Esfandiar Maasoumi [1]

THE APPROXIMATE MOMENTS OF THE 3SLS REDUCED FORM ESTIMATOR AND A MELO COMBINATION OF OLS-3SLS FOR PREDICTION

1. INTRODUCTION

This paper develops the small sample approximations to the moments of the 3SLS *reduced form* estimator in a general linear simultaneous equations model under the classical assumptions. These approximations easily specialize to the case of k-class reduced form estimators, and they may be used to evaluate conditional forecasts based on the k-class or 3SLS estimators.

The method of derivation and its interpretation are discussed in Section 2. It can be seen that the techniques are more generally applicable and may be used to derive the approximate moments of any standard reduced form estimator.

Sections 2 and 3 of the paper discuss examples of how these approximate expressions have been used, for instance, to evaluate estimators or mixtures of estimators under quadratic loss and other moment-based criteria. In particular, an approximate Minimum Expected Loss (MELO) mixture of unrestricted OLS and 3SLS reduced form estimators is described.

Derivation of asymptotic expansions for the moments of econometric estimators began with Nagar (1959) who dealt with the k-class estimators of the *structural* coefficients in a linear simultaneous equations model. Nagar also showed how his formula for the bias may be used to develop "almost unbiased" estimators. Sawa (1973a) developed the "almost unbiased" application of these formulae, and (e.g.) Nagar and Carter (1976), Sawa (1973b) and Zellner and Vandaele (1975) have utilized these approximations to develop Minimum Variance, Minimum Mean Squared Error (MMSE) and MELO estimators. Such expansions to estimator biases have more recently been

[1] Department of Economics, Indiana University, Bloomington, Indiana 47405, also University of Southern California.

I. B. MacNeill and G. J. Umphrey (eds.), Time Series and Econometric Modelling, 359–371.

used by Rothenberg (1984) and others for correction of ML, GLS and other statistics in order to discuss questions of second and higher order efficiency.

Considerably less attention has been given to estimators and forecasts in the ***reduced form*** context. While asymptotic properties of the reduced form estimators can be evaluated simply by the methods of Goldberger *et al.* (1961) and Dhrymes (1973), higher order approximations do not seem to have been developed for many reduced form estimators. One exception is the work of Nagar and Sahay (1978), who were concerned with the Partially Restricted Reduced Form and the 2SLS estimators. The existence of the moments of several reduced form estimators is discussed by McCarthy (1972), Sargan (1976b), Maasoumi (1977, 1978, 1985, 1986), and Knight (1977).

2. APPROXIMATE MOMENTS OF THE 3SLS

Following Rothenberg (1984), we consider a random sample of size T from a population with a continuous density function which depends on an unknown parameter θ. Let $\hat{\theta}_T$ be the standardized estimator of θ and $F_T(k) = Pr[\hat{\theta}_T \leq k]$. For estimators which are Best Asymptotic Normal (BAN) and admit an Edgeworth expansion to order T^{-1}, we can write:

$$F_T(k) = F(k) + O(T^{-1}), \tag{1}$$

$$F(k) = \eta(k) + \frac{\mu(k)}{\sqrt{T}} + \frac{\gamma(k)}{T}. \tag{2}$$

In (2), $\eta(\cdot)$ is the standard normal distribution function and $\mu(\cdot)$ and $\gamma(\cdot)$ are usually polynomials multiplied by $\eta(k)$. η is referred to as a "first order" approximation, and F a "second order" approximation to F_T. Consequently, estimators that satisfy (1)–(2) are first-order efficient and so may be compared on the basis of the moments of F, i.e., the "second-order" moments. In this section we derive the moments of $F(\cdot)$ for the case of the 3SLS reduced form estimator. These approximate moments are thus well defined as the moments of the distributions which approximate the exact distribution, $F_T(\cdot)$. [The moments of η are first-order approximate moments of F_T in exactly the same sense. The only distinction with the traditional "asymptotic theory" is therefore one of degree of approximation and not of interpretation.]

The validity of Edgeworth approximations to F_T is discussed by Sargan (1976a) and generally does not depend on the *existence* of the moments. When the moments of F_T do not exist, however, care must be taken in interpreting the moments of F (or η) as approximations to the moments of F_T. In this situation the value of the approximate moments derives from the value of F (or η) as a representation of F_T. To the extent that this is done adequately by F, the study of its moments is of value in characterizing both F *and* F_T. We emphasize that, once again, there is no distinction between the interpretations of the moments of the asymptotic distribution of estimators and those of higher order approximate distributions.

The Linear Simultaneous Equations Model (SEM) is defined for Y $(T\times n)$ endogenous variables and Z $(T \times m)$ non-stochastic exogenous variables as follows:

$$AX' = BY' + \Gamma Z' = U', \tag{3}$$

where $X = (Y, Z)$ is a $T\times(n+m)$ matrix of all observations, $A = (B, \Gamma)$ is the $(n\times(n+m))$ matrix of the unknown coefficients, and U is the $(T\times n)$ matrix of the random disturbances such that each row, U_t, satisfies the following assumption:

$A1: \ U_t \sim \text{i.i.d.}(0, \Sigma)$.

Further assumptions of the classical SEM are:

$A2: \ \lim_{T\to\infty} T^{-1}Z'Z = M$, a constant matrix of rank m;

$A3: \ B$ is non-singular.

The reduced form of (3) is given by:

$$Y' = PZ' + V', \tag{4}$$

where $P = -B^{-1}\Gamma$, $V' = B^{-1}U'$ and, from $A1$, rows of V have zero mean and a common covariance matrix $\Omega = B^{-1}\Sigma^{-1}B'^{-1}$. When we need to, we assume that the *a priori* (identifying) restrictions on A are of the exclusion (zero order) type that may be represented as follows:

$$s - S\boldsymbol{\alpha} = \text{Vec } A. \tag{5}$$

Here Vec denotes stacking by *rows*, $\boldsymbol{\alpha}$ is the vector of the non-zero elements of A, $S = \text{diag}(S_1, \ldots, S_n)$ such that $XS_i = X_i$ represents the selection from X of only those columns that appear on the right hand side of the ith equation. The selection vector s represents the "normalization" restriction since its ith subvector, s_i, is such that $Xs_i = y_i$, the left hand side endogenous variable of the ith equation.

We define the following estimators of P:

$$\hat{P} = (Y'Z)(Z'Z)^{-1} \tag{6}$$

is the Unrestricted Least Squares (ULS) estimator of P in (4). Under the classical assumptions it is unbiased but less efficient than some Restricted Reduced Form (RRF) estimators,

$$P^+ = -B^{+^{-1}}\Gamma^+, \tag{7}$$

where B^+ and Γ^+ may be such full information estimators as 3SLS or FIML. In this paper I assume that the estimator A^+ satisfies the following property:

$$A4: \ \Delta A = A^+ - A = O_p(T^{-1/2}).$$

Define ΔB and $\Delta\Gamma$ as ΔA was defined. Then:

$$\begin{aligned} P^+ &= -(B + \Delta B)^{-1}(\Gamma + \Delta\Gamma) \\ &= -(B^{-1}\Delta B + I)^{-1} B^{-1}(\Gamma + \Delta\Gamma). \end{aligned} \tag{8}$$

Expanding $(B^{-1}\Delta B + I)^{-1}$,

$$\begin{aligned} (B^{-1}\Delta B + I)^{-1} = I - B^{-1}\Delta B + B^{-1}\Delta B \cdot B^{-1}\Delta B \\ + O_p(T^{-3/2}) \end{aligned} \tag{9}$$

and using this in (8), we have:

$$\begin{aligned} \Delta P = {} & P^+ - P \\ = {} & -B^{-1}\Delta\Gamma + B^{-1}\Delta B \; B^{-1}\Delta\Gamma + B^{-1}\Delta B \cdot B^{-1}\Gamma \\ & - B^{-1}\Delta B \cdot B^{-1}\Delta B \cdot B^{-1}\Delta\Gamma \\ & - B^{-1}\Delta B \cdot B^{-1}\Delta B \cdot B^{-1}\Gamma + O_p(T^{-2}). \end{aligned} \tag{10}$$

If one defines $Q = \left(\frac{P}{Im}\right)$ and rearrange, one finds that:

$$\Delta P = -B^{-1}\Delta A \cdot Q + B^{-1}\Delta B \cdot B^{-1}\Delta A \cdot Q + O_p(T^{-3/2}). \tag{11}$$

Let $b^+ = E$ (Vec ΔP) and $V(p^+)$ denote, respectively, the bias and the variance-covariance matrix of $p^+ =$ Vec P^+. Then:

$$V(p^+) = E\left[(\text{Vec}\Delta P)(\text{Vec}\Delta P)'\right] - b^+ b^{+\prime}, \tag{12}$$

where the first term on the r.h.s. of (12) is the MSE of p^+. Approximations to b^+ are obtained by taking expectations of (11), term by term. For the first term, we note that

$$\text{Vec}\left(-B^{-1} \cdot \Delta A \cdot Q\right) = -\left(B^{-1} \otimes Q'\right) \text{Vec}\Delta A, \tag{13}$$

where $\otimes$ denotes the Kronecker product. The r.h.s. of (13) is the basic relationship used by (e.g.) Dhrymes (1973) and Goldberger *et al.* (1961) to obtain the asymptotic properties of P^+ from those of A^+. We will employ higher order expansions of E(Vec ΔA) in (13), and will combine these with the expectation of the second term of (11). To obtain the latter expectation, it is easier to work first with simpler linear functions of the required terms. We will thus first obtain:

$$E\left[\operatorname{tr}\left(\Phi B^{-1}\Delta B\cdot B^{-1}\Delta A\cdot Q\right)\right] \tag{14}$$

for a known arbitrary $n \times n$ matrix Φ, and then recover the desired expectation from (14). We first note that:

$$\begin{aligned}(14) &= E\left\{\operatorname{tr}\left[B'^{-1}\Delta B'\left(Q\Phi B^{-1}\right)'\Delta A'\right]\right\}\\ &= E\left\{(\text{Vec }\Delta A)'\left(B'^{-1}\otimes Q\Phi B^{-1}\right)(\text{Vec }\Delta B')\right\}.\end{aligned} \tag{15}$$

Let Π be a permutation matrix such that, for a matrix D,

$$\text{Vec } D' = \Pi\cdot\text{Vec } D. \tag{16}$$

We also define the following "slash" product, $\oslash$, for any two matrices F_1 and F_2:

$$(F_1\oslash F_2) = (F_1\otimes F_2)\Pi. \tag{17}$$

Subsequently, using (15)–(17), we conclude:

$$\begin{aligned}(14) &= E\left[(\text{Vec }\Delta A)'\left(B'^{-1}\oslash Q\Phi B^{-1}\right)(\text{Vec }\Delta B)\right]\\ &= \operatorname{tr}\left\{\left(B'^{-1}\oslash Q\Phi B^{-1}\right)\cdot E\left[(\text{Vec }\Delta B)(\text{Vec }\Delta A)'\right]\right\}.\end{aligned} \tag{18}$$

The expression inside { } and, in particular, the covariance between ΔB and ΔA is to be evaluated. Let the asymptotic variance of ΔA be denoted by G. Then (an approximation to) the variance of ΔA is given by

$$E\left[(\text{Vec }\Delta A)(\text{Vec }\Delta A)'\right] \doteq \frac{1}{T}G, \tag{19}$$

where G is $n(n+m)\times n(n+m)$ and, (e.g.) for 3SLS, we have

$$G = S\left[S'\left(\Sigma^{-1}\otimes\overline{R}\right)S\right]^{-1}S', \tag{20}$$

where $\overline{R} = QMQ'$; see Sargan (1978) or Maasoumi (1978). [It is not necessary at this stage to consider approximations to the variance of ΔA. But this approximation is inevitable at the final stage.] Define $\overline{\psi} = (I_n 0_m)$ and $\psi^* = (I\otimes\overline{\psi})$. It follows that:

$$\text{Vec }\Delta B = \psi^*\cdot\text{Vec }\Delta A$$

and

$$E\left[(\text{Vec}\ \Delta B)(\text{Vec}\ \Delta A)'\right] \doteq \frac{1}{T}\psi^{*}G = \overline{G}, \quad \text{say.} \tag{21}$$

Using (21) in (18), we may assert:

$$\begin{aligned}(14) &\doteq \text{tr}\left\{\left[B'^{-1} \otimes Q\Phi B^{-1}\right]\overline{G}\right\} \\ &= \text{tr}\left\{\left[B'^{-1} \otimes Q\Phi B^{-1}\right]\Pi_n^n\overline{G}\right\},\end{aligned} \tag{22}$$

where

$$\Pi_n^n\overline{G} = \begin{pmatrix} \overline{G}_{11} & \cdots & \overline{G}_{1n} \\ \vdots & & \vdots \\ \overline{G}_{n1} & \cdots & \overline{G}_{nn} \end{pmatrix} \tag{23}$$

with each submatrix $\overline{G}_{ij}$ having $n \times (n+m)$ dimensions. From (22) it may be verified that:

$$(14) \doteq \sum_{j=1}^{n} b^{jj} \cdot \text{tr}\left(Q\Phi B^{-1}\overline{G}_{jj}\right), \tag{24}$$

where b^{ii} is the ith diagonal element of B'^{-1}, and noting that $E[\text{tr}(\cdot)] = \text{tr}[E(\cdot)]$, we have:

$$E\left(B^{-1}\Delta B \cdot B^{-1}\Delta A \cdot Q\right) \doteq \sum_{j=1}^{n} b^{jj}B^{-1}\overline{G}_{jj}Q. \tag{25}$$

Gathering terms from (25) and (13), we have:

$$b^{+} = \left(B^{-1} \otimes Q'\right)\left[\sum_{j=1}^{n} b^{jj}(\text{Vec}\ \overline{G}_{jj}) - E(\text{Vec}\ \Delta A)\right] + \ldots. \tag{26}$$

Approximations to Vec $\overline{G}_{jj}$ and $E(\text{Vec}\ \Delta A)$ should retain all terms of order T^{-1} or larger in order to correspond to (11). Lower order approximations for b^{+} will necessarily omit the first term of (26). It remains to replace $E(\text{Vec}\ \Delta A)$ with the approximate bias of A^{+}. For a variety of structural estimators, such as k-class, FIML and 3SLS, these approximations have been given in the literature. For 3SLS, in particular, these are given by Sargan (1976a). To order T^{-2}, the bias is as follows:

$$\begin{aligned}E(\text{Vec}\ \Delta A) &\doteq mG\left(\Sigma^{-1} \otimes I\right)q + \Psi'\left(I \otimes \overline{Q}_s\right)q \\ &\quad + \overline{Q}q - 2G\left(\Sigma^{-1} \otimes I\right)\Psi^{T}q - G\left[\left(\Sigma^{-1}H\Sigma^{-1}\right) \otimes I\right]q \\ &\quad - G\left(\Sigma^{-1} \otimes \Psi_s\right)q,\end{aligned} \tag{27}$$

where Ψ^T is the block transpose of $\Psi = (\Sigma^{-1} \otimes \overline{X}'\overline{X})G$, $\Psi_s = \sum_j \Psi_{jj}$, $\overline{X} = (\overline{Y}\ Z)$, $\overline{Y} = Y - V$, $q = (q_1', q_2', \ldots, q_n')'$, $q_i = 1/T\ E(V^{*\prime}u_i)$, $V^* = (UB'^{-1}0)$ is $T \times (n+m)$, u_i is the T-element vector of disturbances in the ith equation, $H = \mathrm{tr}[G(I \otimes \overline{X}'\overline{X})]$, and $\overline{Q} = \mathrm{diag}.(\overline{Q}_1, \overline{Q}_2, \ldots, \overline{Q}_n)$, where $\overline{Q}_i$ is obtained from $(\overline{X}_i'\overline{X}_i)^{-1}$ by adding rows and columns of zeros corresponding to the excluded variables in the ith equation, and $\overline{Q}_S = \sum_{i=1}^{n} \overline{Q}_i$.

The above techniques may be used to obtain higher order approximations to the second and higher order moments of P^+. These expressions can be used to obtain almost unbiased estimators, and are also useful in comparing higher order efficiency of the estimators. We will proceed to demonstrate their use by considering an "optimal" mixture of the ULS estimator ($\hat{P}$) and the 3SLS (P^+) reduced form estimators under quadratic loss. The latter (or its expectation) is approximated using the moment expansions given above.

2. APPROXIMATE MINIMUM EXPECTED LOSS (MELO) COMBINATIONS OF OLS-3SLS

This and the next section are based on Maasoumi (1985).

Let $\hat{P} = Y'Z(Z'Z)^{-1}$ denote the ULS estimator of P and $P^+ = -B^{+-1}\Gamma^+$ another estimator derived from such restricted estimators of B and Γ as 2SLS, 3SLS, FIML, etc. We propose the following mixed estimator of P:

$$P^* = \lambda\hat{P} + (1-\lambda)P^+ \tag{28}$$

$$= P^+ + \lambda(\hat{P} - P^+). \tag{29}$$

In the remainder of this paper lower case letters p, $\hat{p}$, p^+ denote vec P, vec $\hat{P}$ and vec P^+, respectively. Let the bias in p^* be denoted by b^*. It is readily seen that, since $E(\hat{p}) - p = 0$,

$$|b^*| = |E(p^*) - p| = |(1-\lambda)b^+| \le |b^+|, \qquad \text{if } 0 \le \lambda \le 1. \tag{30}$$

Using the following identity, the variance matrix of p^*, $V(p^*)$, is obtained in terms of the variances of $\hat{p}$, p^+ and their covariance:

$$p^* - E(p^*) = \lambda(\hat{p} - p) + (1-\lambda)(p^+ - E(p^+)); \tag{31}$$

$$V(p^*) = \lambda^2 V(\hat{p}) + (1-\lambda)^2 V(p^+) + 2\lambda(1-\lambda)\mathrm{cov}(\hat{p}, p^+). \tag{32}$$

Under the standard assumptions of our model, and since the mixing parameter λ is a constant, the bias and variance of p^* would be finite if p^+ has finite moments. For instance, when p^+ represents the FIML estimator, b^*

and $V(p^*)$ are finite so long as $T - n - m \geq 2$; see Sargan (1976b). If p^+ represents either 2SLS or 3SLS, then b^* and $V(p^*)$ will not be finite unless $\lambda = 1 (p^* \equiv \hat{p})$. For exactly identified models $P^* = \hat{P}$ since $P^+ \equiv \hat{P}$ in that case.

Since $\hat{p}$ is a consistent estimator under our assumptions, it is seen that p^* will be consistent if p^+ is consistent. If not, the inconsistency in p^* will be smaller than that in p^+ as long as $\lambda \in [0,1]$. If both $\hat{p}$ and p^+ are inconsistent but have the same limit in probability, then p^* will also be inconsistent with the same plim as $\hat{p}$ (or p^+).

As for asymptotic efficiency, the derivations given in the next section may be used to verify that:

$$\begin{aligned} AV(p^*) &= \lambda^2 AV(\hat{p}) + (1 - \lambda^2) AV(p^+) \\ &= AV(p^+) + \lambda^2 \left[AV(\hat{p}) - AV(p^+) \right], \end{aligned} \tag{33}$$

where $AV(\cdot)$ denotes the asymptotic variance. From (33) it is clear that p^* is more efficient than the ULS so long as $\hat{p}$ is less efficient than the restricted estimator p^+. While this is the case for the full information estimators such as the 3SLS and FIML, it is not always so for the limited information estimators such as 2SLS and LIML; see Dhrymes (1973). The latter statement holds even if p^+ is replaced by the Partially Restricted Reduced Form (PRRF) estimator of Amemiya (1966) and Kakwani and Court (1972). For while PRRF has finite moments (see Knight, 1977), it is not necessarily more asymptotically efficient than 2SLS. On the other hand, p^* is less efficient than 3SLS and FIML, but can be more efficient than (e.g.) 2SLS whenever $\hat{p}$ is.

3. MIXED PREDICTION UNDER QUADRATIC LOSS

Let $Y_f^* = P^* Z_f$ be the predictor of Y_f conditional on Z_f under the assumption that $Y_f = P Z_f + v_f$, where v_f denotes the forecast period random disturbance with the same properties as v_t, $t = 1, \ldots, T$. The forecast error and a general quadratic loss are given as follows:

$$Y_f^* - Y_f = (P^* - P) Z_f - v_f, \tag{34}$$

and

$$\begin{aligned} L(Y_f^*, Y_f) &= (Y_f^* - Y_f)' W^* (Y_f^* - Y_f) \\ &= \operatorname{tr} \left[W^* (Y_f^* - Y_f)(Y_f^* - Y_f)' \right], \end{aligned} \tag{35}$$

where W^* is a symmetric, positive definite matrix of known weights. From (34)–(35) the expected loss (risk) is derived as follows:

$$\begin{aligned} R(Y_f^*) &= E[L(\cdot)] \\ &= \mathrm{tr}[W \mathrm{MSE}(P^*)] + \mathrm{tr}[W\Omega], \end{aligned} \tag{36}$$

where $W = (W^* \otimes Z_f Z_f')$ and $\mathrm{MSE}(P^*)$ is the MSE matrix of p^*. Since the second term of (36) is common to all conditional forecasts, we focus on the first term which is a well known ***estimation*** risk function. Consequently, minimization of $R(Y_f^*)$ is equivalent to minimization of $R(p^*) = \mathrm{tr}[W\mathrm{MSE}(p^*)]$ with respect to λ. We note that, from (30) to (32):

$$\begin{aligned} R(p^*) &= \lambda^2 \mathrm{tr}[WV_0] + (1-\lambda)^2 \mathrm{tr}[WV^+] \\ &\quad + 2\lambda(1-\lambda)\mathrm{tr}[W\mathrm{cov}(\hat{p}, p^+)] + (1-\lambda)^2 b^{+\prime} W b^+, \end{aligned} \tag{37}$$

where $V_0 = V(\hat{p})$ and $V^+ = V(p^+)$. To minimize $R(p^*)$ with respect to λ consider

$$\begin{aligned} \frac{\partial R(\cdot)}{\partial \lambda} &= 2\lambda \ \mathrm{tr}[WV_0] - 2(1-\lambda)\ \mathrm{tr}[WV^+] \\ &\quad + 2(1-2\lambda)\ \mathrm{tr}[W\ \mathrm{cov}(\cdot)] - 2(1-\lambda) b^{+\prime} W b^+ \end{aligned} \tag{38}$$

and

$$\begin{aligned} \frac{\partial^2 R(\cdot)}{\partial \lambda^2} &= 2\ \mathrm{tr}[WV_0] + 2\ \mathrm{tr}[WV^+] - 4\ \mathrm{tr}[W\ \mathrm{cov}(\cdot)] + 2b^{+\prime} W b^+ \\ &= 2\ \mathrm{tr}[WV(\hat{p} - p^+)] + 2b^{+\prime} W b^+ \geq 0, \end{aligned} \tag{39}$$

where $V(\hat{p} - p^+)$ denotes the variance of $(\hat{p} - p^+)$. From (38), the optimal value λ_1^* of λ is obtained by solving $\partial R(\cdot)/\partial\lambda = 0$,

$$\lambda_1^* = \frac{b^{+\prime} W b^+ + \mathrm{tr}\left[W(V^+ - \mathrm{cov}(\cdot))\right]}{b^{+\prime} W b^+ + \mathrm{tr}\left[WV(\hat{p} - p^+)\right]}. \tag{40}$$

It may be observed that the denominator of λ_1^* is non-negative, and $\lambda_1^* \leq 1$ if $[V_0 - \mathrm{cov}(\hat{p}, p^+)]$ is positive semidefinite. In what follows we demonstrate that this condition, as well as the range of possible values for λ_1^*, depends on the level of approximation considered for the otherwise unknown moments entering in (40). Equivalently, these issues depend on the order of finite sample approximations for the $L(\cdot)$ and $R(\cdot)$ functions.

Strictly speaking, since 2SLS and 3SLS reduced form estimators have no finite moments, $\lambda_1^* = 1 \Rightarrow p^* = \hat{p}$ is the only member of the corresponding mixtures that has finite quadratic risk. For FIML, on the other hand, all the corresponding mixtures have finite risk if $T-n-m \geq 2$. In either case, when we consider Nagar-type approximations to these moments we are in effect evaluating the risk functions with respect to finite sample *approximations* to the exact sampling distributions of the p^+ estimator.

We note the following well known results:

$$\operatorname*{plim}_{T\to\infty} T(\hat{p}-p)(\hat{p}-p)' = (\Omega \otimes M^{-1}) = \lim\ TV_0 = O(1). \tag{41}$$

In other words, under our standard assumptions, $V_0 = O(T^{-1})$. When p^+ is the 3SLS estimator, it follows that:

$$\begin{aligned}\operatorname*{plim}_{T\to\infty} T(p^+-p)(p^+-p)' &= (B^{-1} \otimes Q')G(B'^{-1} \otimes Q)\\ &= O(1)\\ &= \lim\ TV_a^+, \quad \text{say.}\end{aligned} \tag{42}$$

Consequently $V_a^+ = O(T^{-1})$. Maasoumi (1978) has shown that, if p^+ denotes the 3SLS estimator,

$$\operatorname*{plim}_{T\to\infty} T(p^+-p)(\hat{p}-p^+)' = 0 \tag{43}$$

and

$$\operatorname*{plim}_{T\to\infty} T(p^+-p)(\hat{p}-p)' = \lim\ TV_a^+ = O(1). \tag{44}$$

The *asymptotic* properties given in (43)–(44) hold for both the 3SLS and FIML estimators and may also be deduced from a Rao-Blackwell lemma; see, for example, Hausman (1978). They do not hold for the asymptotically less efficient 2SLS or LIML reduced form estimators.

In λ_1^*, if we replace all terms with their $O(T^{-1})$ approximations and utilize the results in (41)–(44), we find

$$\lambda_2^* = b_a^{+\prime} W b_a^+ / \left\{\mathrm{tr}[W(V_0 - V_a^+)] + b_a'^+ W b_a^+\right\}, \tag{45}$$

where b_a^+ is the approximate bias of p^+ obtained by retaining terms of $O_p(T^{-1/2})$ in the expansion of $p^+ - p$. We note that, if $b_a^+ = O(T^{-1/2})$, $\lambda_2^* = O(1)$ since $V_0 = (\Omega \otimes (Z'Z)^{-1}) = O(T^{-1})$ and

$$\begin{aligned}V_a^+ &= (B^{-1} \otimes Q')\, S \left[S' \left(\Sigma^{-1} \otimes \hat{R}\right) S\right]^{-1} S' \left(B'^{-1} \otimes Q\right)\\ &= O(T^{-1}),\end{aligned} \tag{46}$$

where $\hat{R} = (X'Z)(Z'Z)^{-1}(Z'X)$, and under these conditions we have:

$$0 \leq \lambda_2^* \leq 1 \tag{47}$$

whenever $V_0 - V_a^+$ is non-negative definite. This last condition is clearly satisfied for the full information estimators which permitted the simple formula in (45).

The mixing parameter λ_2^* has several desirable properties:

(i) As the efficiency gain of the restricted estimator over the ULS decreases, $\lambda_2^* \to 1$ and the corresponding mixed estimator (p^*) moves closer to the simple ULS estimator.

(ii) As the bias of the efficient estimator increases, $\lambda_2^* \to 1$ and $p^* \to \hat{p}$. This is evidently desirable since this bias would be large either due to structural misspecification or due to poor finite sample properties of the efficient estimator (even as judged by its approximate distribution), or both. On the other hand, $p^* \to p^+$ as $\lambda_2^* \to 0$ which occurs as $b^+ \to 0$.

(iii) The formula for λ_2^* is seen to provide a mechanism for pooling of estimators (predictors) which accounts for the efficiency-bias trade-offs.

(iv) Under correct specification p^+ is a consistent estimator. Therefore, when the sample size is "sufficiently" large it is reasonable to expect b_a^+ to be close to zero. This will also pool the mixed estimator toward the asymptotically desirable estimator (p^+). This pattern of large sample behavior for b^+ has been confirmed by numerous Monte Carlo studies—for example, see Maasoumi (1977), and Rhodes and Westbrook (1980).

The formula given for λ_1^* in (40) may of course be approximated at a higher level. It can be verified that the next possible degree of approximation will retain terms of $O(T^{-1})$. The resulting value for λ will behave more like λ_1^* while exhibiting only some of the properties enumerated for λ_2^*. While these higher order expressions may be computed from Section 2 of this paper and the moments given by Sargan (1976a), *improved* approximation is by no means guaranteed by the additional terms. Some have argued that if $O(T^{-1})$ terms are of significance then the sample size is too small to allow reliable inferences in reasonably sized simultaneous systems. Nevertheless, there is a higher level of approximation for b^+ that results in an interesting variant of λ_2^*. This is obtained from λ_2^* by replacing b_a^+ with an $O(T^{-1})$ approximation of b^+ given in Section 2, and maintaining the $O(T^{-1})$ approximations for variances and covariances. This approximation produces a mixing parameter, λ^*, which is $O(T^{-1})$, and therefore a mixed estimator (predicator) which is asymptotically equivalent to the asymptotically desirable method based on p^+. A further justification for this choice of λ is that, if p^+ is consistent, the odd order terms ($O_p(T^{-1/2})$, $O_p(T^{-3/2})$ etc.) in the expansion of $p^+ - p$ have zero expectations under the normality assumption and may be dropped in obtaining an $O(T^{-1})$ approximation for b^+.

ACKNOWLEDGMENTS

I wish to thank the Centre of Decision Sciences and Econometrics, Department of Economics, The University of Western Ontario, for their invitation and financial support to attend the Symposia in Statistics and the Festschrift in honor of Professor V. M. Joshi. In particular, I should like to thank Professors R. A. L. Carter and Aman Ullah for their hospitality. I am also grateful to Mrs. Alice Gilmore for her skillful typing.

REFERENCES

Amemiya, T. (1966), "On the use of principal components of independent variables in two-stage least-squares estimation". *International Economic Review* **7**, 283–303.

Dhrymes, P. (1973), "Restricted and unrestricted reduced forms: asymptotic distribution and relative efficiency". *Econometrica* **41**, 119–134.

Goldberger, A., A. L. Nagar, and H. S. Odeh (1961), "The covariance matrices of reduced form coefficients and of forecasts in a structural model". *Econometrica* **29**, 556–573.

Hausman, J. (1978), "Specification tests in econometrics". *Econometrica* **46**, 1251–1271.

Kakwani, N. C., and R. H. Court (1972), "Reduced form coefficient estimation and forecasting from a simultaneous equation model". *Australian Journal of Statistics* **14**, 143–160.

Knight, J. L. (1977), "On the existence of moments of the partially restricted reduced-form estimators from a simultaneous-equation model". *Journal of Econometrics* **5**, 315–321.

Maasoumi, E. (1977), "A study of improved methods of estimating the reduced form coefficients based upon 3SLS". Ph.D. Thesis, London School of Economics and Political Science.

Maasoumi, E. (1978), "A modified Stein-like estimator for the reduced form parameters of simultaneous equations". *Econometrica* **46**, 695–703.

Maasoumi, E. (1985), "Mixing forecasts in linear simultaneous equations under quadratic loss". Mimeograph, Indiana University, Bloomington.

Maasoumi, E. (1986), "Reduced form estimation and prediction from uncertain structural models: a generic approach". *Journal of Econometrics*, **31**, 3–29.

McCarthy, M. (1972), "A note on the forecasting properties of 2SLS restricted reduced forms". *International Economic Review* **13**, 757–761.

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

Nagar, A. L., and R. Carter (1974), "Minimum second moment estimation in simultaneous equation systems". *International Economic Review* **15**, 31–38.

Nagar, A. L., and S. N. Sahay (1978), "The bias and mean squared error of forecasts from partially restricted reduced form". *Journal of Econometrics* **7**, 227–243.

Rhodes, G. F., and D. Westbrook (1980), "An experimental investigation of the properties of simultaneous equations estimators". (July), Colorado State University, Fort Collins, technical report (mimeo).

Rothenberg, T. (1984), "Approximating the distributions of econometric estimators and test statistics". In *Handbook of Econometrics*, Volume 2, ed. Z. Griliches and M. Intriligator, Ch. 15. Amsterdam: North-Holland.

Sargan, J. D. (1970), "The finite sample distribution of the FIML estimator". Presented at the World Congress of Econometric Society, Cambridge, England.

Sargan, J. D. (1976a), "Econometric estimators and Edgeworth approximations". *Econometrica* **44**, 421–448.

Sargan, J. D. (1976b), "The existence of the moments of estimated reduced form coefficients". Discussion Paper No. A6, LSE Econometrics Programme.

Sawa, T. (1973a), "Almost unbiased estimator in simultaneous equations systems". *International Economic Review* **14**, 97–106.

Sawa, T. (1973b), "The mean square error of a combined estimator and numerical comparison with the TSLS estimator". *Journal of Econometrics* **1**, 115–132.

Zellner, A., and W. Vandaele (1975), "Bayes-Stein estimators for k-means, regression and simultaneous equation models". In *Studies in Bayesian Econometrics and Statistics in Honor of Leonard J. Savage*, ed. S. E. Feinberg and A. Zellner, pp. 627–653. Amsterdam: North-Holland.