Sankhyā : The Indian Journal of Statistics
1983, Volume 45, Series B, Pt. 3, pp. 398–412.

CONVERGENCE OF THE MOMENTS OF THE MODIFIED K-CLASS ESTIMATORS*

By P. A. V. B. SWAMY, J. S. MEHTA and N. S. IYENGAR
Federal Reserve Board, Temple University
and
Indian Statistical Institute

SUMMARY. In this paper it is shown that a modified k-class estimator with $k < 1$ is preferable to the modified LIML estimator in terms of convergence in probability or in law to normal. The Central Limit Theorem for the density of this modified k-class estimator can be strengthened considerably, since its higher order moments exist in small samples.

1. Introduction

Consider the k-class estimators of the coefficients of a structural equation as defined by Theil (1971, p. 504). For these estimators k can be either fixed or random. The much heed paid by the econometricians to the asymptotic properties—such as weak consistency and covergence in law to normal—of these estimators justifies an explicit reminder of the work by Anderson and Sawa (1979) who were the first to prove that these asymptotic properties of the "fixed" k-class estimator with $k = 1$ or the two stage least squares (2SLS) estimator are not necessarily expected to be relevant to the cases that appear in practice. The further work done by Anderson, Kunitomo and Sawa (1982) shows that the limited information maximum likelihood (LIML) estimator which is a "random" k-class estimator with $k > 1$ (almost surely) is essentially median unbiased and its distribution approaches normality much faster than that of the 2SLS estimator with finite lower order moments, due primarily to the bias of the latter.[1] More importantly, these authors demonstrate that the 2SLS estimator with finite lower order moments in generally very biased and, as a result, is likely to substantially underestimate the parameter. In an earlier paper, Anderson (1976, p. 10) uses the approximate distributions

---

* The views expressed herein are solely those of the authors and do not necessarily represent the views of the organisations to which they belong.

[1] In this comparison it is important to remember that the sample sizes required for the existence of the LIML estimator are larger than those required for the existence of the fixed k-class estimators, see Swamy (1980, p. 172).

to prove that under certain conditions the cumulative distribution function of the absolute error of the LIML estimator is everywhere above that of the 2SLS estimator.[2] Zaman (1981, p. 297) obtains similar results for a simple case and concludes that if the signs of errors do not have special significance, if the loss function is bounded, if the probability of an estimator falling in an interval about the true value of the parameter is the primary criterion, and if the sample size is sufficiently large, then the maximum likelihood estimator and its truncations perform reasonably well. This is consistent with Anderson and Sawa's (1979, p. 175) [or Anderson, Kunitomo and Sawa's, (1982)] conclusion that the LIML estimator would be certainly preferred to the 2SLS estimator in terms of the asymptotic normality. But the important problem is that the size of the currently available sample may not be large enough to guarantee our reliance on the asymptotic normality of LIML.

In small samples, the LIML estimator may not be preferable to the fixed k-class estimators because, as shown by Hatanaka (1973), Mariano and Sawa (1972) and others, there are situations in which the fixed k-class estimators with $k < 1$ possess finite second-order moments whereas the LIML estimator does not possess finite moments of any positive integer order. In these situations the former dominates the latter under squared error loss.[3] Another implication of the results of these authors is that the necessary and sufficient conditions given in Lukacs (1975, p. 43, Theorem 2.3.4) for the convergence of a sequence of moments of a k-class estimator to the corresponding moment of the limiting random variable are not always satisfied. We can also appeal to a theorem in Lukacs (1975, p. 65, Theorem 3.1.3) to draw the conclusion that the convergence in probability of a k-class estimator is in general incompatible with the existence of a norm.

However, if we make the stronger assumption that an estimator converges in the r-th mean (instead of in probability or in law) to a random variable, then a sequence of the p-th moments of the estimator converges to the p-th

[2] See also Fujikoshi, Morimune, Kunitomo, and Taniguchi (1982, p. 197).

[3] Zaman (1981) observes that the non-existence of sampling moments may not be a serious problem if a bounded loss function is appropriate and the sample size is sufficiently large. He goes on to show that with a bounded loss function, asymptotic efficiency is a desirable property in large samples and that the use of a quadratic loss function can conflict with the objective of obtaining an estimator that has high probability of falling in an interval about the true value of the parameter for most intervals of interest. Apart from the well known difficulties of determining when a sample may be considered large in actual econometric models, the appropriateness of a bounded loss function is very hard to determine, particularly when we are unable to choose a fairly small region of the parameter space in which the true value of a parameter lies.

moment of the limiting random variable for $p \leq r$, see Lukacs (1975, 40-41). Also, convergence in the quadratic mean which is a special case of convergence in the r-th mean implies convergence in probability which in turn implies convergence in law. While there exists no implication between the concepts of almost sure convergence and convergence in the r-th mean, these modes of convergence are compatible, as shown by Lukacs (1975, p. 36). [For an excellent discussion on the direction of the implications between the modes of convergence, see Lukacs (1975, p. 37]. It is also demonstrated by Lukacs (1975, p. 69) that convergence in the sense of a distance is equivalent to convergence in the r-th mean. If a finite sample risk converges to the asymptotic risk which can happen under convergence in the r-th mean, then asymptotic efficiency is also a relevant property for unbounded loss functions, see Zaman (1981, p. 290).

The purpose of this paper is to present a simple modification of the k-class estimators and to show under certain conditions that the modified estimators converge in the r-th mean. The modified estimators have the advantage that the Nagar (1959) type approximations for their moments may be developed. These approximations are useful for obtaining a better approximation than the asymptotic approximation even for quite small samples, at least in some cases.

The model and the modified k-class estimators are given in Section 2. The principal results on the existence of the moments for these estimators are established in Section 3. It is also shown in this section that the modified "fixed" k-class estimators converge in the r-th mean under a mild restriction on k whereas the modified LIML estimator may not. Conclusions of the study are presented in Section 4.

2. The model and the modified k-class estimators.

Consider the following i-th structural equation of a G-equation model :

$$\begin{aligned} \mathbf{y}_i &= Y_i \boldsymbol{\gamma}_i + X_i \boldsymbol{\beta}_i + \mathbf{u}_i \\ &= Z_i \boldsymbol{\delta}_i + \mathbf{u}_i \qquad \dots \quad (1) \end{aligned}$$

where $\mathbf{y}_i$ is the $T \times 1$ vector of observations on a left-hand endogenous variable, Y_i is the $T \times G_i$ matrix of observations on the endogenous variables which appear in the equation, X_i is the $T \times K_i$ matrix of observations on the exogenous variables that appear in the equation, $Z_i = (Y_i \vdots X_i)$, $\mathbf{u}_i$ is the $T \times 1$ vector of disturbances, $\boldsymbol{\gamma}_i$ and $\boldsymbol{\beta}_i$ are respectively $G_i \times 1$ and $K_i \times 1$ coefficient vectors.

and $\delta_i = (\gamma_i', \beta_i')'$. In the full model there are $K-K_i$ other exogenous variables whose observations are given in the $T \times (K-K_i)$ matrix X_i^*. Let $X = (X_i \vdots X_i^*)$.

We write the reduced form equations for y_i and Y_i as

$$Y_0 = (y_i \vdots Y_i) = X_i(\pi_i \vdots \Pi_i) + X_i^*(\pi_i^* \vdots \Pi_i^*) + (v_i \vdots V_i). \qquad \ldots \quad (2)$$

To identify δ_i, we assume that rank $(\Pi_i^*) = \text{rank}(\pi_i^* \vdots \Pi_i^*) = G_i$, see Theil (1971, 491–2). This imposes the restriction $K_i^* = K-K_i > G_i < G$.

We make the following assumptions :

Assumption 1. (i) The vector u_i is distributed, independent of X, with mean zero and covariance matrix $\sigma_{ii}I$.

(ii) The rows of $(v_i \vdots V_i)$ are independent drawings from a multivariate distribution with zero mean vector and positive definite covariance matrix

$$\Omega = \begin{bmatrix} \omega_{11} & \omega_{12}' \\ \omega_{21} & \Omega_{22} \end{bmatrix}.$$

(iii) Both the vector u_i and the rows of $(v_i \vdots V_i)$ are normally distributed.

Assumption 2. (i) The $T \times K$ matrix X is "fixed" such that $K = \inf\{T : X'X \text{ is nonsingular}\} < \infty$ and $T > K$.[4] (ii) For each $j = 1, 2, \ldots, K$, $\max_{1 \leq i \leq T} |x_{ij}|$ is bounded, where x_{ij} is the ij-th element of X. (iii) The matrix $T^{-1}X'X$ converges, as $T \to \infty$, to a finite, positive definite matrix.

Define

$$M_i = I - X_i(X_i' X_i)^{-1}X_i' \text{ and } M = I - X(X'X)^{-1}X'.$$

The k-class estimator of γ_i is

$$(g_i)_k = (Y_i'QY_i)^{-1}Y_i'Qy_i, \qquad \ldots \quad (3)$$

where $Q = M_i - kM$, and that of β_i is

$$(b_i)_k = (X_i'X_i)^{-1}X_i'[y_i - Y_i(g_i)_k]. \qquad \ldots \quad (4)$$

We now consider the Mehta and Swamy (1978) type modifications of the estimators (3) and (4) where the modifications guarantee the existence of the moments at least for certain sample sizes. These modifications are

$$(g_i)_{k,\mu} = [Y_i'Q_\mu Y_i + (\mu_1 + \mu_2)I]^{-1}Y_i'Q_\mu y_i \qquad \ldots \quad (5)$$

[4] This condition is relaxed in Swamy (1980). See also Swamy, and Mehta (1983).

and

$$(b_i)_{k,\mu} = [X_i'X_i+(\hat{\beta}_1+\mu_2)I]^{-1}X_i'[y_i - Y_i(g_i)_{k,\mu}] \qquad (6)$$

respectively, where $Q_\mu = I - X_i[X_i'X_i+(\hat{\beta}_1+\mu_2)I]^{-1}X_i' - kM$, $\mu_2 > 0$ is a fixed real number and $\hat{\beta}_1$ is the solution to the equation

$$d_i'(2\hat{\beta}_1^{-1}I+\hat{\omega}_{ii}[Z_i'X(X'X)^{-1}X'Z_i]^{-1})^{-1}d_i = 1 \qquad \dots \quad (7)$$

with $d_i = [Z_i'X(X'X)^{-1}X'Z_i]^{-1}Z_i'X(X'X)^{-1}X'y_i$ and $\hat{\omega}_{ii} = y_i'My_i/T$ if $d_i'Z_i'X(X'X)^{-1}X'Z_id_i < \hat{\omega}_{ii}$ and $\hat{\beta}_1 = 0$ if $d_i'Z_i'X(X'X)^{-1}X'Z_id_i \leqslant \hat{\omega}_{ii}$. It is easy to check that the left-hand side of eq. (7) is a monotonically increasing function of $\hat{\beta}_1 \geqslant 0$ and $\Pr(\hat{\beta}_1 \geqslant 0) = 1$, see Swamy, Mehta and Rappoport (1978, p. 1145). The solution $\hat{\beta}_1$ can be found by numerical methods. Combining the estimators (5) and (6), we can write the modified k-class estimator of δ_i as

$$(d_i)_{k,\mu} = [Z_i'(I-kM)Z_i+(\hat{\beta}_1+\mu_2)I]^{-1}\, Z_i'(I-kM)\, y_i \qquad \dots \quad (8)$$

where k can be either fixed or random.

The argument of Mehta and Swamy (1978, pp. 3-4) can be utilized here to show that the estimator (8) approximates the restricted least squares estimator of δ_i subject to a boundedness restriction on the squared length of δ_i. This provides the motivation for our derivations. It should also be noted that the left-hand side of eq. (7) is a reasonable estimate of a necessary and sufficient condition given in Swamy (1980, p. 178) for a restricted estimator to be better than an unrestricted estimator. The advantages of the modifications in (5) and (6) in the context of the standard linear regression model are pointed out by Swamy and Mehta (1980), Swamy, Mehta and Thurman (1981) and Thurman, Swamy and Mehta (1981).

The modifications in (5) and (6) preserve the errors—in-variables nature of the problem which Zellner (1970) and Anderson (1976) bring out explicitly. To see this postmultiply both sides of eq. (2) by $(1 \vdots -\gamma_i')'$ and equate the resulting coefficients with those appearing in eq. (1). This gives

$$\pi_i^* = \Pi_i^*\gamma_i. \qquad \dots \quad (9)$$

If π_i^* and Π_i^* were known an estimator of γ_i is

$$(\Pi_i^{*\prime}X_i^{*\prime}M_iX_i^*\Pi_i^*)^{-1}\Pi_i^{*\prime}X_i^{*\prime}M_iX_i^*\pi_i^*. \qquad \dots \quad (10)$$

Since π_i^* and Π_i^* are unobservable, it is clear from eq. (2) that we can approximate the estimator (10) by the estimator (5) because $[Y_i'Q_\mu Y_i+(\hat{\beta}_1+\mu_2)I]$ is an estimator of $\Pi_i^{*\prime}X_i^{*\prime}M_iX_i^*\Pi_i^*$ and $Y_i'Q_\mu y_i$ is an estimator of $\Pi_i^{*\prime}X_i^{*\prime}M_iX_i^*\pi_i^*$.

Thus Zellner (1970)[7] and Anderson (1976) are able to establish a connection between the formulae defining estimators for the errors-in-variables models and simultaneous equations. As pointed out by Anderson (1976, p. 8), the estimator (5) is consistent as $T \to \infty$ and K_i^* remains fixed. This clarifies the conditions under which the estimators for the errors-in-variables models are consistent.

Before we proceed to study the properties of (8), a comparison of the estimators (5) with Fuller's (1977) modification of the 2SLS and LIML estimators of γ_i may be made. Let $\hat{f}$ be the smallest root of

$$|K_i^{*-1}Y_i'(M_i-M)Y_i-f(T-K)^{-1}Y_i'MY_i| = 0,$$

let $\alpha > 0$ be a fixed real number and let $\hat{\lambda}$ be the smallest root of

$$|K_i^{*-1}Y_0'(M_i-M)Y_0-\lambda(T-K)^{-1}Y_0'MY_0| = 0.^{8}$$

Then Fuller's (1977, p. 942) modified 2SLS estimator of γ_i is

$$\hat{\gamma}_i = \hat{H}^{-1}\hat{N} \qquad \ldots \quad (11)$$

where

$$\hat{H} = \frac{Y_i'(M_i-M)Y_i}{K_i^*} - \frac{(K_i^*-G_i-\alpha)Y_i'MY_i}{K_i^*(T-K)} \quad \text{if } \hat{f} > \frac{(K_i^*-G_i+1)}{K_i^*}$$

$$= \frac{Y_i'(M_i-M)Y_i}{K_i^*} - \left(\hat{f} - \frac{1+\alpha}{K_i^*}\right)\frac{Y_i'MY_i}{(T-K)} \quad \text{otherwise},$$

$$\hat{N} = \frac{Y_i'(M_i-M)y_i}{K_i^*} - \frac{(K_i^*-G_i-\alpha)}{K_i^*}\frac{Y_i'My_i}{(T-K)} \quad \text{if } \hat{f} > \frac{(K_i^*-G_i+1)}{K_i^*}$$

$$= \frac{Y_i'(M_i-M)y_i}{K_i^*} - \left(\hat{f} - \frac{1+\alpha}{K_i^*}\right)\frac{Y_i'My_i}{(T-K)} \quad \text{otherwise}.$$

The modified LIML estimator of γ_i introduced by Fuller (1977, p. 942) is

$$\tilde{\gamma}_{iL} = \hat{H}_L^{-1}N_L \qquad \ldots \quad (12)$$

where

$$\hat{H}_L = \frac{Y_i'(M_i-M)Y_i}{K_i^*} - \left(\hat{\lambda} - \frac{\alpha}{K_i^*}\right)\frac{Y_i'MY_i}{(T-K)}$$

and

$$\hat{N}_L = \frac{Y_i'(M_i-M)y_i}{K_i^*} - \left(\hat{\lambda} - \frac{\alpha}{K_i^*}\right)\frac{Y_i'My_i}{(T-K)}.$$

---

[8]The smallest root $\hat{\lambda}$ will not be finite unless T-rank $(X) > G_i+1$, see Swamy (1980, p. 172).

The modifications in (11) and (12), it is clear, are different from those in (5) and are designed to guarantee the existence of moments for sufficiently large samples. Since the LIML does not possess moments of any positive integer order, the motivation for the modifications in (12) is clear. There is no similar motivation for the modifications in (11) because the first two moments of the usual 2SLS estimator exist for the degree of over identification of eq. (1) greater than or equal to 2. The modified 2SLS (11), moreover, does not reduce to the usual 2SLS whenever the latter possesses finite second-order moments. In fact, it is misleading to call (11) the modified 2SLS estimator because the 2SLS estimator is a member of the fixed k-class estimator and the estimator (11) is not.

Although it turns out that the modifications in (8) also guarantee the existence of moments, they are primarily designed to give a satisfactory approximation to a restricted least squares estimator of δ_t. In the next section we compare the operating characteristics of (5) with those of (11) and (12).

3. Principal results

We first prove a unique property of the estimator (8) with fixed $k \leqslant 1$.

Lemma 1 : *Given the model* (1) *and assumptions* 1(i), 1(ii), 1(iii), 2(i) *and* 2(ii), *the modified fixed k-class estimator* (8) *possesses moments of all orders for every* $\mu_2 > 0$ *and every* $k \leqslant 1$.

Proof: Let $0 < r < \infty$. Then the r-th absolute moment of an arbitrary linear combination of the elements of $(g_t)_{k,\mu}$ is given by

$$E|l'(g_t)_{k,\mu}|^r = E|l'[Y_t'Q_\mu Y_t+(\beta_1+\mu_2)I]^{-1}Y_t'Q_\mu y_t|^r \quad \dots \quad (13)$$

where l is an arbitrary $G_t \times 1$ vector of fixed elements, not all of which are zeros.

It has been shown by Mehta and Swamy (1978, p. 11) that the matrix Q_μ has at most $2+K_t$ district roots, two of them being $(1-k)$ of multiplicity $(T-\text{rank}(X))$ and 1 of multiplicity rank $(X)-K_t$. The remaining K_t eigenvalues of Q_μ lie in the interval (0, 1) with probability 1. Therefore $Y_t'Q_\mu Y_t$ is positive semidefinite with probability 1 for $k \leqslant 1$. This shows that

$$E|l'(g_t)_{k,\mu}|^r \leqslant E|\mu_2^{-1}l'Y_t'Q_\mu y_t|^r < E|\mu_2^{-1}l'Y_t'y_t|^r|\lambda_{\max}(Q_\mu)|^r \quad \dots \quad (14)$$

where $\lambda_{\max}(Q_\mu)$ is the maximum eigenvalue of Q_μ. The expectation on the right-hand side of the second inequality sign can be easily shown to be finite because y_t and Y_t are normal variables. It remains now to show that the estimator (6) possesses moments of all orders. Defining $E|l'(b_t)_{k,\mu}|^r$ and applying the inequality in Rao (1973, p. 149, Problem 8(a)), we see that

$$E|l'(b_t)_{k,\mu}|^r \leqslant C_r(E|l'[X_t'X_t+(\beta_1+\mu_2)I]^{-1}X_t'y_t|^r + E|l'[X_t'X_t+(\beta_1+\mu_2)I]^{-1}X_t'Y_t(g_t)_{k,\mu}|^r) \quad \ldots \quad (15)$$

where $C_r = 1$ for $r \leqslant 1$ and $= 2^{r-1}$ for $r > 1$. Consequently, $E|l'(b_t)_{k,\mu}|^r$ is finite if the expectations on the right-hand side of the inequality (15) are finite. Now it can be easily seen that

$$E|l'[X_t'X_t+(\beta_1+\mu_2)I]^{-1}X_t'y_t|^r \leqslant E|\mu_2^{-1}l'X_t'y_t|^r < \infty \quad \ldots \quad (16)$$

and

$$\begin{aligned} E|l'[X_t'X_t+(\beta_1+\mu_2)I]^{-1}X_t'Y_t(g_t)_{k,\mu}|^r &\leqslant E|\mu_2^{-1}l'X_t'Y_t(g_t)_{k,\mu}|^r \\ &\leqslant E|\mu_2^{-2}l'X_t'Y_tY_t'Q_\mu y_t|^r \\ &\leqslant E|\mu_2^{-2}l'X_t'Y_tY_t'y_t|^r|\lambda_{\max}(Q_\mu)|^r < \infty. \quad \text{Q.E.D.} \quad \ldots \quad (17) \end{aligned}$$

Theorem 1 : *If the conditions of Lemma* 1 *hold, if assumption* 2(iii) *is true, if* $\mu_2 > 0$ *and if the nonstochastic* k *with* $k \leqslant 1$ *is such that*

$$p\lim \sqrt{T}(k-1) = 0 \text{ as } T \to \infty,$$

then the modified fixed k-class estimator (8) *converges to a normal variable in the r-th mean.*

Proof: By Laha and Rohatgi's (1979, p. 64) Problem 37 and our Lemma 1 above, the result is true if every arbitrary linear combination of the elements of a fixed k-class estimator of the form (8) that has $\mu_2 > 0$, $k \leqslant 1$ and $p\lim \sqrt{T}(k-1) = 0$ converges in probability to a normal variable. This linear combination can be written as

$$\begin{aligned} l'[(d_t)_{k,\mu}-\delta_t] &= l'[Z_t'(I-M)Z_t-(k-1)Z_t'MZ_t+(\beta_1+\mu_2)I]^{-1} \\ &\quad [Z_t'(I-M)u_t-(k-1)Z_t'Mu_t] \\ &\quad -(\beta_1+\mu_2)l'[Z_t'(I-M)Z_t-(k-1)Z_t'MZ_t+(\beta_1+\mu_2)I]^{-1}\delta_t \end{aligned} \quad \ldots \quad (18)$$

where l is a $(G_t+K_t)\times 1$ fixed vector, not all elements of which are zero. It can be shown that $p\lim_{T\to\infty} (\beta_1/\sqrt{T}) = 0$. This together with the method of proof given in Theil (1971, pp. 487, 498 and 505-506) implies that

$$\begin{aligned} &p\lim_{T\to\infty} [T^{-1}Z_t'(I-M)Z_t-(k-1)T^{-1}Z_t'MZ_t+T^{-1}(\beta_1+\mu_2)I]^{-1} \\ &= \lim_{T\to\infty} (T^{-1}\bar{Z}_t'\bar{Z}_t)^{-1} \end{aligned} \quad \ldots \quad (19)$$

where $\bar{Z}_i = (X\Pi_i \vdots X_i)$. Now let

$$w_i = l'\left(\lim_{T\to\infty} T^{-1}\bar{Z}_i'\bar{Z}_i\right)^{-1/2}\left(\lim_{T\to\infty} T^{-1/2}\bar{Z}_i' u_i\right).$$

Then under assumptions 1(iii) and 2 the variable w_i is distributed as normal with mean zero and variance $\sigma_{ii}l'l$ for every $T-K > 1$. Once again, the method of proof of Theil (1971, pp. 498 and 505-506) establishes that

$$\operatorname*{plim}_{T\to\infty} (l'[T^{-1}Z_i'(I-M)Z_i-(k-1)T^{-1}Z_i'MZ_i+T^{-1}(\beta_1+\mu_2)I]^{-1/2}[(\mathbf{d}_i)_{k,\mu}-\delta_i]-w_i)$$

$$= -\operatorname*{plim}_{T\to\infty} T^{-1/2}(\beta_1+\mu_2)l'[T^{-1}Z_i'(I-M)Z_i-(k-1)T^{-1}Z_i'MZ_i+T^{-1}(\beta_1+\mu_2)I]^{-1}\delta_i$$

$$+l'\Big(\operatorname*{plim}_{T\to\infty}[T^{-1}Z_i'(I-M)Z_i-(k-1)T^{-1}Z_i'MZ_i+T^{-1}(\beta_1+\mu_2)I]^{-1/2}$$

$$-\lim_{T\to\infty}(T^{-1}\bar{Z}_i'\bar{Z}_i)^{-1/2}\Big)\Big(\lim_{T\to\infty} T^{-1/2}\bar{Z}_i'u_i\Big)$$

$$+l'\Big(\operatorname*{plim}_{T\to\infty}[T^{-1}Z_i'(I-M)Z_i-(k-1)T^{-1}Z_i'MZ_i+T^{-1}(\beta_1+\mu_2)I]^{-1/2}\Big)$$

$$\times\Big(\operatorname*{plim}_{T\to\infty}[T^{-1/2}Z_i'(I-M)u_i-\sqrt{T}(k-1)T^{-1}Z_i'Mu_i]$$

$$-\lim_{T\to\infty} T^{-1/2}\bar{Z}_i'u_i\Big) = 0 \qquad \text{Q.E.D.} \qquad \dots (20)$$

We call the estimator (8) the modified LIML estimator if $k = \lambda^*$ which is the smallest root of $|Y_0'M_iY_0-\lambda Y_0'MY_0| = 0$. Let $l'(\mathbf{d}_i)_{\lambda^*,\mu}$ denote an arbitrary linear combination of the elements of the modified LIML estimator (8) with $k = \lambda^*$.

Lemma 2 : *Suppose that the model* (1) *and assumptions* 1(i), 1(ii), 1(iii), 2(i) *and* 2(ii) *hold. Then the r-th,* $r = 1, 2, \dots$, *moment of the modified LIML estimator* (8) *is bounded for all T greater than* $K+4r$ *and every* $\mu_2 > 0$.

Proof: The modified LIML estimator of γ_i can be written as

$$(g_i)_{\lambda^*,\mu} = [Y_i'Q_\mu Y_i+(\beta_1+\mu_2)I]^{-1}Y_i'Q_\mu y_i \qquad \dots (21)$$

where

$$Q_n = I - X_i[X_i'X_i + (\hat{\beta}_1+\mu_2)I]^{-1}X_i' - M + (1-\lambda^*)M.$$

Defining

$$Q_{n1} = I - X_i[X_i'X_i + (\hat{\beta}_1+\mu_2)I]^{-1}X_i' - M,$$

we have

$$\begin{aligned} l'(g_i)_{\lambda^*, n} &= l'[Y_i'Q_{n1}Y_i + (1-\lambda^*)Y_i'MY_i + (\hat{\beta}_1+\mu_2)I]^{-1} \\ &\quad [Y_i'Q_{n1}y_i + (1-\lambda^*)Y_i'My_i] \\ &= R_1 + R_2, \end{aligned} \qquad \dots \quad (22)$$

where

$$R_1 = l'[Y_i'Q_{n1}Y_i + (1-\lambda^*)Y_i'MY_i + (\hat{\beta}_1+\mu_2)I]^{-1}Y_i'Q_{n1}y_i$$

and

$$R_2 = l'[Y_i'Q_{n1}Y_i + (1-\lambda^*)Y_i'MY_i + (\hat{\beta}_1+\mu_2)I]^{-1}(1-\lambda^*)Y_i'My_i.$$

From the inequality used in (15) we find that

$$|l'(g_i)_{\lambda^*, n}|^r \leqslant C_r(E|R_1|^r + E|R_2|^r). \qquad \dots \quad (23)$$

Consequently, the r-th absolute moment of $l'(g_i)_{\lambda^*, n}$ is finite if the r-th absolute moments of R_1 and R_2 are finite. It is found by Kadiyala's (1970) method that with probability 1, $k^* = \inf_{c \neq 0} (c'Y_i'M_iY_i'c/c'Y_i'MY_ic) > \lambda^* \geqslant 1$ and the matrix $Y_i'(M_i - kM)Y_i$ is positive definite for $k < k^*$. It follows from these results that

$$\begin{aligned} E|R_1|^r &\leqslant E|l'[Y_i'Q_{n1}Y_i + (1-k^*)Y_i'MY_i + (\hat{\beta}_1+\mu_2)I]^{-1}Y_i'Q_{n1}y_i|^r \\ &\leqslant E|\mu_2^{-1}l'Y_i'y_i|^r < \infty \end{aligned} \qquad \dots \quad (24)$$

where the second inequality is based on the results that

$$\begin{aligned} &Y_i'[Q_{n1} + (1-k^*)M]Y_i \\ &= Y_i'[X_i(X_i'X_i)^{-1}X_i' - X_i(X_i'X_i + (\hat{\beta}_1+\mu_2)I)^{-1}X_i' + M_i - k^*M]Y_i \end{aligned}$$

is positive definite with probability 1 and the maximum eigenvalue of Q_{n1} is 1. By the preceding reasoning it is also true that

$$\begin{aligned} E|R_2|^r &\leqslant E|(1-\lambda^*)l'[Y_i'Q_{n1}Y_i + (1-k^*)Y_i'MY_i + (\hat{\beta}_1+\mu_2)I]^{-1}Y_i'My_i|^r \\ &\leqslant E|(1-\lambda^*)\mu_2^{-1}l'Y_i'y_i|^r \\ &\leqslant E|k^*\mu_2^{-1}l'Y_i'y_i|^r. \end{aligned} \qquad (25)$$

Since k^* is obtained by minimizing the ratio $c'Y_i'M_iY_ic/c'Y_i'MY_ic$ with respect to c, it follows that $k^* \leqslant c'Y_i'M_iY_ic/c'Y_i'MY_ic$. Therefore,

$$E|R_2|^r \leqslant E|(c'Y_i'M_iY_ic/c'Y_i'MY_ic)\mu_2^{-1}l'Y_i'y_i|^r. \quad \ldots \quad (26)$$

It now follows by application of the Cauchy-Schwarz inequality (Lukacs, 1975, p. 12) that

$$(E|R_2|^r)^2 \leqslant E\left(\frac{c'Y_i'M_iY_ic}{c'Y_i'MY_ic}\right)^{2r} E(\mu_2^{-1}l'Y_i'y_i)^{2r}. \quad \ldots \quad (27)$$

In view of assumption 1(iii), it suffices to show that the first factor on the right-hand side of the inequality sign in (27) is finite. To show this, we write

$$E\left(\frac{c'Y_i'M_iY_ic}{c'Y_i'MY_ic}\right)^{2r} = E\left[1+\frac{c'Y_i(M_i-M)Y_ic}{c'Y_i'MY_ic}\right]^{2r}$$

$$\leqslant 2^{2r-1}\left[1+E\left(\frac{c'Y_i'(M_i-M)Y_ic}{c'Y_i'MY_ic}\right)^{2r}\right]. \quad \ldots \quad (28)$$

The variables $c'Y_i'(M_i-M)Y_ic$ and $c'Y_i'MY_ic$ are independent chi-square variables with $K-K_i$ and $T-K$ degrees of freedom respectively because the matrices M and M_i-M are idempotent and $M(M_i-M) = 0$, see Theil (1971, p. 682). Hence the first factor on the right-hand side of the inequality sign in (27) is finite if $T-K > 4r$.

It remains now to establish the finiteness of the r-th absolute moment of the modified LIML estimator of $l'\beta_i$ which is

$$l'(b_i)_{\lambda^*,\mu} = l'[X_i'X_i+(\hat{\beta}_1+\mu_2)I]^{-1}X_i'[y_i-Y_i(g_i)_{\lambda^*,\mu}]. \quad \ldots \quad (29)$$

This result can be shown to hold by the use of an argument parallel to that underlying (15)–(17) and (21)–(28). Q.E.D.

Theorem 2: *Under the conditions of Lemma 2 if assumption 2(iii) is true, and if $T-K > 4p$, then for $r < p$ and every $\mu_2 > 0$, the r-th absolute moment of $l'(d_i)_{\lambda^*,\mu}$ tends to the r-th absolute moment of the limiting normal distribution of $l'(d_i)_{\lambda^*,\mu}$.*

Proof: Anderson's (1976) method of proof can be used to show that $l'(d_i)_{\lambda^*,\mu}$ converges in distribution to normal. Hence the result follows immediately from our Lemma 2 above and Chung's (1974, p. 95) Theorem 4.5.2. Q.E.D.

It should be noted that convergence in the r-th mean implies convergence of a sequence of r-th moments to the r-th moment of the limiting variable but the converse is not true unless the conditions of Chung's (1974, p. 97) Theorem 4.5.4 hold. These conditions are not satisfied by the modified LIML estimator (8). The modified fixed k-class estimators (8) with $k \leq 1$ and the modified LIML estimator (8) with $k = \lambda$ do not seem to share the same properties. The former may be preferred in terms of convergence in the r-th mean. Another disadvantage of the modified LIML estimator (8) is that the existence of its r-th moment requires a larger sample size than the existence of the r-th moment of the modified fixed k-class estimators (8) with $k \leq 1$. Consequently, when $T < K+8$ the modified LIML estimator (8), if it exists, is inadmissible under a squared error loss function. We have already pointed out in footnote 5 that the condition, $T \geq K+G_t+1$, is needed for the existence of the (modified) LIML estimator.

Both the modified 2SLS(11) and LIML(12) estimators due to Fuller (1977) and the modified LIML estimator (8) have one property in common : they do not possess the r-th moment unless the sample size is sufficiently large. Perhaps the reason is that Fuller's modified 2SLS(11), unlike the usual 2SLS, depends on the smallest root f and is strictly the random k-class type estimator. Fuller's modified 2SLS estimator, in consequence, behaves differently from the usual 2SLS estimator, that is, they do not share the same small-sample properties. This statement is based on Anderson and Sawa's (1979, p. 179) result that on the whole, the small sample properties of the 2SLS and LIML are substantially different from each other. It follows from Ghosh, Sinha and Wieand's (1980) study that Fuller's estimator (12) may not dominate (3) with fixed $k \leq 1$ even asymptotically unless (12) is used to correct itself by estimating $E[(g_t)_k - \bar{\gamma}_{tL}]$, provided this expectation exists. That is, when $E\bar{\gamma}_{tL} < \infty$, the estimator $\bar{\gamma}_{tL}+\zeta(\bar{\gamma}_{tL})$, where $\zeta(\bar{\gamma}_{tL})$ is $E[(g_t)_k - \bar{\gamma}_{tL}]$ evaluated at $\gamma_t = \bar{\gamma}_{tL}$, may asymptotically dominate (g_t) with $E(g_t)_k < \infty$, see Efron (1982, p. 350, Remark D). On the basis of Fuller's investigations, we cannot make some comparative statements between the usual 2SLS and his modified LIML(12) estimators by comparing the mean square error of the estimator (11) with that of the estimator (12). It will come as no surprise to learn that the estimator (11) is inadmissible (Fuller, 1977, p. 950) because it is not analytic. (Brown, 1975 establishes that analyticity is needed.) The usual 2SLS estimator and the estimator (8) are analytic.

We further remark that Lemmas 1 and 2 above can also be proved by establishing the conditions of Lemma B in Fuller (1977, p. 944) for the estimator (8). While we accept Lemmas A and B in Fuller (1977, p. 944), we find it convenient to use the standard inequalities and the conditions of Chung's (1974, p. 95) Theorem 4.5.2.

Part of the appeal of the modified fixed k-class estimator (8) is obvious from Theorem 1. It possesses moments of all orders for every $k \leqslant 1$, every $\mu_2 > 0$, and all $T > K$ and converges in the r-th mean to normal even when the value $(\hat{\beta}_1 + \mu_2)$ added to the diagonal elements of $Z_i'(I - kM)Z_i$ in (8) is minuscule. For this reason, we consider the modifications in (8) as minimal. In contrast, the r-th moments of Fuller's modified estimators (11) and (12) are finite for all T greater than some $T(r)$, the modifications in (11) and (12) can be substantial and it is not known whether the estimators (11) and (12) converge in the r-th mean to normal.

Feller (1966, p. 506) points out that the central limit theorem for densities can be strengthened considerably when some higher order moments exist. We cannot take advantage of this fact if some higher order moments do not exist, as is true of Fuller's modified estimators in small samples. Our preference for the modified fixed k-class estimators (8) with $k \leqslant 1$ is consistent with Zellner's (1978) remarkable finding that the MELO estimator of $\boldsymbol{\delta}_i$ is the same as the estimator (8) with fixed $k < 1$, $\mu_2 = 0$, and $\Pr(\hat{\beta}_1 = 0) = 1$ and Nagar's (1959) finding that the optimal value of k in the usual fixed k-class estimator is less than 1. Since the orders of matrices are not changed by the modifications in (8), we may employ the optimal values of k indicated by Zellner's (1978) and Nagar's (1959) results. While the modified LIML estimator, $(\boldsymbol{d}_i)_{\lambda^*,\mu}$, may not dominate the modified fixed k-class estimator (8) with $k = \bar{k} \leqslant 1$, denoted by $(\boldsymbol{d}_i)_{\bar{k},\mu}$, even asymptotically, the estimator $\hat{\boldsymbol{d}}_i = (\boldsymbol{d}_i)_{\lambda^*,\mu} + \boldsymbol{\zeta}(\boldsymbol{d}_i)$, where $\boldsymbol{\zeta}(\boldsymbol{d}_i)$ is $\boldsymbol{\zeta}(\boldsymbol{\delta}_i) = E[(\boldsymbol{d}_i)_{\bar{k},\mu} - (\boldsymbol{d}_i)_{\lambda^*,\mu}]$ evaluated at $\boldsymbol{\delta}_i = (\boldsymbol{d}_i)_{\lambda^*,\mu}$, may asymptotically dominate $(\boldsymbol{d}_i)_{\bar{k},\mu}$, as Ghosh, Sinha and Wieand's (1980) study of a simpler situation suggests. Here we are using $(\boldsymbol{d}_i)_{\lambda^*,\mu}$ to correct itself by estimating $\boldsymbol{\zeta}(\boldsymbol{\delta}_i)$, see Efron (1982, p. 350, Remark D).

It is important to remember that our discussion thus far is limited to the case where $T > K$. In the undersized sample case where $T \leqslant K$, the modified fixed k-class estimator (8) can be generalized straightforwardly as in Swamy (1980) and Swamy and Mehta (1982). The condition $T \geqslant \text{rank}(X) + G_i + 1$, which is required for the existence of the (modified) LIML estimator, may not be satisfied when $T \leqslant K$ or the samples are undersized.

4. Conclusions

In this paper we argue that the LIML estimator is certainly preferable to the 2SLS in terms of the speed with which the distribution approaches normality as the sample size tends to infinity, as shown by Anderson and Sawa (1979). But the size of the currently available sample may not be large enough to guarantee our reliance on the asymptotic normality of LIML. Although both the fixed k-class (of which the 2SLS is a member) and the LIML estimators can be modified to ensure finite moments, the modifications of the former converge in the r-th mean for every $k \leq 1$ and not the modifications of the latter. A modified fixed k-class estimator with $k \leq 1$ is preferable to the modified LIML estimator in terms of convergence in the r-th mean to normal, which is a stronger property than convergence in probability or in law to normal. The central limit theorem for the density of this modified fixed k-class estimator can be strengthened considerably, since its higher order moments exist in small samples.

References

Anderson, T. W. (1976): Estimation of linear functional relationships: approximate distributions and connections with simultaneous equations in econometrics (with discussion). *J. Roy. Statist. Soc.*, B, **38**, 1–36.

Anderson, T. W. and Sawa, T. (1979): Evaluation of the distribution function of the two-stage least squares estimate. *Econometrica*, **47**, 163–182.

Anderson, T. W., Kunitomo, N. and Sawa, T. (1982): Evaluation of the distribution function of the limited information maximum likelihood estimator. *Econometrica*, **50**, 1009–1028.

Brown, L. D. (1971): Admissible estimators, recurrent diffusions, and insoluble boundary value problems. *Ann. Math. Statist.*, **42**, 855–904.

Chung, Kai Lai (1974): *A Course in Probability*, 2nd ed., Academic Press, New York.

Efron, B. (1982): The 1981 Wald memorial lectures: Maximum likelihood and decision theory. *Ann. Statist.*, **10**, 340–356.

Feller, W. (1966): *An Introduction to Probability Theory and its Applications*, Vol. 2, Wiley, New York.

Fujikoshi, Y., Morimune, K., Kunitomo, N. and Taniguchi, M. (1982): Asymptotic expansions of the distributions of the estimates of coefficients in a simultaneous equation system. *Journal of Econometrics*, **18**, 191–205.

Fuller, W. A. (1977): Some properties of a modification of the limited information estimator. *Econometrica*, **45**, 939–954.

Ghosh, J. K., Sinha, B. K. and Wieand, H. S. (1980): Second order efficiency of the MLE with respect to any rounded bowl-shaped loss function. *Ann. Statist.*, **8**, 506–521.

Hatanaka, M. (1973): On the existence and the approximation formulas for the moments of the k-class estimators. *Economic Studies Quarterly*, **24**, 1–15.

Kadiyala, K. R. (1970): An exact small sample property of the k-class estimators. *Econometrica*, **38**, 930–932.

Laha, R. and Rohatgi, V. K. (1979): *Probability Theory*. Wiley, New York.

Lukacs, Eugene (1975): *Stochastic Convergence*, 2nd ed., Academic Press, New York.

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. *Jour. Amer. Statist. Assoc.*, **67**, 159–163.

Mehta, J. S. and Swamy, P. A. V. B. (1978): The existence of moments of some simple Bayes estimators of coefficients in a simultaneous equation model. *Journal of Econometrics*, **7**, 1–13.

Nagar, A. L. (1959): The bias and moment matrix of the general k-class estimators of the parameters in simultaneous equations. *Econometrica*, **27**, 575–595.

Rao, C. R. (1973): *Linear Statistical Inference and its Applications*, 2nd ed., Wiley, New York.

Swamy, P. A. V. B. (1980): A comparison of estimators for undersized samples. *Journal of Econometrics*, **14**, 161–181.

Swamy, P. A. V. B., Mehta, J. S. and Rappoport, P. N. (1978): Two methods of evaluating Hoerl-Kennard ridge regression estimator. *Communications in Statistics*, **A7**, 1133–1172.

Swamy, P. A. V. B. and Mehta, J. S. (1980): Ridge regression estimation of the Rotterdam model. *Special Studies Paper*, Federal Reserve Board, Washington, D.C., to appear in *Journal of Econometrics*.

Swamy, P. A. V. B., Mehta, J. S. and Thurman, S. S. (1981): A generalized multicollinearity index for estimation. *Special Studies Paper*, Federal Reserve Board, Washington, D.C.

Paper received : *September*, 1982.