
- Takeuchi, Kei. [1978]. "Asymptotic Higher Order Efficiency of the ML Estimators of Parameters in Linear Simultaneous Equations," Paper presented at the Kyoto Econometrics Seminar Meeting, Kyoto University.
- Takeuchi, Kei and K. Morimune. [1979]. "Asymptotic Completeness of the Extended Maximum Likelihood Estimators in a Simultaneous Equation System," unpublished manuscript.