Discussion Paper No. 501

A THIRD ORDER OPTIMUM PROPERTY OF THE ML ESTIMATOR IN LINEAR FUNCTIONAL RELATIONSHIPS AND SIMULTANEOUS EQUATION SYSTEMS

by

Naoto Kunitomo*

September 1981

*Department of Economics--Northwestern University

This paper is originally Chapter 6 of the author's Ph.D. dissertation submitted to Stanford University. The author is grateful to his principal advisor, Professor T.W. Anderson, for his encouragement and comments in the development of the thesis.

Abstract

The Maximum Likelihood (ML) estimator and its modification in the linear functional relationships model are shown to be third-order asymptotically efficient among a class of almost median-unbiased and almost mean-unbiased estimators, respectively. This implies that the Limited Information Maximum Likelihood (LIML) estimator in the simultaneous equation system is third-order asymptotically efficient when the number of excluded exogenous variables is growing along with the sample size. That is, the LIML estimator has an optimum property when the system of simultaneous equation is large.

Key Words

Maximum Likelihood Estimator, Third-Order Efficiency, Linear Functional Relationship, Large Econometric Model

LIML Estimator

1. Introduction

The concept of asymptotic higher order efficiencies has been recently developed by some theoretical statisticians. Among them, Ghosh et al [1980], Pfanzagle and Wefelmeyer [1978], and Akahira and Takeuchi [1979] are basic references and Efron et al [1980] is an excellent review to find how leading statisticians have different opinions on this subject. The Maximum Likelihood (ML) estimator and the Bayesian estimator with smooth priors have third-order asymptotic efficiency under some regularity conditions. This means that given an estimator we can always construct a modified ML estimator which has the same asymptotic bias and smaller asymptotic loss than the estimator to be compared. Hence, there is no reason why we should choose other estimators except the ML estimator or its modifications.

The purpose of the present article is to show that the ML estimator itself has a third-order optimum property among almost median-unbiased estimators in linear functional relationship models. We also show that a modification of the ML estimator has a third-order optimum property among almost mean-unbiased estimators.

In the linear functional relationship, the number of parameters increases together with the sample size and so we cannot simply apply general theorems in the regular asymptotic theory. In fact, the Least Squares (LS) estimator is inconsistent while the ML estimator is consistent but the ML does not attain the Cramér-Rao lower bound in the linear functional relationships. However, Takeuchi [1972] proved that the ML estimator attains the lower bound of asymptotic variance among a certain class of consistent estimators (Morimune and Kunitomo [1980]). Therefore, further comparison of estimators should be made in terms of higher-order terms of the asymptotic expansions of their distributions.

Anderson [1976] first shed light on connections between the estimation problem of linear functional relationships and that of structural equation in a simultaneous equation system in econometrics. The ML estimator of the slope in the linear functional relationships is mathematically equivalent to the Limited Information Maximum Likelihood (LIML) estimator of a structural coefficient when the covariance matrix of the reduced form is known in simultaneous equation, and the LS estimator in the former is equivalent to the Two-Stage Least Squares (TSLS) estimator in the latter. Further, as Anderson [1976] showed, the parameter sequence, in which the noncentrality parameter increases while the sample size N stays fixed in the linear functional relationships, is the one on which the regular asymptotic theory in econometrics has been concentrating. In this situation, the LIML and TSLS estimators are the best asymptotically normal (BAN) estimators, namely, the two methods yield consistent estimators and the two sets of estimators normalized by the square root of the sample size T have the same limiting joint normal distributions with the covariance of the standardized Fisher information matrix. Here we should note that the sample size T in the simultaneous equation system is different from the sample size N in the linear functional relationships. Since there exist two BAN estimators, several modifications have been proposed in order to improve LIML and (or) TSLS in some sense. For example, see Nagar [1959], Kadane [1971], Sawa [1973], Fuller [1977], Morimune [1978], Takeuchi [1978], Takeuchi and Morimune [1979].

On the other hand, Kunitomo [1980] clarified the meaning of another parameter sequence in which both the noncentrality parameter and the sample size N increase. It may be appropriate in the linear functional relationships. Anderson [1976] showed that the sample size N minus one is the number of excluded exogenous variables in the structural equation of interest in the simultaneous equation system, say K_2. (Although Anderson [1976] uses the two endogenous variables case, his arguments hold in the general case. See Kunitomo [1981 a], for instance.) Since recent macro-econometric models are more or less large in their size and hence K_2 is fairly large, the above parameter sequence can be interpreted as a new asymptotic theory, called the large-K_2 asymptotics, for large econometric models (Kunitomo, 1981b).

As K_2 increases along with the sample size T, the LIML estimator is consistent and asymptotically efficient while both the TSLS estimator and the Ordinary Least Squares (OLS) estimator are inconsistent under appropriate regularity conditions. Furthermore, the modifications of the LIML estimator by Fuller (1977) and Morimune (1978) are shown to improve the LIML estimation in terms of the asymptotic mean squared error, which is defined by the mean squared error of the asymptotic expansion of distributions. Hence there was some ambiguity on the higher order asymptotic optimality of estimator in single equation methods when K_2 is large.

The results obtained in Section 2 imply that the LIML estimator is third order efficient among almost median-unbiased estimators and a modification of the LIML estimator is third order asymptotically efficient among almost mean-unbiased estimators. Therefore the LIML estimation method gives the best estimator if we adjust the asymptotic bias according to our choice of criterion: the median unbiasedness or the mean-unbiasedness etc. We note that almost mean-unbiased estimators when K_2 increases are different from those in the usual large sample asymptotic theory.

One important approach studied in econometrics in the past is small sample theory. Anderson and Sawa [1979], and Anderson, Kunitomo, and Sawa [1981] evaluated the exact distribution functions of the TSLS and LIML estimators, respectively, with systematic computation for a limited number of suitably chosen cases with different values of the <u>key</u> parameters in the simultaneous equation system. The most important finding in their studies is that the TSLS estimator is badly biased while the distribution of the LIML estimator is centered at the parameter value when K_2 is large. In this respect, the results reported in this paper can justify their findings theoretically.

We shall present the model, the assumptions, and the statement of theorems in Section 2. A general model and some implications of our results in econometrics will be discussed in Section 3. Proofs of theorems are given in Section 4 and the validity of asymptotic expansions are discussed in the Appendix.

2. Main Results

Suppose $(x_g\ y_g)$ is an observation from a bivariate normal distribution with mean $(\mu_g\ \nu_g)$, variance σ_{xx} and σ_{yy}, and covariance σ_{xy}, $g = 1,\ldots,N$, for $N > 1$, and suppose that the observations are independent. The parameters $(\mu_g\ \nu_g)$ are assumed to satisfy a linear relationship:

$$\nu_g = \alpha + \beta\mu_g \quad , \quad g = 1,\ldots,N \quad . \tag{2.1}$$

The angle between the line (2.1) and the μ_g-axis may replace the slope $\beta = \tan\theta$. It will be convenient to write

$$x_g = \mu_g + u_g \quad , \quad g = 1,\ldots,N \quad , \tag{2.2}$$

$$y_g = \nu_g + v_g \quad , \quad g = 1,\ldots,N \quad , \tag{2.3}$$

where u_g and v_g are normally distributed random variables with means zero and covariance matrix

$$\underset{\sim}{\Sigma} = \sigma^2\begin{pmatrix}\sigma_{xx} & \sigma_{xy}\\ \sigma_{xy} & \sigma_{yy}\end{pmatrix} .$$

A case of special interest is the model of $\sigma_{xy} = 0$ and $\sigma_{xx} = \sigma_{yy} = 1$. In this case each pair of errors has a normal distribution with vector of means 0 and the covariance matrix $\sigma^2 \underset{\sim}{I}$. If the covariance matrix is a function of more than one unknown parameter, estimation methods are more or less arbitrary. See Anderson [1976], and Kendall and Stuart ([1973], Ch. 29) for discussion.

The stimator of α in any method here is $\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}$, where $\hat{\beta}$ is an estimator of β and

$$\bar{x} = N^{-1} \sum_{g=1}^{N} x_g \quad , \quad \bar{y} = N^{-1} \sum_{g=1}^{N} y_g \quad . \tag{2.4}$$

The estimator of β is then defined in terms of deviations from the means $(\bar{x}\ \bar{y})$. Let

$$s_{xx} = \frac{1}{n} \sum_{g=1}^{N} (x_g - \bar{x})^2 \quad , \tag{2.5}$$

$$s_{yy} = \frac{1}{n} \sum_{g=1}^{N} (y_g - \bar{y})^2 \quad , \tag{2.6}$$

$$s_{xy} = \frac{1}{n} \sum_{g=1}^{N} (x_g - \bar{x})(y_g - \bar{y}) \quad , \tag{2.7}$$

and $n = N - 1$.

Other notations used here are the standardized estimator

$$\hat{e} = \frac{\lambda}{(1 + \beta^2)} (\hat{\beta} - \beta) \quad , \tag{2.8}$$

and the noncentrality parameter

$$\lambda^2 = \frac{(1 + \beta^2)}{\sigma^2} \sum_{g=1}^{N} (\mu_g - \bar{\mu})^2 \quad , \tag{2.9}$$

where

$$\bar{\mu} = N^{-1} \sum_{g=1}^{N} \mu_g \quad .$$

The parameter λ^2 may be interpreted as a measure of the spread of the true values about their means. The assumption we shall make here to derive asymptotic distributions of estimators is the following.

<u>Assumption A</u>: There exists a finite positive number ρ such that

$$\frac{\sum_{g=1}^{N} (\mu_g - \bar{\mu})^2}{n\sigma^2} = \rho + O(n^{-1}) \quad , \tag{2.10}$$

and $\delta = (1 + \beta^2)\rho$, where $n = N - 1$.

Assumption A means that the noncentrality parameter (the spread of the true values) increases with the same order as the sample size N.

In the following analysis, λ^2 is replaceable with $n\delta$, and λ^2 is employed instead of N wherever we are able to avoid complexity of expressions. We note that it is possible to extend our results to alternative parameter sequences instead of Assumption A, which will be briefly discussed in Section 4.

Define a class of estimators, called the extended regular efficient estimator (See Akahira and Takeuchi [1979]), by

$$\hat{\beta} = \phi(s_{yy}, s_{yx}, s_{xx}) + \frac{1}{N}\psi(s_{yy}, s_{yx}, s_{xx}) \quad , \tag{2.11}$$

where s_{yy}, s_{yx}, and s_{xx} are given by (2.4)-(2.7) and $\phi(\cdot)$ is four times continuously differentiable, $\psi(\cdot)$ is twice continuously differentiable, and both are independent of $N = n + 1$, and all of the derivatives are bounded around true parameters.

This class includes the ML estimator, the LS estimator, and their modifications. Also we define the third-order asymptotic median-unbiased (AMDU) estimator by

$$\lim_{n\to\infty} \sup_{\beta\in U_\delta} n\left|\Pr(\hat{\beta} \leq \beta) - \frac{1}{2}\right| = 0 \quad , \tag{2.12}$$

and

$$\lim_{n\to\infty} \sup_{\beta\in U_\delta} n\left|\Pr(\hat{\beta} > \beta) - \frac{1}{2}\right| = 0 \quad , \tag{2.13}$$

where U_δ is a neighborhood $|\beta - \beta_0| \leq \delta$ for some $\delta > 0$ and any β_0.

Theorem 1: For all $\xi_1 \geq 0$ and $\xi_2 \geq 0$,

$$(2.14)\quad \lim_{n\to\infty} n[\Pr\{-\xi_1 < \sqrt{n}(\hat{\beta}_{ML} - \beta) \leq \xi_2\} - \Pr\{-\xi_1 < \sqrt{n}(\hat{\beta} - \beta) \leq \xi_2\}] \geq 0,$$

where $\hat{\beta}$ is any AMDU estimator and $\hat{\beta}_{ML}$ is given by

$$(2.15)\qquad \hat{\beta}_{ML} = \frac{s_{yy} - s_{xx} + [(s_{yy} - s_{xx})^2 + 4s_{xy}^2]^{1/2}}{2s_{xy}} \quad .$$

Corollary 1: The ML estimator has a third-order optimum property among AMDU estimators with respect to any bounded (bowl-shaped) loss function $L_n(\beta,a) = h(n^{1/2}(a - \beta))$ whose minimum value is zero at $\beta = a$ and which increases with $|a - \beta|$.

Similarly, we define the third-order asymptotic mean-unbiased (AMNU) estimator by

$$(2.16)\qquad \lim_{n\to\infty} \sup_{\beta\in U_\delta} n|AM_n(\hat{\beta} - \beta)| = 0 \quad ,$$

where $AM_n(\cdot)$ stands for the expectation with respect to the Edgeworth expansion of $\hat{\beta}$.

Theorem 2: For all $\xi_1 \geq 0$ and $\xi_2 \geq 0$,

$$(2.17)\quad \lim_{n\to\infty} n[\Pr\{-\xi_1 < \sqrt{n}(\hat{\beta}^* - \beta) \leq \xi_2\} - \Pr\{-\xi_1 < \sqrt{n}(\hat{\beta} - \beta) \leq \xi_2\}] \geq 0,$$

where $\hat{\beta}$ is any AMNU estimator and $\hat{\beta}^*$ is given by

$$(2.18)\qquad \hat{\beta}^* = \frac{2s_{xy}}{s_{xx} - s_{yy} + 2\hat{c} + [(s_{xx} - s_{yy})^2 + 4s_{xy}^2]^{1/2}}$$

and

$$(2.19)\qquad n\hat{c} = 1 + \frac{\ell_1}{\ell_2 - \ell_1} \quad ,$$

where ℓ_1 and ℓ_2 are the smaller and the larger characteristic roots of the equation: $\left|\begin{pmatrix} s_{xx} & s_{xy} \\ s_{xy} & s_{yy} \end{pmatrix} - \sigma^2 \ell \underset{\sim}{I}\right| = 0.$

Corollary 2: The estimator $\hat{\beta}^*$ has a third-order optimum property among AMNU estimators with respect to any bounded (bowl-shaped) loss function $L_n(\beta,a)$.

Turning to the estimation of the angle, the following theorem holds.

Theorem 3: For all $\xi_1 \geq 0$ and $\xi_2 \geq 0$,

$$(2.20)\quad \underline{\lim}_{n\to\infty}\, n[\Pr\{-\xi_1 < \sqrt{n}(\hat{\theta}_{ML} - \theta) \leq \xi_2\} - \Pr\{-\xi_1 < \sqrt{n}(\hat{\theta} - \theta) \leq \xi_2\} \geq 0 \ ,$$

where $\hat{\theta}$ is any AMDU or AMNU estimator.

Corollary 3: The ML estimator of angle has a third-order optimum property among AMDU and AMNU estimators with respect to any bounded bowl-shaped loss function whose minimum value is zero at $\theta = a$ and which increases with $|a - \theta|$.

3. A Generalization of Takeuchi-Theorem and Simultaneous Equations System

The model we have been considering so far may be unrealistic in practical situations since the covariance matrix is assumed to be known to a proportionality constant. However, this assumption is not essential in our results. In this section, we shall discuss the linear functional relationships model with an arbitrary (known or unknown) covariance matrix and its connnection with simultaneous equation systems in econometrics.

For an arbitrary covariance matrix $\underset{\sim}{\Omega} = (\omega_{ij})$, the likelihood is given by

$$\log L = -\frac{N}{2}\log(2\pi) - \frac{N}{2}\log[\sigma_1^2\sigma_2^2(1-\tau^2)] - \frac{1}{2(1-\tau^2)}\sum_{i=1}^{N} \times \left\{\left(\frac{x_i-\mu_i}{\sigma_1}\right)^2 - 2\tau\left(\frac{x_i-\mu_i}{\sigma_1}\right)\left(\frac{y_i-\alpha-\beta\mu_i}{\sigma_2}\right) + \left(\frac{y_i-\alpha-\beta\mu_i}{\sigma_2}\right)^2\right\}, \tag{3.1}$$

where $\omega_{11} = \sigma_1^2$, $\omega_{12} = \tau\sigma_1\sigma_2$, and $\omega_{22} = \sigma_2^2$.

Then the information matrix for $(\beta,\alpha,\mu_1,\ldots,\mu_N)$ is

$$I(\beta,\alpha,\mu_1,\ldots,\mu_N) \tag{3.2}$$

$$= -\mathcal{E}\begin{bmatrix} \frac{\partial^2 \log L}{\partial^2\beta} & \frac{\partial^2 \log L}{\partial\beta\partial\alpha} & \frac{\partial^2 \log L}{\partial\beta\partial\mu_1} & \cdots & \frac{\partial^2 \log L}{\partial\beta\partial\mu_N} \\ \frac{\partial^2 \log L}{\partial\beta\partial\alpha} & \frac{{}^2 \log L}{\partial^2\alpha} & \frac{{}^2 \log L}{\partial\alpha\partial\mu_1} & \cdots & \frac{{}^2 \log L}{\partial\alpha\partial\mu_N} \\ \frac{{}^2 \log L}{\partial\beta\partial\mu_1} & \frac{{}^2 \log L}{\partial\alpha\partial\mu_1} & \frac{{}^2 \log L}{\partial^2\mu_1} & \cdots & \frac{{}^2 \log L}{\partial\mu_1\partial\mu_N} \\ \vdots & \vdots & \vdots & & \vdots \\ \frac{\partial^2 \log L}{\partial\beta\partial\mu_N} & \frac{\partial^2 \log L}{\partial\alpha\partial\mu_N} & \frac{\partial^2 \log L}{\partial\mu_1\partial\mu_N} & & \frac{\partial^2 \log L}{\partial^2\mu_N} \end{bmatrix}$$

$$
= \frac{1}{(1-\tau^2)\sigma_1^2}
\begin{bmatrix}
\sum_i \mu_i^2 & \sum_i \mu_i & \mu_1(\beta - \tau\frac{\sigma_2}{\sigma_1}) & \cdots & \mu_N(\beta - \tau\frac{\sigma_2}{\sigma_1}) \\
\sum_i \mu_i & N & \beta - \tau\frac{\sigma_2}{\sigma_1} & & \beta - \tau\frac{\sigma_2}{\sigma_1} \\
\mu_1(\beta - \tau\frac{\sigma_2}{\sigma_1}) & \beta - \frac{\sigma_2}{\sigma_1} & & & \\
\vdots & \vdots & & & \\
\mu_N(\beta - \tau\frac{\sigma_2}{\sigma_1}) & \beta - \tau\frac{\sigma_2}{\sigma_1} & & (\beta^2 + \frac{\sigma_2^2}{\sigma_1^2} - 2\beta\tau\frac{\sigma_2}{\sigma_1}) \cdot \underset{\sim}{I}_N &
\end{bmatrix},
$$

and so the partial information for β is

$$
(3.3) \qquad I(\beta) = \frac{\sum_{i=1}^{N} (\mu_i - \bar{\mu})^2}{(\omega_{11}\beta^2 - 2\beta\omega_{12} + \omega_{22})} .
$$

Here we note that the partial information for β when $\underset{\sim}{\Omega}$ is unknown is the same as $I(\beta)$ because $(\partial^2 \log L/\partial\beta\partial\omega_{ij}) = (\partial^2 \log L/\partial\alpha\partial\omega_{ij})$ $= (\partial^2 \log L/\partial\mu_k\partial\omega_{ij}) = 0$ for $i, j = 1,2$ and $k = 1,\ldots,N$.

Now we can prove the next theorem by using arguments similar to those in the proof of Lemma 1 in Section 4.

<u>Theorem 4:</u> Let an estimator of β be $\hat{\beta} = \phi(s_{yy}, s_{xy}, s_{xx})$, where ϕ is continuously differentiable.

(i) If n is fixed and λ^2 goes to infinity, or alternatively $n << \lambda^2$, then

$$\text{(3.4)} \qquad AM_n \{\sqrt{I(\beta)}(\hat{\beta} - \beta)\}^2 \geq 1 \quad ,$$

and the equality holds for the ML estimator.

(ii) If both n and λ^2 go to infinity while the ratio goes to a constant, then

$$\text{(3.5)} \qquad AM_n \{\sqrt{I(\beta)}(\hat{\beta} - \beta)\}^2 \geq 1 + \frac{1}{\delta} \quad ,$$

and the equality holds for the ML estimator.

(iii) If λ^2 is fixed and n goes to infinity, or alternatively $\lambda^2 << n$, then there does not exist any consistent estimator.

Takeuchi [1972] proved (ii) for the case $\underset{\sim}{\Omega} = \sigma^2 \underset{\sim}{I}$ (Morimune and Kunitomo [1980]). In the case of (i), which corresponds to the parameter sequence Anderson [1976] considered, the ML attains the Cramér-Rao lower bound. However, the regular asymptotic properties of the ML estimator such as the Cramér-Rao lower bound cannot be applied to the case of (ii) and (iii) since the number of parameters increases along with the sample size. For the parameter sequence of (ii), which Kunitomo [1980], and Morimune and Kunitomo [1980] investigated, the ML estimator attains a possible lower bound which is larger than the information quantity. For the parameter sequence of (iii), a consistent estimator cannot be constructed since the number of parameters grows too fast to give enough information for estimation. Hence the ML estimator loses even consistency in this situation.

Let a structural equation in time period t in a simultaneous equations system be

$$(3.6) \qquad y_{2t} = \beta y_{1t} + \sum_{k=1}^{K_1} \gamma_{2k} z_{kt} + u_{1t} \quad , \quad t = 1,\ldots,T \quad ,$$

where $(y_{1t} \quad y_{2t})$ are two endogenous variables and u_{1t} is an unobservable variable with mean zero and variance τ^2. The reduced form of the system of structural equations includes

$$(3.7) \qquad y_{it} = \sum_{k=1}^{K} \pi_{ik} z_{kt} + v_{it} \quad , \quad i = 1,2 \quad , \quad t = 1,\ldots,T \quad ,$$

where $(\pi_{1k} \quad \pi_{2k})$ are the coefficients of the reduced form, and $(v_{1t} \quad v_{2t})$ are unobservable random variables with mean zero and unknown covariance matrix

$$(3.8) \qquad \underset{\sim}{\Omega} = \sigma^2 \begin{pmatrix} \omega_{11} & \omega_{12} \\ \omega_{12} & \omega_{22} \end{pmatrix} .$$

The variables $z_{1t},\ldots,z_{Kt}$ $(K > K_1)$ are nonstochastic exogenous variables which may include a constant term. We define $K_2 = K - K_1$, which is the number of excluded exogenous variables in the structural equation of interest. Then a similar result can be obtained for the simultaneous equation system.

<u>Theorem 5:</u> Let an estimator of β be $\hat{\beta} = \phi(s_{yy}, s_{xy}, s_{xx}; \hat{\omega}_{11}, \hat{\omega}_{12}, \hat{\omega}_{22})$, where ϕ is continuously differentiable. Let $q = T - K$ be the degrees of freedom for ω_{ij}.

(i) If n is fixed and λ^2 goes to infinity, or alternatively $n << \lambda^2$, then

$$AM_n \{\sqrt{I(\beta)}(\hat{\beta} - \beta)\}^2 \geq 1 \quad , \tag{3.9}$$

and the equality holds for the LIML estimator.

(ii) If both n and λ^2 go to infinity while the ratio goes to a constant, then

$$AM_n \{\sqrt{I(\beta)}(\hat{\beta} - \beta)\}^2 \geq 1 + \frac{1}{\delta} \tag{3.10}$$

for $q >> n$, and

$$AM_n \{\sqrt{I(\beta)}(\hat{\beta} - \beta)\}^2 \geq 1 + \frac{1}{\delta}(1 + \frac{\nu^2}{\delta}) \tag{3.11}$$

for $q = O(n)$, where $\nu^2 = \lim_{q\to\infty} \lambda^2/q$. The equality holds for the LIML estimator.

(iii) If λ^2 is fixed and n goes to infinity, or alternatively $\lambda^2 << n$, then there does not exist any consistent estimator.

The proof is similar to that of Lemma 1 in Section 4. In the simultaneous equation systems one can estimate the covariance matrix by using the residual matrix of the regression estimates of the reduced form parameters π_{ij}. Therefore, if we have enough observations (or equivalently the sample size T) to estimate the covariance matrix consistently, the results for the linear functional relationships are still valid.

So far we discussed the first-order efficiency of the ML and LIML estimators in this section. Further, the third-order asymptotic optimality of the ML and LIML estimators can be proven with some minor modifications of the proof in Section 4 for the alternative parameter sequences we discussed. In the more general case where the parameter of interest is a vector or matrix (such as some subsystem of simultaneous equations), the Edgeworth expansion of distribution is very complicated. However, in principle, similar results can be obtainable.

4. Proofs of theorems

Measuring all x_i, y_i, and μ_i from their means, we construct the following vectors: $\underset{\sim}{x}' = (x_1,\ldots,x_N)\underset{\sim}{P}$, $\underset{\sim}{y}' = (y_1,\ldots,y_N)\underset{\sim}{P}$, and $\underset{\sim}{\mu}' = (\mu_1,\ldots,\mu_N)\underset{\sim}{P}$ where $\underset{\sim}{P} = \underset{\sim}{I}_N - (1/N)\underset{\sim}{e}\underset{\sim}{e}'$ and $\underset{\sim}{e}' = (1,\ldots,1)$. Since there exists an $N \times N$ an orthogonal matrix R such that

$$(4.1) \qquad \underset{\sim}{R}(1 + \beta^2)^{1/2} \frac{\underset{\sim}{\mu}}{\sigma} = (\lambda,0,\ldots,0)' \quad ,$$

where the N-th row is $(1/\sqrt{N})\underset{\sim}{e}'$.

We define N $(= n + 1)$ vectors

$$(4.2) \qquad \underset{\sim}{u}^* = \underset{\sim}{R}\underset{\sim}{P}(1 + \beta^2)^{-1/2}\left(\frac{\underset{\sim}{x} + \beta\underset{\sim}{y}}{\sigma}\right)$$

$$= (\lambda,0,\ldots,0)' + (u_1,\ldots,u_n,0)' \quad ,$$

$$(4.3) \qquad \underset{\sim}{v}^* = \underset{\sim}{R}\underset{\sim}{P}(1+\beta^2)^{-1/2}\left(\frac{-\beta \underset{\sim}{x} + \underset{\sim}{y}}{\sigma}\right) = (v_1,\ldots,v_n,0)' \quad ,$$

where $\mathcal{E}(u_1^*) = \mathcal{E}(v_i^*) = \mathcal{E}(u_i^* v_j^*) = 0$, and $\mathcal{E}(u_i^*)^2 = \mathcal{E}(v_i^*)^2 = 1$. Then $\underset{\sim}{x}$ and $\underset{\sim}{y}$ can be written in terms of $\underset{\sim}{u}^*$ and $\underset{\sim}{v}^*$ as follows.

$$(4.4) \qquad \underset{\sim}{x} = \sigma(1+\beta^2)^{-1/2} \; \underset{\sim}{R}'(\underset{\sim}{u}^* - \beta\underset{\sim}{v}^*) \quad ,$$

$$(4.5) \qquad \underset{\sim}{y} = \sigma(1+\beta^2)^{-1/2} \; \underset{\sim}{R}'(\beta\underset{\sim}{u}^* + \underset{\sim}{v}^*) \quad .$$

Defining $s_{uu} = \underset{\sim}{u}^{*\prime}\underset{\sim}{u}^*$, $s_{vv} = \underset{\sim}{v}^{*\prime}\underset{\sim}{v}^*$, and $s_{uv} = \underset{\sim}{u}^{*\prime}\underset{\sim}{v}^*$, we have

$$(4.6) \qquad s_{xx} = \sigma^2(1+\beta^2)^{-1}\left(\frac{s_{uu} + \beta^2 s_{vv} - 2\beta s_{uv}}{n}\right) \quad ,$$

$$(4.7) \qquad s_{yy} = \sigma^2(1+\beta^2)^{-1}\left(\frac{\beta^2 s_{uu} + s_{vv} + 2\beta s_{uv}}{n}\right) \quad ,$$

$$(4.8) \qquad s_{xy} = \sigma^2(1+\beta^2)^{-1}\left(\frac{\beta s_{uu} - \beta s_{vv} + (1-\beta^2) s_{uv}}{n}\right) \quad .$$

In this section we shall derive formal asymptotic expansions of the distributions of estimators in two lemmas. The validity of the formal expansions will be seen in the Appendix.

<u>Lemma 1</u>: A necessary and sufficient condition for an efficient estimator among the consistent estimators is

$$\phi_1 = -\phi_3 = \frac{\beta}{(1+\beta^2)\rho}\ , \tag{4.9}$$

$$\phi_2 = \frac{1-\beta^2}{(1+\beta^2)\rho}\ ,$$

where $\phi_i = \partial\phi/\partial h_i$, $i = 1,\ldots,3$, $h_1 = s_{yy}$, $h_2 = s_{xy}$, and $h_3 = s_{xx}$.

<u>Proof</u>: Taking the probability limits of s_{yy}, s_{xy}, and s_{xx}, consistency implies that

$$\beta = \phi(1+\beta^2\rho,\ \beta\rho,\ 1+\rho)\ . \tag{4.10}$$

Then differentiating (4.10) with respect to β and ρ gives

$$\begin{pmatrix} 2\beta & 1 & 0 \\ \beta^2 & \beta & 1 \end{pmatrix} \begin{pmatrix} \phi_1 \\ \phi_2 \\ \phi_3 \end{pmatrix} = \begin{pmatrix} \frac{1}{\rho} \\ 0 \end{pmatrix}\ , \tag{4.11}$$

and hence

$$\begin{pmatrix} \phi_2 \\ \phi_3 \end{pmatrix} = \begin{pmatrix} -2\beta \\ \beta^2 \end{pmatrix} \phi_1 + \begin{pmatrix} \frac{1}{\rho} \\ -\frac{\beta}{\rho} \end{pmatrix}\ . \tag{4.12}$$

Then the asymptotic variance of $\hat{\beta}$ is given by

$$\text{(4.13)} \qquad AV(\hat{\beta}) = (\phi_1 \quad \phi_2 \quad \phi_3)\begin{pmatrix} 4\beta^2\rho + 2 & 2\beta\rho & 0 \\ 2\beta\rho & \rho + \beta^2\rho + 1 & 2\beta\rho \\ 0 & 2\beta\rho & 4\rho + 2 \end{pmatrix}\begin{pmatrix} \phi_1 \\ \phi_2 \\ \phi_3 \end{pmatrix}$$

$$= 2[(1 + \beta^2)\phi_1 - \frac{\beta}{\rho}]^2 + \frac{1}{\rho^2} + \frac{1 + \beta^2}{\rho}$$

$$\geq \frac{1}{\rho^2} + \frac{1 + \beta^2}{\rho} \quad .$$

The equality holds if and only if $\phi_1 = \beta/[(1 + \beta^2)\rho]$. (QED)

Lemma 2: Any efficient estimator can be expressed in the canonical form:

$$\text{(4.14)} \qquad \sqrt{n}(\phi - \beta) = U_0 + \frac{U_1}{\sqrt{n}} + \frac{U_2}{n} + R_1 \quad ,$$

where U_i, $i = 0,1,2$ are given by (4.20), (4.28), and (4.37), and R_1 is a remainder term of the order $o_p(n^{-1})$.

Proof: First, define the random variables

$$\text{(4.15)} \qquad y^*_{11} = (n)^{-1/2}(\sum_2^n u_i^2 - n) \quad ,$$

$$\text{(4.16)} \qquad y^*_{12} = \frac{\sum_2^n u_i v_i}{\left(\sum_2^n u_i^2\right)^{1/2}} \quad .$$

$$(4.17) \qquad y^*_{22} = (n)^{-1/2} \left(\sum_2^n v_i^2 - y_{12}^2 - n \right) .$$

We shall expand the estimators by Taylor's Theorem in y^*_{11}, y^*_{12}, y^*_{22}, u_1, and v_1 in the set J_n for which

$$(4.18) \qquad |y^*_{ii}| < 2 \log n \ , \quad i = 1,2 \quad , \quad |y^*_{12}| < 2(\log n)^{1/2} \ ,$$

$$|u_1| < 2(\log n)^{1/2} \quad \text{and} \quad |v_1| < 2(\log n)^{1/2} .$$

A Taylor expansion of the estimator yields

$$(4.19) \qquad \sqrt{n}(\phi - \beta) = \sum_{i=1}^{3} \phi_i h^*_i + \frac{1}{2\sqrt{n}} \sum_{i,j=1}^{3} \phi_{ij} h^*_i h^*_j$$

$$+ \frac{1}{6n} \sum_{i,j,k=1}^{3} \phi_{ijk} h^*_i h^*_j h^*_k + R_1 \ ,$$

where $h^*_1 = \sqrt{n}(h_1 - 1 - \beta^2\rho)$, $h^*_2 = \sqrt{n}(h_2 - \beta\rho)$, $h^*_3 = \sqrt{n}(h_3 - 1 - \rho)$, and $\phi_{ij} = \partial^2\phi/\partial h_i \partial h_j$, $\phi_{ijk} = \partial^3\phi/\partial h_i \partial h_j \partial h_k$ evaluated at $h_1 = 1 + \beta^2\rho$, $h_2 = \beta\rho$, $h_3 = 1 + \rho$, and R_1 is a polynomial of degree 3 in h^*_i, which is $O(n^{-3/2})$ and is $O[(\log \sqrt{n}/\sqrt{n})^3]$ uniformly in J_n.

From (4.9) and (4.19), we have

$$(4.20) \qquad U_0 = \frac{Z_2}{\rho} \ ,$$

where $Z_2 = s_{uv}/\sqrt{n}$.

Now the differentiation of (4.9) gives

$$2\beta\phi_{11} + \phi_{12} = \frac{1 - \beta^2}{\rho^2(1 + \beta^2)^2} , \tag{4.21}$$

$$\beta^2\phi_{11} + \beta\phi_{12} + \phi_{22} = \frac{-\beta}{\rho^2(1 + \beta^2)} , \tag{4.22}$$

$$2\beta\phi_{12} + \phi_{22} = \frac{-4\beta}{\rho^2(1 + \beta^2)^2} , \tag{4.23}$$

$$\beta^2\phi_{12} + \beta\phi_{22} + \phi_{23} = \frac{\beta^2 - 1}{\rho^2(1 + \beta^2)} , \tag{4.24}$$

$$2\beta\phi_{31} + \phi_{23} = \frac{\beta^2 - 1}{\rho^2(1 + \beta^2)^2} , \tag{4.25}$$

$$\beta^2\phi_{31} + \beta\phi_{32} + \phi_{33} = \frac{\beta}{\rho^2(1 + \beta^2)} . \tag{4.26}$$

Then rearranging each term, we have

$$\begin{pmatrix} \phi_{12} \\ \phi_{13} \\ \phi_{22} \\ \phi_{23} \\ \phi_{33} \end{pmatrix} = \begin{pmatrix} -2\beta \\ \beta^2 \\ 4\beta^2 \\ -2\beta^3 \\ \beta^4 \end{pmatrix} \phi_{11} + \frac{1}{\rho^2(1 + \beta^2)^2} \begin{pmatrix} 1 - \beta^2 \\ -2\beta \\ 2\beta(\beta^2 - 3) \\ 5\beta^2 - 1 \\ 2\beta(1 - \beta^2) \end{pmatrix} . \tag{4.27}$$

Substituting (4.27) into (4.19) we have

$$(4.28)\qquad U_1 = \frac{1}{2}\left\{(h_1^* \; h_2^* \; h_3^*)\begin{pmatrix} 1 & -2\beta & \beta^2 \\ -2\beta & 4\beta^2 & -2\beta^3 \\ \beta^2 & -2\beta^3 & \beta^4 \end{pmatrix}\begin{pmatrix} h_1^* \\ h_2^* \\ h_3^* \end{pmatrix}\phi_{11}\right.$$

$$\left. + \frac{1}{\rho^2(1+\beta^2)^2}(h_1^* \; h_2^* \; h_3^*)\begin{pmatrix} 0 & 1-\beta^2 & -2\beta \\ 1-\beta^2 & 2\beta(\beta^2-3) & 5\beta^2-1 \\ -2\beta & 5\beta^2-1 & 2\beta(\beta^2-3) \end{pmatrix}\begin{pmatrix} h_1^* \\ h_2^* \\ h_3^* \end{pmatrix}\right\}$$

$$= \frac{(1+\beta^2)^2}{2}\phi_{11}Z_3^2 + \frac{1}{\rho^2(1+\beta^2)}\{-Z_1Z_2 + Z_2Z_3 - \beta Z_3^2 + \beta Z_2^2\},$$

where $Z_1 = \sqrt{n}[s_{uu}/n - (1+\beta^2)\rho - 1]$, and $Z_3 = \sqrt{n}(s_{vv}/n - 1)$.

Next, again the differentiation of (4.27) yields

$$(4.29)\qquad 2\beta(\phi_{121} + 2\beta\phi_{111}) + (\phi_{122} + 2\beta\phi_{112}) = \frac{-2\phi_{11}}{\rho} - \frac{2\beta(3-\beta^2)}{\rho^3(1+\beta^2)^3},$$

$$(4.30)\qquad \beta^2(\phi_{121} + 2\beta\phi_{111}) + \beta(\phi_{122} + 2\beta\phi_{112}) + (\phi_{123} + 2\beta\phi_{113}) = \frac{2(\beta^2-1)}{\rho^3(1+\beta^2)^2},$$

$$(4.31)\qquad \beta^2(\phi_{131} - \beta^2\phi_{111}) + \beta(\phi_{132} - \beta^2\phi_{112}) + (\phi_{133} - \beta^2\phi_{113}) = \frac{4\beta}{\rho^3(1+\beta^2)^2},$$

$$(4.32)\qquad 2\beta(\phi_{221} - 4\beta^2\phi_{111}) + (\phi_{222} - 4\beta^2\phi_{112}) = \frac{8\phi_{11}\beta}{\rho} - \frac{2(\beta^4 - 12\beta^2 + 3)}{\rho^3(1+\beta^2)^3},$$

$$(4.33)\quad \beta^2(\phi_{221} - 4\beta^2\phi_{111}) + \beta(\phi_{222} - 4\beta^2\phi_{112}) + (\phi_{223} - 4\beta^2\phi_{113})$$

$$= \frac{4\beta(3 - \beta^2)}{\rho^3(1 + \beta^2)^2} \quad ,$$

$$(4.34)\quad 2\beta(\phi_{331} - \beta^4\phi_{111}) + (\phi_{332} - \beta^4\phi_{112}) = \frac{4\beta^3\phi_{11}}{\rho} + \frac{2(1 - 12\beta^2 + 3\beta^4)}{\rho^3(1 + \beta^2)^3} \quad ,$$

$$(4.35)\quad \beta^2(\phi_{331} - \beta^4\phi_{111}) + \beta(\phi_{332} - \beta^4\phi_{112}) + (\phi_{333} - \beta^4\phi_{113})$$

$$= \frac{4\beta(\beta^2 - 1)}{\rho^3(1 + \beta^2)^2} \quad .$$

Hence rearranging each term, we have

$$(4.36)\begin{pmatrix} \phi_{122} \\ \phi_{123} \\ \phi_{133} \\ \phi_{222} \\ \phi_{223} \\ \phi_{233} \\ \phi_{333} \end{pmatrix} = \begin{pmatrix} -4\beta^2 & -4\beta & 0 \\ 2\beta^3 & \beta^2 & -2\beta \\ -\beta^4 & 0 & 2\beta^2 \\ 16\beta^3 & 12\beta^2 & 0 \\ -8\beta^4 & -4\beta^3 & 4\beta^2 \\ 4\beta^5 & \beta^4 & -4\beta^3 \\ -2\beta^6 & 0 & 3\beta^4 \end{pmatrix} \begin{pmatrix} \phi_{111} \\ \phi_{112} \\ \phi_{113} \end{pmatrix} + \begin{pmatrix} -2 \\ 2\beta \\ -2\beta^2 \\ 12\beta \\ -10\beta^2 \\ 8\beta^3 \\ -6\beta^4 \end{pmatrix} \frac{\phi_{11}}{\rho}$$

$$+ \frac{1}{\rho^3(1+\beta^2)^3} \begin{pmatrix} 2\beta(\beta^2-3) \\ 2(3\beta^2-1) \\ 2\beta(3-\beta^2) \\ -6(\beta^4-6\beta^2+1) \\ 2\beta(9-11\beta^2) \\ 2(1-12\beta^2+3\beta^4) \\ 6\beta(3\beta^2-1) \end{pmatrix} .$$

Substituting (4.36) into (4.19) and rearranging each term, we finally obtain

$$(4.37) \qquad U_2 = \frac{(1+\beta^2)}{6}\Big\{3(\beta^2\phi_{111} + \beta\phi_{112} + \phi_{113} + \frac{4\beta}{[\rho^3(1+\beta^2)^3]})Z_1Z_3^2$$

$$+ 3(2\beta\phi_{111} + (1-\beta^2)\phi_{112} - 2\beta\phi_{113} + \frac{4\beta\phi_{11}}{\rho} - \frac{6\beta^2}{\rho^3(1+\beta^2)^3})Z_2Z_3^2$$

$$+ ((1-4\beta^2-2\beta^4)\phi_{111} - 3\beta\phi_{112} + 3\beta^2\phi_{113} - \frac{6\beta^2\phi_{11}}{\rho})Z_3^3$$

$$+ 6(\frac{-\phi_{11}}{\rho} + \frac{4\beta}{\rho^3(1+\beta^2)^3})Z_2^2Z_3\Big\}$$

$$+ \frac{1}{\rho^3(1+\beta^2)^3}(-(1-\beta^2)Z_2^3 + Z_1^2Z_2 - 2\beta Z_1Z_2^2 - 2Z_1Z_2Z_3) .$$

In the next two lemmas, we explicitly use the assumption of normality.

<u>Lemma 3:</u> An asymptotic expansion of the distribution of any efficient third-order AMDU estimator as n and λ^2 increase is given by

$$(4.38) \qquad \Pr\{\frac{\hat{e}}{\tau} \le \xi\}$$

$$= \Phi(\xi) - \frac{\beta\tau\xi^2\phi(\xi)}{\lambda}$$

$$+ \frac{\xi\phi(\xi)}{2\lambda^2\lambda^2}\{\frac{1}{2}(\frac{3}{\tau^2}-1) - 2a^2 + \xi^2[2\tau^4(\beta^2-1)+\frac{1}{2}(3-\frac{1}{\tau^2})] - \beta^2\tau^4\xi^4\}$$

$$+ O(\lambda^{-3}) \quad ,$$

where

$$(4.39) \qquad a = \rho(1+\beta^2)^2\{\phi_{11} - \frac{2\beta}{\rho^2(1+\beta^2)^3}\} \quad .$$

Proof: Let $T_0 = \text{plim}\, \psi(s_{yy}, s_{xy}, s_{xx})$

$$(4.40) \qquad = \psi(1+\beta^2\rho, \beta\rho, 1+\rho) \quad .$$

Then from (4.28), the third-order asymptotic median-unbiasedness requires

$$(4.41) \qquad (1+\beta^2)^2\phi_{11} - \frac{2\beta}{\rho^2(1+\beta^2)} + T_0 = 0 \quad .$$

Then differentiating (4.41) with respect to β and ρ,

$$(4.42) \qquad 2\beta\psi_1 + \psi_2 = \frac{2(1-\beta^2)}{\rho^3(1+\beta^2)^2} - \frac{4\beta(1+\beta^2)\phi_{11}}{\rho} - (1+\beta^2)^2(2\beta\phi_{111}+\phi_{112}) \quad ,$$

$$(4.43) \qquad \beta^2\psi_1 + \beta\psi_2 + \psi_3 = \frac{-4\beta}{\rho^3(1+\beta^2)} - (1+\beta^2)^2(\beta^2\phi_{111}+\beta\phi_{112}+\phi_{113}) \quad ,$$

where $\psi_1 = \partial\psi/\partial s_{yy}$, $\psi_2 = \partial\psi/\partial s_{xy}$, and $\psi_3 = \partial\psi/\partial s_{xx}$ evaluated at $s_{yy} = 1 + \beta^2\rho$, $s_{xy} = \beta\rho$, and $s_{xx} = 1 + \rho$.

Hence the conditions can be written as

$$(4.44) \qquad \begin{pmatrix} \psi_2 \\ \psi_3 \end{pmatrix} = \begin{pmatrix} -2\beta \\ \beta^2 \end{pmatrix}\psi_1 + \frac{2}{\rho^3(1+\beta^2)^2}\begin{pmatrix} 1-\beta^2 \\ -3\beta-\beta^3 \end{pmatrix} + \frac{4\beta(1+\beta^2)}{\rho}\begin{pmatrix} -1 \\ \beta \end{pmatrix}\phi_{11}$$

$$+ (1+\beta^2)^2\begin{pmatrix} -2\beta & -1 & 0 \\ \beta^2 & 0 & -1 \end{pmatrix}\begin{pmatrix} \phi_{111} \\ \phi_{112} \\ \phi_{113} \end{pmatrix} .$$

Then by (4.19) and (4.44)

$$(4.45) \qquad T_1 \equiv \sum_{i=1}^{3} \psi_i h_i^*$$

$$= Z_1\left\{\frac{-4\beta}{\rho^3(1+\beta^2)^2} - (1+\beta^2)(\beta^2\phi_{111} + \beta\phi_{112} + \phi_{113})\right\}$$

$$+ Z_2\left\{\frac{2(1+3\beta^2)}{\rho^3(1+\beta^2)^2} - (1+\beta^2)\left(\frac{4\beta\phi_{11}}{\rho} + 2\beta\phi_{111} + (1-\beta^2)\phi_{112}\right.\right.$$

$$\left.\left. - 2\beta\phi_{113}\right)\right\}$$

$$+ Z_3\left\{(1+\beta^2)\psi_1 - \frac{2\beta}{\rho^3(1+\beta^2)} + \beta(1+\beta^2)\left[\frac{4\phi_{11}\beta}{\rho} + \beta(\beta^2+2)\phi_{111}\right.\right.$$

$$\left.\left. + \phi_{112} - \beta\phi_{113}\right]\right\} .$$

Let

$$Q_0^* = \frac{Z_2}{\rho} \quad , \tag{4.46}$$

$$Q_1^* = U_1 + T_0 \quad , \tag{4.47}$$

$$Q_2^* = U_2 + T_1 \quad . \tag{4.48}$$

Then

$$Q_1^* = \frac{1}{\rho^2(1+\beta^2)} \{Z_2(Z_3 - Z_1) + \beta Z_2^2 + [\frac{\rho^2(1+\beta^2)^3 \phi_{11}}{2} - \beta](Z_3^2 - 2) \quad , \tag{4.49}$$

and

$$\begin{aligned} Q_2^* = (\frac{Z_3^2}{2} - 1)(1+\beta^2)\{ & Z_1(\beta^2\phi_{111} + \beta\phi_{112} + \phi_{113} + \frac{4\beta}{\rho^3(1+\beta^2)^3}) \\ & + Z_2[2\beta\phi_{111} + (1-\beta^2)\phi_{112} - 2\beta\phi_{113} + \frac{4\beta\phi_{11}}{\rho}] \\ & - \beta Z_3[\beta(\beta^2+2)\phi_{111} + \phi_{112} - \beta\phi_{113} + \frac{4\beta\phi_{11}}{\rho}] - \frac{6\beta^2 Z_2}{\rho^3(1+\beta^2)^3} \Big\} \\ & + Z_3^3[\frac{(1+\beta^2)\beta^2\phi_{11}}{\rho} + \frac{(1+\beta^2)^3\phi_{111}}{6}] \\ & + Z_3[(1+\beta^2)\psi_1 - \frac{2\beta}{\rho^3(1+\beta^2)}] + \frac{2Z_2}{\rho^3(1+\beta^2)^2} \end{aligned} \tag{4.50}$$

$$+ Z_2^2 Z_3 \{ \frac{4\beta}{\rho^3(1+\beta^2)^2} - \frac{\phi_{11}}{\rho}(1+\beta^2) \}$$

$$+ \frac{1}{\rho^3(1+\beta^2)^2} \{ -(1-\beta^2) Z_2^3 + Z_1^2 Z_2 - 2\beta Z_1 Z_2^2 - 2 Z_1 Z_2 Z_3 \} \quad .$$

Also let

$$\frac{\hat{e}}{\tau} = \frac{\lambda}{(1+\beta^2)}(\hat{\beta} - \beta) = Q_0 + \frac{Q_1}{\lambda} + \frac{Q_2}{\lambda^2} + R_2 \quad , \tag{4.51}$$

where R_2 is a remainder term of the order $o_p(n^{-1})$.

By the Cornish-Fisher expansion of χ^2-random variables (Cornish and Fisher [1937])

$$\begin{pmatrix} Z_1 & Z_2 \\ Z_2 & Z_3 \end{pmatrix} = \begin{pmatrix} \sqrt{2}\, y_{11} + 2\sqrt{\rho(1+\beta^2)}\,u_1 & y_{12} + \sqrt{\rho(1+\beta^2)}\,v_1 \\ y_{12} + \sqrt{\rho(1+\beta^2)}\,v_1 & \sqrt{2}\, y_{22} \end{pmatrix}$$

$$+ \frac{1}{\sqrt{n}} \begin{pmatrix} \frac{2}{3}(y_{11}^2 - \frac{5}{2}) + u_1^2 & \frac{y_{11}y_{12}}{\sqrt{2}} + u_1 v_1 \\ \frac{y_{11}y_{12}}{\sqrt{2}} + u_1 v_1 & \frac{2}{3}(y_{22}^2 - 4 + \frac{3}{2}y_{12}^2) + v_1^2 \end{pmatrix}$$

$$+ \frac{1}{n} \begin{pmatrix} \frac{y_{11}^3 - 16 y_{11}}{9\sqrt{2}} & \frac{(y_{11}^2 - 10) y_{12}}{12} \\ \frac{(y_{11}^2 - 10) y_{12}}{12} & \frac{y_{22}^3 - 25 y_{22}}{9\sqrt{2}} \end{pmatrix} + R_3 \quad , \tag{4.52}$$

where each component of u_1, v_1, y_{11}, y_{12}, and y_{22} are mutually independent standard normal random variables, and R_3 is $(1/\sqrt{n})^3$ times a polynomial of degree 4 or 2 in y_{11}, y_{12}, and y_{22} plus a remainder term, which is $O(n^{-2})$ and is $O[(\log \sqrt{n}/\sqrt{n})^4]$ in J_n.

Define a standard normal random variable as

$$W = \frac{(v_1 + \frac{y_{12}}{\sqrt{\delta}})}{\tau} \quad , \tag{4.53}$$

where $\tau = (1 + 1/\delta)^{1/2}$, W is independent of u_1, v_1, y_{11}, and y_{22}, and $z = (v_1/\sqrt{\delta} - y_{12})/\tau$. Transforming v_1 by $(\sqrt{\delta}W + z)/(1 + \delta)^{1/2}$ and y_{12} by $(W - \sqrt{\delta}z)/(1 + \delta)^{1/2}$,

$$\mathcal{E}(Q_0|W) = W, \tag{4.54}$$

$$\mathcal{E}(Q_1|W) = \beta\tau W^2 \quad , \tag{4.55}$$

$$\begin{aligned} \mathcal{E}(Q_2|W) = {} & \frac{-\frac{3}{4}W}{\tau^2} + \frac{2W}{\tau^2} + \frac{W}{\tau^2\delta} - \frac{1 - W^2}{\tau} \\ & -(1 - \beta^2)\tau^2 W^3 + 4\tau^2 W \quad , \end{aligned} \tag{4.56}$$

$$(4.57) \qquad \mathcal{E}(Q_1^2|W) = \frac{1+\delta W^2}{1+\delta} + \frac{\delta + W^2}{2(1+\delta)} + 4\tau^4 W^2 + \beta^2\tau^4 W^4$$

$$- 2(2 + \frac{1}{\delta})W^2 + \frac{a^2}{\tau^2} \quad ,$$

where the expectations are in terms of u_1, v_1, and z in the whole space, which differs from the expectation in J_n by $O(\lambda^{-4})$. Finally by Fourier inversion formulae we find (4.38). The validity of expansion is seen in the Appendix. (QED)

Lemma 4: An asymptotic expansion of the distribution of any third-order AMNU estimator as n and λ^2 increase is given by

$$(4.58) \qquad \Pr\{\frac{\hat{e}}{\tau} \le \xi\} = \Phi(\xi) - \frac{\beta\tau(\xi^2-1)}{\lambda}\phi(\xi)$$

$$+ \frac{\xi\phi(\xi)}{2\lambda^2\tau^2}\{\frac{1}{2}(\frac{3}{\tau^2}-1) - 2a^2 + 2\tau^4(1+\beta^2+\beta^2\xi^2) - \tau^4\beta^2$$

$$+ \xi^2[2\tau^4(\beta^2-1) + \frac{1}{2}(3-\frac{1}{\tau^2})] - \beta^2\tau^4\xi^4\} + O(\lambda^{-3}) \quad .$$

Proof: From (4.28), the third-order mean-unbiasedness requires

$$(4.59) \qquad (1+\beta^2)^2\phi_{11} - \beta[\frac{1}{\rho^2(1+\beta^2)} - \frac{1}{\rho}] + T_0 = 0 \quad .$$

Then differentiating (4.59) with respect to β and ρ,

$$(4.60) \qquad 2\beta\psi_1 + \psi_2 = \frac{1-\beta^2}{\rho^3(1+\beta^2)^2} - \frac{1}{\rho^2} - \frac{4\beta(1+\beta^2)\phi_{11}}{\rho} - (1+\beta^2)^2(2\beta\phi_{111} + \phi_{112}) \quad ,$$

$$(4.61) \qquad \beta^2\psi_1 + \beta\psi_2 + \psi_3 = \frac{-2\beta}{\rho^3(1+\beta^2)} + \frac{\beta}{\rho^2} - (1+\beta^2)^2(\phi_{111}\beta^2 + \phi_{112}\beta + \phi_{113}) \quad .$$

Rearranging each term, we have

$$(4.62) \qquad \begin{pmatrix}\psi_2\\ \psi_3\end{pmatrix} = \begin{pmatrix}-2\beta\\ \beta^2\end{pmatrix}\psi_1 + \frac{1}{\rho^3(1+\beta^2)^2}\begin{pmatrix}1-\beta^2\\ -\beta(\beta^2+3)\end{pmatrix} + \frac{1}{\rho^2}\begin{pmatrix}-1\\ 2\beta\end{pmatrix} + \frac{4\beta(1+\beta^2)}{\rho}\begin{pmatrix}-1\\ \beta\end{pmatrix}\phi_{11} + (1+\beta^2)^2\begin{pmatrix}-2\beta & -1 & 0\\ \beta^2 & 0 & -1\end{pmatrix}\begin{pmatrix}\phi_{111}\\ \phi_{112}\\ \phi_{113}\end{pmatrix} .$$

Then substitution of (4.62) into (4.45) gives

$$(4.63) \qquad T_1 = Z_1\left\{\frac{-2\beta}{\rho^3(1+\beta^2)^2} + \frac{\beta}{\rho^2(1+\beta^2)} - (1+\beta^2)(\beta^2\phi_{111} + \beta\phi_{112} + \phi_{113})\right\} + Z_2\left\{\frac{1+3\beta^2}{\rho^3(1+\beta^2)^2} - \frac{1+3\beta^2}{\rho^2(1+\beta^2)} - (1+\beta^2) \times \left[\frac{4\beta\phi_{11}}{\rho} + 2\beta\phi_{111} + (1-\beta^2)\phi_{112} - 2\beta\phi_{113}\right]\right\}$$

$$+ Z_3\{(1 + \beta^2)T_0 - \frac{\beta}{\rho^3} + \frac{\beta(2\beta^2 + 1)}{\rho^2(1 + \beta^2)}$$

$$+ \beta(1 + \beta^2)[\frac{4\beta\phi_{11}}{\rho} + \beta(\beta^2 + 2)\phi_{111} + \phi_{112} - \beta\phi_{113}]\} \quad .$$

Also let

$$(4.64) \qquad Q_1^* = U_1 + T_0 \;,\; Q_2^* = U_2 + T_1 \;.$$

Then similarly by the same transformation of W, and z, we have

$$(4.65) \qquad \mathcal{E}(Q_0|W) = W \quad ,$$

$$(4.66) \qquad \mathcal{E}(Q_1|W) = \beta\tau(W^2 - 1) \quad ,$$

$$(4.67) \qquad \mathcal{E}(Q_2|W) = -\frac{\frac{3}{4}W}{\tau^2} + \frac{2W}{\tau^2} + \frac{W}{\tau^2\delta} - \frac{1 - W^2}{\tau} - (1 + 3\beta^2)\tau^2 W \quad ,$$

$$(4.68) \qquad \mathcal{E}(Q_1^2|W) = \beta^2\tau^2(1 - 2W^2) + \{\frac{1 + \delta W^2}{1 + \delta} + \frac{(\delta + W^2)}{2(1 + \delta)} + 4\tau^4 W^2$$

$$+ \beta^2\tau^4 W^4 - 2(2 + \frac{1}{\delta})W^2\} + \frac{a^2}{\tau^2} \quad ,$$

where the expectations are in terms of u_1, v_1, and z in the whole space, which differs from the expectation in J_n by $O(\lambda^{-4})$. Finally by Fourier inversion formulae we find (4.58). (QED)

Proof of Theorem 1: For the ML estimator Kunitomo [1980] gives

$$\text{(4.69)} \quad \Pr\{\frac{\hat{e}_{ML}}{\tau} \leq \xi\} = \Phi(\xi) - \frac{\beta\tau\xi^2\phi(\xi)}{\lambda}$$

$$+ \frac{\xi\phi(\xi)}{2\tau^2\lambda^2}\{\frac{1}{2}(\frac{3}{\tau^2}-1)+\xi^2[2\tau^4(\beta^2-1)+\frac{1}{2}(3-\frac{1}{\tau^2})]-\beta^2\tau^4\xi^4\}$$

to terms of order N^{-1}, where e_{ML} is the standardized ML estimator. Then from Lemma 3 for any $\xi_1 \geq 0$ and $\xi_2 \geq 0$,

$$\text{(4.70)} \quad \Pr\{-\xi_1 < \frac{\hat{e}_{ML}}{\tau} \leq \xi_2\} - \Pr\{-\xi_1 < \frac{\hat{e}}{\tau} \leq \xi_2\}$$

$$= \frac{a^2}{2\tau^2\lambda^2}\{\xi_1\phi(\xi_1) + \xi_2\phi(\xi_2)\} \geq 0 \quad ,$$

to terms of order N^{-1}, where $\hat{e}$ is any standardized AMDII estimator. The equality holds if and only if $a = 0$. (QED)

Proof of Theorem 2: The asymptotic expansion of the distribution of the estimator $\hat{\beta}^*$ when n and λ^2 increase is given in Chapter 5 of Kunitomo (1981a). Then from Lemma 4, for any $\xi_1 \geq 0$ and $\xi_2 \geq 0$,

$$\text{(4.71)} \quad \Pr\{-\xi_1 \leq \frac{\hat{e}^*}{\tau} \leq \xi_2\} - \Pr\{-\xi_1 < \frac{\hat{e}}{\tau} \leq \xi_2\}$$

$$= \frac{a^2}{2\tau^2\lambda^2}\{\xi_1\phi(\xi_1) + \xi_2\phi(\xi_2)\}$$

$$\geq 0 \quad ,$$

to terms of order N^{-1} (or λ^{-2}), where $\hat{e}$ is any standardized AMNU estimator. The equality holds if and only if $a = 0$. (QED)

Proof of Theorem 6.3: Let $\hat{\beta} = \tan \hat{\theta}$. Then putting $= x + \beta\tau x^2/\lambda + \tau^2(\beta^2 + 1/3)x^3/\lambda^2 + \ldots$, we have

$$(4.72) \qquad \Pr\{\frac{\lambda}{\tau}(\hat{\theta} - \theta) \leq x\} = \Phi(x) - \frac{\delta(1+\beta^2)}{\tau\lambda}\phi(x)\{\phi_{11} - \frac{2\beta}{\rho^2(1+\beta^2)^3}\} + O(\lambda^{-2}) \quad .$$

Then the asymptotic median-unbiasedness or mean-unbiasedness requires $\phi_{11} = 2\beta/[\delta^2(1 + \beta^2)]$. Hence Theorem 3 follows from Theorem 1 and Theorem 2. (QED)

Proof of Corollaries: The proof follows immediately from the fact that

$$(4.73) \qquad \mathcal{E} L_n(\hat{\beta},\beta) = \int_0^\infty (1 - \Pr\{\sqrt{n}(\hat{\beta} - \beta) \leq y\})dh(y) - \int_\infty^0 (\Pr\{\sqrt{n}(\hat{\beta} - \beta) \leq y\})dh(y) \quad . \qquad \text{(QED)}$$

Appendix : The Validity of the Asymptotic Expansions

The purpose of this appendix is to make our derivations more rigorous. Following Anderson [1974], to control the errors of approximation we define the set J_n as (4.18). Then Anderson showed that $\Pr(J_n) = 1 - O(n^{-2})$, and we shall ignore the tail probability in J_n^c. A Taylor expansion of $\hat{\beta}$ about the probability limits of (s_{yy}, s_{xy}, s_{xx}) gives

$$\hat{\beta} = \beta + \sum_{j=1}^{4} n^{-j/2} \beta^{(j)} + r \quad , \tag{A.1}$$

where each element of $\beta^{(j)}$ is a homogeneous polynomial of degree j in the elements of v_1, u_1^2, v_1^2, $u_1 v_1$, and y_{ij}, and r is the usual remainder which is $O[(\log n/n)^{5/2}]$ uniformly in J_n and is $O(n^{-5/2})$ for fixed u_1, v_1, and y_{ij}.

We write

$$\frac{\hat{e}}{\tau} = \frac{\sqrt{n}(\hat{\beta} - \beta)}{\tau} = W + \frac{e^{(1)}}{\sqrt{n}} + \frac{e^{(2)}}{n} + R \quad , \tag{A.2}$$

where each element of $e^{(j)}$ is a homogeneous polynomial of degree $j + 1$ in the elements of $u_1, W, u_1 W, u_1^2$, and $z, u_1 z$ and y_{ii} $(i=1,2)$, R is a remainder term which is $O(n^{-3/2})$ and is $O[(\log n/n)^{-3/2}]$ uniformly in J_n. Let

$$\hat{C}(t) = \mathcal{E}(A \exp(it\, W)) \ . \tag{A.3}$$

where $A = 1 + (1/\sqrt{n})ite^{(1)} + (1/n)\{ite^{(2)} + (ite^{(1)})^2/2\}$.

We know that $|\exp(it\hat{e}/\tau) - A(itW)|$ is bounded by $|B|^3/6 + |1 + B + B^2/2 - A|$ and hence is $0(n^{-3/2})$ in J_n where $B = it\{e^{(1)}/\sqrt{n} + e^{(2)}/n + R\}$, we have

$$(A.4) \qquad C(t) - \hat{C}(t) = 0(n^{-3/2})$$

where $C(t)$ is the characteristic function of $\hat{e}/\tau$.

To complete the justification of our formal derivations, we need to show that the Fourier inverse transform of the terms $0(n^{-3/2})$ in (A.3) is $0(n^{-3/2})$. We use the existence of a valid asymptotic expansion for the distribution function of $\hat{e}/\tau$ such that

$$(A.5) \qquad \Pr\{\hat{e}/\tau \leq a\} = \int_{\xi \leq a} \hat{f}(\xi)d\xi + 0(n^{-3/2}) \quad ,$$

where $\hat{f}(\xi) = \phi(\xi) + f_1(\xi)/\sqrt{n} + f_2(\xi)/n$ and $f_i(\xi)$ are polynomial multiples of $\phi(\xi)$ whose coefficients do not depend on n. We omit the details of the proof of the existence of (A.5), which can be done by applying arguments similar to those in Anderson [1974], Sargen [1975], and Phillips [1977]. To sketch the outline, by a valid expansion $\hat{g}(Y)$ for the density function of $\underset{\sim}{Y} = (y_{ij})$ from Battacharya and Ghosh [1978], we can write

$$(A.6) \qquad \Pr\{\frac{\hat{e}}{\tau} \leq a\} = \int_{\hat{e} \leq a} h(v_1)\hat{g}(Y)dv_1 dY + 0(n^{-3/2}) \quad ,$$

where $\hat{g}(\underset{\sim}{Y}) = g_0(\underset{\sim}{Y}) + g_1(\underset{\sim}{Y})\sqrt{n} + g_2(\underset{\sim}{Y})/n$, $g_0(\underset{\sim}{Y})$ is the two-dimensional normal density, and $g_1(Y)$ are polynomial multiples of $g_0(\underset{\sim}{Y})$. (See (4.52) under Assumption A.) Making the change of variables from $(v_1 \ \underset{\sim}{Y})$ to $(\dot{W}, z, y_{11}, y_{22})$ in J_n with large enough n and then integrating with respect to (z, y_{11}, y_{22}), we will obtain (A.5). From (A.5) we have

(A.7) $$\hat{C}(t) = \mathcal{E}(\exp\,(it\xi)\hat{f}(\xi)) + O(n^{-3/2})$$

for any fixed t. Then (A.4) and (A.7) imply that $\hat{C}(t) = \exp\,(it\xi)\hat{f}(\xi)$ and hence $\hat{f}(\xi)$ and the inverse Fourier transform of $C(t)$ are identical. This gives the validity of our asymptotic expansion.

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 Lieklihood Estimate of a Coefficient in a Simultaneous Equation System," Journal of the American Statistical Association, 69, 565-572.

Anderson, T.W. [1976]. "Estimation of Linear Functional Relationships: Approximate Distributions and Connections with Simultaneous Equations in Econometrics," Journal of the Royal Statistical Society, Series B 38, 1-38.

Anderson, T.W., Naoto Kunitomo and Takamitsu Sawa. [1981]. "Evaluation of the Distribution Function of the Limited Information Maximum Likelihood Estimator," Forthcoming in Econometrica.

Anderson, T.W. and Takamitsu Sawa. [1979]. "Evaluation of the Distribution Function of the Two-Stage Least Squares Estimate," Econometrica, 47, 163-183.

Battacharya, R.N. and J.K. Ghosh. [1978]. "On the Validity of the Formal Edgeworth Expansion," The Annals of Statistics, 6, 434-451.

Cornish, E.R. and R.A. Fisher. [1937]. "Moments and Cumulants in the Specification of Distributions," Review of International Statistical Institute, 5, 307.

Efron, Bradley, J.K. Ghosh, L. Le Cam, Johann Pfanzagle and Radhakrishna Rao. [1980]. "Discussion of Professor Berkson's Paper by Efron, Bradley, et al," The Annals of Statistics, 8, 469-487.

Fuller, W.A. [1977]. "Some Properties of a Modification of the Limited Information Estimator," Econometrica, 45, 939-953.

Ghosh, J.K., B.K. Sinha and H.S. Wieand. [1980]. "Second Order Efficiency of the MLE With Respect to Any Bounded Bowl-Shaped Loss Function," Annals of Statistics, 8, 506-521.

Kadane, Joseph B. [1971]. "Comparison of k-class Estimators When the Disturbances Are Small," Econometrica, 65, 723-737.

Kendall, M.G., and A. Stuart. [1973]. The Advanced Theory of Statistics, Vol. 2, London: Charles Griffin and Company.

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. [1981a]. "Asymptotic Efficiency and Higher Order Efficiency of the Limited Information Maximum Likelihood Estimator in Large Econometric Models," unpublished Ph.D. Dissertation, Stanford University.

Kunitomo, Naoto. [1981b]. "Asymptotic Optimality of the Limited Information Maximum Likelihood Estimator in Large Econometric Models," forthcoming in the Economic Studies Quarterly.

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 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.

Sargan, John D. [1975]. "Gram-Charlier Approximation Applied t-Ratios of k-Class Estimators," Econometrica, 43, 327-346.

Sawa, Takamitsu. [1973]. "Almost Unbiased Estimators in Simultaneous Equations System," International Economic Review, 14, 97-106.

Takeuchi, Kei. [1972]. Contributions to the Theory of Statistical Inference in Econometrics, Tokyo: Toyokeizai-Shinposha. (In Japanese).

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.