

Further consequences of viewing LIML as an iterated Aitken estimator

Chuanming Gao, Kajal Lahiri^{*}

Department of Economics, State University of New York at Albany,
Albany, NY 12222

ABSTRACT

Following an approach originally suggested by Pagan (1979), we have explored some additional consequences of viewing LIML as an iterated Aitken estimator. A curious result we derive in this context is that the Aitken estimator based on 2SLS residuals produces 2SLS estimate of the structural parameters. We derive a simple expression for the difference between 2SLS and LIML in finite samples, which turns out to be very similar to an expression for the approximate bias of 2SLS in small samples [Nagar (1959), Buse(1992) and Bound et al. (1995)]. The expression pinpoints the gain from iteration. Our analytical formula yields many of the results on the difference between 2SLS and LIML that Anderson and Sawa (1979) and Anderson et al. (1982) have obtained through extensive numerical simulations in a model with two endogenous variables. The robustness of 2SLS and LIML with respect to certain deviations from normality is also examined.

JEL classification: C30

Keywords: 2SLS, Limited Information, Iterated GLS, Finite Sample, Simultaneity, \hat{A}^2 and Student-t distribution.

^{*}Corresponding author. Tel: (518) 442 4758; fax: (518) 442 4736; e-mail: KL758@cnsvox.albany.edu. The authors wish to thank Terrence Kinal and two anonymous referees for helpful comments on earlier drafts of the article.

1. Introduction

Since the limited information (LI) specification of a standard simultaneous equation system can be written as a triangular model, Pagan (1979) suggested an alternative formulation of limited information maximum likelihood (LIML) estimation in terms of an iterated Seemingly Unrelated Regression (SUR) procedure.¹ The computation of LIML can be easily accomplished by the key idea of the paper. Using this approach, Pagan also derived an expression for the relationship between LIML and two stage least squares (2SLS) estimates in finite samples.

In this work we show that the first iterate of the feasible SUR for the structural equation produces 2SLS if 2SLS residuals (and OLS for the reduced form equations) are used to estimate the covariance matrix of the disturbances. This means that the analytical expression for the difference between LIML and 2SLS in Pagan (1979) is identically equal to zero for the first iteration of feasible iterated SUR. Given the structure of the LI specification, it also produces 3SLS estimates of all structural parameters. In addition, we show that the finite sample difference between LIML and 2SLS that Pagan (1979) obtained from this approach simplifies to a very familiar expression which is almost identical to the expression for the approximate expected bias of 2SLS obtained recently by Bound, Jaeger and Baker (1995) and Buse (1992). Three fundamental canonical form parameters well-known in the literature — the noncentrality parameter, a standardized structural coefficient, and the degree of overidentification — characterize this difference. Anderson and Sawa (1979) and Anderson, Kunitomo and Sawa (1982) have

¹See Lahiri and Schmidt (1978).

studied the difference between 2SLS and LIML by extensive numerical calculations in a model with two included endogenous variables. The advantage of the current approach is that our expression is simple, exact, and can readily be evaluated for any model specification. We obtain important additional insights on the conditions under which the two estimators will differ. Some Monte Carlo experiments confirm the superiority of LIML over 2SLS under normal and certain non-normal distributions.

2. SUR as LIML Estimator

Consider a structural equation

$$y = Y^{\circ} + X_1^{-} + u; \quad (1)$$

where Y is $T \times G_1$ matrix consisting of other included endogenous variables with reduced form

$$Y = X\gamma + v; \quad (2)$$

where X is $T \times K$ matrix of predetermined variables of the system. X_1 is $T \times K_1$ matrix with $K_2 = K - K_1$, G_1 and the structural equation is identified.

Assume that each row of $(u; v)$ has mean zero and covariance matrix $S = \begin{pmatrix} \gamma^2 & \gamma^0 \\ \gamma^0 & C \end{pmatrix}$. And $S^{-1} = \begin{pmatrix} a & b \\ b & C \end{pmatrix}$; which is partitioned such that a is a scalar, b is a vector, and C is a symmetric square matrix.

Let $Z = (Y; X_1)$, $\pm = \begin{pmatrix} \gamma^0 \\ \gamma^0 \end{pmatrix}$; $Y_R = \text{vec}(Y)$, $\gamma_R = \text{vec}(\gamma)$, $v_R = \text{vec}(v)$.

Applying generalized least squares to SUR model (1) and (2), following Pagan (1979), we have²

²To obtain (3), we use the result that $\text{vec}(ABC) = (C^0 - A) \text{vec}(B)$, see Magnus and Neudecker (1988).

$$\begin{bmatrix} \mathbf{b}_{\text{SUR}} \\ \mathbf{b}_{\text{R;SUR}} \end{bmatrix} = \begin{bmatrix} \frac{1}{2} \mathbf{Z}^0 \mathbf{Q} \mathbf{Z}^0 \mathbf{P}_X \mathbf{y} + \mathbf{a} \mathbf{Q} \mathbf{Z}^0 \mathbf{M}_X \mathbf{y} + \mathbf{Q} \mathbf{Z}^0 \mathbf{M}_X \mathbf{Y} \mathbf{b} \\ (\mathbf{I} - (\mathbf{X}^0 \mathbf{X})^{-1} \mathbf{X}^0) \mathbf{Y}_R + (\mathbf{C} - \mathbf{X}^0 \mathbf{X})^{-1} (\mathbf{b} - \mathbf{X}^0) (\mathbf{y} - \mathbf{Z} \mathbf{b}_{\text{SUR}}) \end{bmatrix}; \quad (3)$$

where

$$\mathbf{Q}^{-1} = \mathbf{a} \mathbf{Z}^0 \mathbf{Z}^{-1} (\mathbf{b}^0 - \mathbf{Z}^0 \mathbf{X}) (\mathbf{C} - \mathbf{X}^0 \mathbf{X})^{-1} (\mathbf{b} - \mathbf{X}^0 \mathbf{Z}) = \frac{1}{2} \mathbf{Z}^0 \mathbf{P}_X \mathbf{Z} + \mathbf{a} \mathbf{Z}^0 \mathbf{M}_X \mathbf{Z};$$

$$\mathbf{P}_A = \mathbf{A} (\mathbf{A}^0 \mathbf{A})^{-1} \mathbf{A}^0; \text{ and } \mathbf{M}_A = \mathbf{I} - \mathbf{P}_A \text{ for any matrix } \mathbf{A};$$

Factoring out $\mathbf{b}_{\text{2SLS}} = (\mathbf{Z}^0 \mathbf{P}_X \mathbf{Z})^{-1} \mathbf{Z}^0 \mathbf{P}_X \mathbf{y}$, (3) becomes

$$\begin{bmatrix} \mathbf{b}_{\text{SUR}} \\ \mathbf{b}_{\text{R;SUR}} \end{bmatrix} = \begin{bmatrix} \mathbf{b}_{\text{2SLS}} + (\mathbf{Z}^0 \mathbf{P}_X \mathbf{Z})^{-1} \mathbf{F} \mathbf{b}_{\text{2SLS}} + \mathbf{a} \mathbf{Q} \mathbf{Z}^0 \mathbf{M}_X \mathbf{y} + \mathbf{Q} \mathbf{Z}^0 \mathbf{M}_X \mathbf{Y} \mathbf{b} \\ (\mathbf{I} - (\mathbf{X}^0 \mathbf{X})^{-1} \mathbf{X}^0) \mathbf{Y}_R + (\mathbf{C} - \mathbf{X}^0 \mathbf{X})^{-1} (\mathbf{b} - \mathbf{X}^0) (\mathbf{y} - \mathbf{Z} \mathbf{b}_{\text{SUR}}) \end{bmatrix}; \quad (4)$$

where $\mathbf{F} = \mathbf{V}^0 (\mathbf{K} \mathbf{I} + \mathbf{V} (\mathbf{Z}^0 \mathbf{P}_X \mathbf{Z})^{-1} \mathbf{V}^0)^{-1} \mathbf{V}$; with $\mathbf{V} = \mathbf{M}_X \mathbf{Z}$; $\mathbf{K}^{-1} = \frac{1}{2} \mathbf{a}$;

3. The First Iterate of SUR Estimator

Up to now, we assumed that the covariance matrix \mathbf{S} is known. Curious results occur at the first iteration if 2SLS residuals are used to compute a consistent initial \mathbf{b} , which is usually suggested in practice. By partitioned inversion of $\mathbf{S}^{(0)}$, we have $\mathbf{b}^{(0)} = (\mathbf{v}^{(0)0} \mathbf{v}^{(0)})^{-1} \mathbf{v}^{(0)0} \mathbf{u}^{(0)} \mathbf{a}^{(0)}$; where the superscript (0) denotes the initial estimate. Typically, $\mathbf{u}^{(0)} = \mathbf{y} - \mathbf{Z} \mathbf{b}_{\text{2SLS}}^{(0)}$; $\mathbf{v}^{(0)} = \mathbf{Y} - \mathbf{X} \mathbf{b}_{\text{OLS}}^{(0)} = \mathbf{M}_X \mathbf{Y}$; where $\mathbf{b}_{\text{OLS}}^{(0)} = (\mathbf{X}^0 \mathbf{X})^{-1} \mathbf{X}^0 \mathbf{Y}$: Then

$$\mathbf{b}^{(0)} = (\mathbf{Y}^0 \mathbf{M}_X \mathbf{Y})^{-1} (\mathbf{Y}^0 \mathbf{M}_X \mathbf{y} - \mathbf{Y}^0 \mathbf{M}_X \mathbf{Z} \mathbf{b}_{\text{2SLS}}^{(0)}) \mathbf{a}^{(0)}; \quad (5)$$

Substituting (5) into (4), we first focus on $\mathbf{b}_{\text{SUR}}^{(1)}$: Notice that its last term is

$$\mathbf{Q}^{(0)} \mathbf{Z}^0 \mathbf{M}_X \mathbf{Y} \mathbf{b}^{(0)} = \mathbf{Q}^{(0)} \mathbf{Z}^0 \mathbf{M}_X \mathbf{Y} (\mathbf{Y}^0 \mathbf{M}_X \mathbf{Y})^{-1} (\mathbf{Y}^0 \mathbf{M}_X \mathbf{y} - \mathbf{Y}^0 \mathbf{M}_X \mathbf{Z} \mathbf{b}_{\text{2SLS}}^{(0)}) \mathbf{a}^{(0)}; \quad (6)$$

Using the fact that $Z^0 M_X Y (Y^0 M_X Y)^{-1} Y^0 M_X = Z^0 M_X$; (6) reduces to

$$Q^{(0)} Z^0 M_X Y b^{(0)} = \frac{1}{2} a^{(0)} Q^{(0)} Z^0 M_X Y + a^{(0)} Q^{(0)} Z^0 M_X Z \frac{1}{2} b^{(0)}$$

Note that the first term from the above expression cancels out the third term of $\pm_{SUR}^{(1)}$ in (4), and the second term cancels out the second term of $\pm_{SUR}^{(1)}$ in (4) since

$$\begin{aligned} & a^{(0)} Q^{(0)} Z^0 M_X Z \frac{1}{2} b^{(0)} - \frac{1}{2} (Z^0 P_X Z)^{-1} F^{(0)} b^{(0)} \\ &= a^{(0)} Q^{(0)} Z^0 M_X Z \frac{1}{2} b^{(0)} + \frac{1}{2} a^{(0)} Q^{(0)} Z^0 P_X Y \frac{1}{2} b^{(0)} \\ &= [Q^{(0)} Q^{-1} - \frac{1}{2} I] \frac{1}{2} b^{(0)} \end{aligned}$$

which is a zero vector. Therefore $\pm_{SUR}^{(1)} = \frac{1}{2} b^{(0)}$; a simple but curious result which seems to be unrecognized in the literature. Also, since the first structural equation can be the only overidentified equation in the LI specification (1)-(2), $\frac{1}{2} b^{(0)} = \frac{1}{2} b^{(0)}$, see Narayanan (1969).³ One implication of this result is that iterative 3SLS always yields 2SLS estimates for the single overidentified equation.

Next consider $\frac{1}{4} b_{R;SUR}^{(1)}$: Following Court (1973) and recognizing that $\pm_{SUR}^{(1)} = \frac{1}{2} b^{(0)}$ in (4); we find that $\frac{1}{4} b_{R;SUR}^{(1)} = \frac{1}{4} b_{R;3SLS}$, which is more efficient than $\frac{1}{4} b_{R;OLS}$. Interestingly, it can be shown that, for a general triangular system [Lahiri and Schmidt (1978)], the first iterate of SUR based on 2SLS residuals does not yield 3SLS for the system, neither does it yield 2SLS for the structural equation.

³It is interesting to note that this result holds independent of the estimated β ; even for general simultaneous equations system with a single overidentified equation. Unfortunately, this fact has not been explicitly recognized and is lost in the proof by Narayanan (1969), and remains unappreciated later on. See Court (1973), Hausman (1983, p. 423), among others.

4. The Gain from Iteration

Although there is no gain in estimating \pm at the first iteration, iterated SUR numerically produces LIML parameter estimates, on convergence. In the following, we examine the gain from iteration. We first derive a simple expression for the difference between LIML and 2SLS.

Based on (4), consider the n -th iteration. We can write

$$\begin{pmatrix} \pm_{SUR}^{(n)} \\ \frac{1}{4}_{R;SUR}^{(n)} \end{pmatrix} = \begin{pmatrix} \mathbf{b}_{2SLS} + \pm_4^{(n)} \\ \mathbf{b}_{R;OLS} + \frac{1}{4}_{R;4}^{(n)} \end{pmatrix}; \quad (7)$$

where

$$\begin{pmatrix} \pm_4^{(n)} \\ \frac{1}{4}_{R;4}^{(n)} \end{pmatrix} = \begin{pmatrix} \mathbf{I} - \mathbf{Z}^0 \mathbf{P}_X \mathbf{Z} \end{pmatrix}^{-1} \mathbf{F}^{(n_i-1)} \mathbf{b}_{2SLS} + \mathbf{a}^{(n_i-1)} \mathbf{Q}^{(n_i-1)} \mathbf{Z}^0 \mathbf{M}_X \mathbf{y} + \mathbf{Q}^{(n_i-1)} \mathbf{Z}^0 \mathbf{M}_X \mathbf{Y} \mathbf{b}^{(n_i-1)} \\ \left(\mathbf{C}^{(n_i-1)} - \mathbf{X}^0 \mathbf{X} \right)^{-1} \left(\mathbf{b}^{(n_i-1)} - \mathbf{X}^0 \mathbf{u}^{(n)} \right); \quad (8)$$

$$\frac{1}{4}_{R;4}^{(n)} = \text{vec}\left(\frac{1}{4}_4^{(n)}\right); \text{ and } \mathbf{u}^{(n)} = \mathbf{y} - \mathbf{Z} \pm_{SUR}^{(n)}; \text{ for } n > 2:$$

Notice that (7) is formulated as the sum of the initial estimates and the cumulative improvements up to the n -th iteration. Pagan (1979) derived this expression for the wedge between LIML and 2SLS, but did not explore its implication any further. What we have demonstrated in Section 3 is that $\pm_4^{(1)} = 0$: A close examination shows that the $O_p(T^{i-1=2})$ terms in $\pm_4^{(n)}$ cancel each other out and we are left only with higher order corrections.

Consider a general overidentified case.⁴ Note that after the $(n_i - 1)$ -th iteration, the LI model (1)-(2) may be decomposed as $\mathbf{y} = \mathbf{Z} \mathbf{b}_{2SLS} +$

⁴In the special case where the structural equation of the LI system is just identified, $\mathbf{K}_2 = \mathbf{G}_1$. Then the first iteration based on 2SLS residuals yields $\pm_{SUR}^{(1)} = \mathbf{b}_{2SLS} = (\mathbf{X}^0 \mathbf{Z})^{-1} \mathbf{X}^0 \mathbf{y}$; and

$$\frac{1}{4}_{R;SUR}^{(1)} = \mathbf{b}_{R;OLS} + \mathbf{C}^{i-1;(0)} \mathbf{b}^{(0)} - (\mathbf{X}^0 \mathbf{X})^{-1} \mathbf{X}^0 \mathbf{u}^{(1)}$$

$Z_{\pm 4}^{(n_i-1)} + u^{(n_i-1)}$, and $Y = Xb_{OLS} + X\frac{1}{4}^{(n_i-1)} + v^{(n_i-1)}$ for $n > 2$: We then write $u^{(n_i-1)} = u^{(0)} + Z_{\pm 4}^{(n_i-1)}$; and $v^{(n_i-1)} = v^{(0)} + X\frac{1}{4}^{(n_i-1)}$: Substituting these into $b^{(n_i-1)} = (v^{(n_i-1)0}v^{(n_i-1)})^{-1}v^{(n_i-1)0}u^{(n_i-1)}a^{(n_i-1)}$, using the fact that $v^{(0)0}X = 0$; and manipulating (8), we get

$$\begin{aligned} \pm_4^{(n)} &= a^{(n_i-1)}Q^{(n_i-1)}\alpha[\frac{1}{4}^{(n_i-1)0}X^0u^{(n_i-1)} + v^{(0)0}Z_{\pm 4}^{(n_i-1)} \\ &\quad + v^{(0)0}v^{(0)}H^{(n_i-1)}v^{(n_i-1)0}u^{(n_i-1)}]; \end{aligned} \quad (9)$$

where

$\alpha = (Z^0M_X Y)(Y^0M_X Y)^{-1} = \begin{pmatrix} I \\ 0 \end{pmatrix}$; which is partitioned conformably with Z : Note $H^{(n_i-1)} = (v^{(0)0}v^{(0)})^{-1} + (v^{(n_i-1)0}v^{(n_i-1)})^{-1}$:

If we focus on the parameters of the endogenous variables in the structural equation, by partitioned inversion of $a^{(n_i-1)}Q^{(n_i-1)} = (k^{(n_i-1)}Z^0P_X Z + Z^0M_X Z)^{-1}$ and using the structure of α ; we get

$$\begin{aligned} \circ_4^{(n)} &= [Y^0M_X Y + k^{(n_i-1)}Y^0(P_X + P_{X_1})Y]^{-1}[\frac{1}{4}^{(n_i-1)0}X^0u^{(n_i-1)} + \\ &\quad Y^0M_X Y \circ_4^{(n_i-1)} + v^{(0)0}v^{(0)}H^{(n_i-1)}v^{(n_i-1)0}u^{(n_i-1)}]; \end{aligned} \quad (10)$$

where $P_{X_1} = X_1(X_1^0X_1)^{-1}X_1^0$: Note $C^{i-1:(n_i-1)}b^{(n_i-1)} = (v^{(n_i-1)0}u^{(n_i-1)}(u^{(n_i-1)0}u^{(n_i-1)})^{-1})^{-1}$; and $v^{(n_i-1)0}v^{(n_i-1)} = v^{(0)0}v^{(0)} + \frac{1}{4}^{(n_i-1)0}X^0X\frac{1}{4}^{(n_i-1)}$. Some algebraic manipulation of (10) yields

$$\begin{aligned} \circ_4^{(n)} &= [Y^0M_X Y + k^{(n_i-1)}Y^0(P_X + P_{X_1})Y]^{-1}[Y^0M_X Y \circ_4^{(n_i-1)} \\ &\quad + v^{(n_i-1)0}u^{(n_i-1)}(u^{(n_i-1)0}u^{(n_i-1)})^{-1}u^{(n_i-1)0}P_X u^{(n_i-1)}k^{(n_i-1)}]; \quad (11) \\ &= b_{R;OLS} + C^{i-1:(0)}b^{(0)} - [(X^0X)^{-1}X^0(y - Z(X^0Z)^{-1}X^0y)] \end{aligned}$$

Note that the term in the brackets is a zero vector, so that $\frac{1}{4}_{R;SUR}^{(1)} = b_{R;OLS}$: Thus the iterative process converges on the ...rst iteration. Therefore we get the well-known result that for a just identified structural equation, 2SLS is identical to LIML, cf. Greene (1997, p. 745). By implication, the gain from performing iterated SUR for a general LI model hinges on the overidentifying restrictions of the structural model.

Upon convergence, $\hat{\beta}_4^{(n)} = \hat{\beta}_4^*$: Evaluated at the convergent values, we get

$$\hat{\beta}_4^* = \left[Y'(P_X - P_{X_1})Y \right]^{-1} v' u^* (u^{*'} u^*)^{-1} u^{*'} P_X u^*; \quad (12)$$

which is the correction of iterated SUR over 2SLS. It provides a simple analytical expression for the difference between LIML and 2SLS (or the n -th iterate of LIML) estimates of β in finite samples.

So far, we have shown that the 2SLS is the first step Aitken estimator and it differs from the iterated Aitken estimator by expression (12) for parameter vector β . It would be interesting to examine the second step Aitken estimator as well. Since 2SLS and LIML are both efficient in large samples, the second step Aitken estimator is also asymptotically as efficient as 2SLS and LIML. In finite samples, the gain of the second step Aitken estimator $\hat{\beta}_{SUR}^{(2)}$ for β over 2SLS, using (11), is given by

$$\hat{\beta}_4^{(2)} = \left[k^{-1} Y' M_X Y + Y'(P_X - P_{X_1})Y \right]^{-1} v^{(1)'} u^{(0)} (u^{(0)'} u^{(0)})^{-1} u^{(0)'} P_X u^{(0)}; \quad (13)$$

since $\hat{\beta}_4^{(1)} = 0$, $u^{(1)} = u^{(0)}$.

In cases where $\hat{\beta}_4^{(2)}$ is very close to $\hat{\beta}_4^*$, it may be computationally acceptable to stop at the second step to avoid further iteration. To get some insight on the relationship, let us compare the components in (13) with those in (12) for a model with two included endogenous variables, i.e. $G_1 = 1$: First note that in this case, $k = 1 - \frac{1}{2}$; where $\frac{1}{2}$ is the correlation coefficient between u and v : Specifically, $k^{(1)} = 1 - (u^{(0)'} u^{(0)})^{-1} u^{(0)'} v^{(1)} (v^{(1)'} v^{(1)})^{-1} v^{(1)'} u^{(0)}$, so in general $0 < k^{(1)} < 1$: Also, $Y' M_X Y$ and $Y'(P_X - P_{X_1})Y$ are positive semi-definite symmetric. Second, since $(u^{*'} u^*)^{-1} u^{*'} P_X u^*$ is a non-negative scalar, $\hat{\beta}_4^*$ is of an opposite sign to that of the covariance term $v' u^*$: Using this fact,

it is easy to show that $v^{(1)0}u^{(0)}$ is always greater than $v^{(1)0}u^{(0)}$ in absolute value. Further, $(u^{(0)0}u^{(0)})i^{-1}u^{(0)0}P_X u^{(0)}$ will be always greater than but very close to $(u^{(0)0}u^{(0)})i^{-1}u^{(0)0}P_X u^{(0)}$. Combined together, we find that $0 < \frac{\sigma_4^{(2)}}{\sigma_4^{(1)}} < 1$ when $\frac{1}{2} \in 0$ and $K_2 > G_1$: In cases where $\frac{1}{2} = 0$ or $K_2 = G_1$, both $\frac{\sigma_4^{(2)}}{\sigma_4^{(1)}}$ and $\sigma_4^{(1)}$ will be close or equal to zero, see also footnote 4.

Note that $Y^0(P_X | P_{X_1})Y$ in (13) represents the variation in Y explained by the instruments excluded from the structural equation, and $Y^0M_X Y = v^{(0)0}v^{(0)}$. If these additional instruments are weak, or if $\frac{1}{2}$ is large, we will expect $\frac{\sigma_4^{(2)}}{\sigma_4^{(1)}}$ in absolute value to be significantly less than $\sigma_4^{(1)}$. Thus, in these cases, there is a need to iterate on the second step Aitken estimator. This may also explain why some researchers found the convergence of iterated SUR to be slow.⁵

5. Relative Performance of LIML vs 2SLS

In addition to (11) and (13), (12) allows us to reexamine some previous results regarding the relative performance of LIML and 2SLS without evaluating the finite sample distributions or approximating the asymptotic distributions of the estimators.

While expressions (11) to (13) permit analysis for a general model, in this section we consider the special case when there are two included endogenous variables. For this case, (12) may then be rewritten as

$$\sigma_4^{(1)} = i \frac{\frac{1}{2}_{\alpha} u^{(0)0} P_X u^{(0)}}{\frac{1}{2}_{\alpha} \frac{1}{4}_{\alpha}!_{\alpha}}; \quad (14)$$

where subscript (α) or superscript (*) denotes values evaluated or estimated at convergence. Typically, $\frac{1}{2}_{\alpha}^2 = u^{(0)0}u^{(0)}=T$; $\frac{1}{4}_{\alpha}^2 = v^{(0)0}v^{(0)}=T$. Note that $\frac{1}{2}_{\alpha}^2 =$

⁵See Hausman (1983, p. 426) and Greene (1997, p. 748).

$Y'(P_X - P_{X_1})Y/n$ is the sample "concentration parameter" or "noncentrality parameter". Alternatively, we find it more convenient to rewrite (14) as

$$\hat{\beta}_4 = \beta + \frac{1/2}{n} \frac{u' P_X u}{3/4 n}; \quad (15)$$

where $\hat{\lambda}^2 = Y'(P_X - P_{X_1})Y/n$; which we call the "modified" sample concentration parameter.

Note that (14) is an exact expression for the difference between LIML and 2SLS in finite samples without any distributional assumption. There have been a number of important studies on the difference between LIML and 2SLS in the case of two included endogenous variables, see Mariano (1982), and Phillips (1983) for excellent surveys.⁶ Various canonical forms are derived in the literature to indicate directly the critical parameter functions which affect the statistical behavior of the two estimators. However, an expression for the exact difference between LIML and 2SLS has not readily emerged from the complicated expressions of the exact or asymptotically approximate distributions of the two estimators.

Remark 1. Assuming normality and using power series approximation methods, Nagar (1959) obtained the expected bias of 2SLS to the order of $1/n$: Recently, Buse (1992) derived the expected bias of instrumental variable estimators (to the order of $1/n$) without assuming normality. In current notations, a little rearrangement of their expressions gives the approximate bias of 2SLS estimator for β as

$$\frac{1/2}{n} \frac{3/4}{n} (K_2 - 2) \quad (16)$$

⁶See also Amemiya (1985, Section 7.3).

or

$$\frac{1}{2}(K_2 - 1); \quad (17)$$

where $\frac{1}{2}$; $\frac{3}{4}$; 1 are unknown population values. As is usually defined in the literature, $\sigma^2 = (EY)'(P_X - P_{X_1})(EY) = \frac{1}{2}$; Correspondingly, $\sigma^2 = (EY)'(P_X - P_{X_1})(EY) = \frac{3}{4}$; More recently, under the assumption of normality, drawing on Sawa's (1969) work on the exact sampling distribution of 2SLS, Bound, Jaeger and Baker (1995) found a result similar to (16) with $(K_2 - 1)$ replaced by K_2 , for $K_2 > 1$; When $K_2 = 1$; $E(\mathbf{b}_{2SLS})$ does not exist.

Notice the interesting similarity between (14) and (16), or between (15) and (17). Actually, $u'P_X u = \frac{3}{4} \sigma^2$ is exclusively a function of K_2 ; the number of excluded predetermined variables in the structural equation; it has an approximate median (its mean may not exist)⁷ close to $(K_2 - 1)$.⁸ Some Monte Carlo simulation confirmed that $u'P_X u = \frac{3}{4} \sigma^2$; on "average", is between $K_2 - 1$ and K_2 ; Therefore (16) almost offsets (14), supporting the conclusion that LIML is median unbiased — a result which has emerged from extensive tabulation of the distributions of 2SLS and LIML by Anderson and Sawa (1979) and Anderson, Kunitomo and Sawa (1982).

Remark 2. Based on (15), we may make the following additional observations:

⁷Since LIML estimator has no finite moments, it is safe to infer that σ_p^2 has no finite moments [see Mariano and Sawa (1972)].

⁸Note that

$$\begin{aligned} E(u^{(0)'} P_X u^{(0)}) &= E(u' (P_X - P_X Z (Z' P_X Z)^{-1} Z' P_X) u) \\ &= \text{tr}(P_X - P_X Z (Z' P_X Z)^{-1} Z' P_X) E(uu') \\ &= (K_2 - 1) G_1 \frac{3}{4} \sigma^2; \end{aligned}$$

where $\frac{3}{4} \sigma_{(0)}^2 = u^{(0)'} u^{(0)} = T$ and currently $G_1 = 1$: See Buse (1992, p. 175) for a similar expression.

(i) For large samples, the difference between LIML and 2SLS is $O_p(1/T)$: It can be shown that $\text{plim } \hat{\beta}_4^{\text{2SLS}} = 0$, $\text{plim } \hat{\beta}_4^{\text{LIML}} = 0$; and $T \hat{\beta}_4^{\text{2SLS}} = O_p(1)$: The latter holds under the condition that $\text{plim } \hat{\beta}_4^{\text{LIML}} = \beta_4$ exists and is finite, which is usually assumed in the literature. However, it is difficult to derive the asymptotic distribution of $T \hat{\beta}_4^{\text{2SLS}}$ because the elements in $\hat{\beta}_4^{\text{2SLS}}$ are not mutually independent. These properties also hold for $\hat{\beta}_4^{(2)}$ in (13).

(ii) If LIML is median unbiased, then 2SLS is median biased unless, of course, the structural equation is just identified (see footnote 4) or $\beta_4 = 0$: Obviously, the second step Aitken estimator will be less (median) biased than 2SLS.

(iii) $\hat{\beta}_4^{\text{2SLS}}$ is of the opposite sign to β_4 : This result was found in Anderson and Sawa (1979). They tabulated the asymmetry and skewness in the 2SLS distribution through Monte Carlo experiments and found that the probability is close to 1 that the 2SLS estimators will be on one side of the true value for some combinations of parameter values (such as $K_2 \rightarrow 20$; low R^2 and high $j|j$):

(iv) The absolute value of $\hat{\beta}_4^{\text{2SLS}}$ is directly proportional to $j|j$; inversely proportional to R^2 ; and an increasing function of $u^{\text{plim}} P_X u^{\text{plim}} = \frac{1}{n} \sum u^2$. Note that in general, (15) does not imply that $j|j$ will necessarily rise as K_2 increases. As K_2 increases, the concentration parameter may increase more than proportionately if the additional instruments are important, and thus $j|j$ may actually fall. In other words, if the additional instruments are weak such that the first stage R^2 rises less than proportional to the increase in K_2 ; then $j|j$ may increase; see Buse (1992).

(v) $\hat{\beta}_4^{\text{2SLS}}$ does not explicitly depend on the degrees of freedom ($T - K$) or

sample size T , which appears in the estimator of the covariance matrix of the reduced form or the concentration parameter. Anderson et al. (1982) noticed that the distribution of LIML estimator does not depend much on $(T \downarrow K)$. The observed effect of the degrees of freedom on the distribution of LIML estimator in their tabulation, which is significant when the concentration parameter is small, can be attributed to the implied weakening of the instruments. In a Monte Carlo setup, in order to attain a given value for the concentration parameter (λ^2 or λ^{-2}), the reduced form coefficients, ceteris paribus, have to be reduced as sample size increases. As the instruments get weaker, the first stage R^2 gets smaller. Then σ_{β}^2 , which is the sample counterpart of λ^{-2} and actually appears in (15); will have more variability, and will thereby affect the distribution of LIML estimator.

Remark 3. As an illustration, we examined the Angrist and Krueger (1991) data on earning and schooling, which has recently motivated considerable research on models with weak instruments, see Blomquist and Dahlberg (1999). Using the data for Angrist, Imbens and Krueger (1999) obtainable from the Journal of Applied Econometrics Data Archive, we replicated their 2SLS and LIML estimates as the first step and iterated Aitken estimators respectively. Table 1 reports the values of the components of σ_{β}^2 . Note that σ_{β}^2 computed from expression (15) coincides exactly with the difference between LIML and 2SLS estimates computed directly. Interestingly, the correlation coefficient ρ is estimated to be ≈ 0.11 for the version of their model with 30 instruments, and ≈ 0.19 for their model with 180 instruments. Therefore it is not surprising to observe that the 2SLS and LIML estimates are very close to each other although the degree of overidentification is very high.

6. Robustness to Non-normality

Since expression (12) does not depend on any assumption on the distribution of errors in (1) and (2), it would be instructive to investigate the effect of non-normal disturbances on (12) through a Monte Carlo study. To compare with normal disturbances, we considered two alternatives: a standardized \hat{A}^2 distribution with a degree of freedom of 1 and a scaled Student-t distribution with degrees of freedom of 6.⁹ Both non-normal distributions have mean zero and variance one, but the standardized $\hat{A}^2(1)$ distribution is skewed, while the scaled Student-t distribution is leptokurtic. For the Monte Carlo experiments presented in this section, we followed the specification in Anderson et al. (1982) and focused on setup corresponding to their Figures 2 and 3. In their Figure 2, $T_j = K = 10$; $K_2 = 10$; $\beta = 1.0$; $\sigma^2 = 100$ (specification A, hereafter). In their Figure 3, $T_j = K = 30$; $K_2 = 30$; $\beta = 1.0$; $\sigma^2 = 100$ (specification B, hereafter): Here β is a standardized structural coefficient as defined in Anderson et al. (1982). Figures 1 to 6 plot distributions for estimates of parameter β computed from first step Aitken (2SLS), second step Aitken, third step Aitken (LIML), and the difference β_4 . All estimates have been normalized as in Anderson et al. (1982) such that their asymptotic distributions are standard normal under normal disturbances. The standard normal cumulated distribution function (cdf) is also plotted. The number of replications is 20,000 in each case.

Figures 1 and 4 use normal disturbances and duplicate figures 2 and 3 in Anderson et al. (1982) fairly well for the cdfs of LIML and 2SLS. They show

⁹A standardized $\hat{A}^2(1)$ distribution is obtained as $x = (\sum_{i=1}^p \epsilon_i^2)^{-1/2} \epsilon_1$, where $\epsilon \sim \hat{A}^2(1)$, cf. Blundell and Bond (1998, Table 7). A scaled Student-t distribution with degrees of freedom n is obtained as $x = (\frac{n-2}{n})^{1/2} t$, where $t \sim t_n$: We choose $n = 6$, which is typical for many macroeconomic series, see Lahiri and Teigland (1987) and Geweke (1993).

that the cdf of LIML is closer to normal than 2SLS and median unbiased, as noted in Anderson et al. (1982). With the same value for the concentration parameter, as K_2 increases from 10 to 30, 2SLS moves further away from the true value and its median shifts from $\hat{\beta}_1:0.582$ to $\hat{\beta}_1:1.606$: The normalized (median) difference between LIML and 2SLS estimates triples from 0.518 to 1.519, which matches exactly with the prediction based on (14). Note that a prediction of the bias of 2SLS based on (16) gives a value of $\hat{\beta}_1:0.566$ for specification A, and a value of $\hat{\beta}_1:1.980$ for specification B, which are greater than the estimated values and the prediction based on (14). The superiority of (14) is that it uses the sample concentration parameter. Clearly, the second step Aitken estimator has loci between 2SLS and LIML, with median bias less than that of 2SLS. In specification A, the median bias of the second step Aitken estimator is about half of that of 2SLS. In specification B, it is about 65%.

Figures 2 and 5 depict cdfs when standardized $\hat{A}^2(1)$ disturbances are used in the data generating process (DGP). Compared to Figures 1 and 4, it is noted that this type of non-normality has little impact. LIML remains median unbiased and its cdf is close to normal. The median bias of 2SLS decreases from $\hat{\beta}_1:0.582$ to $\hat{\beta}_1:0.509$ in specification A, and from $\hat{\beta}_1:1.606$ to $\hat{\beta}_1:1.516$ in specification B. As a result, the median of the difference $\hat{\beta}_1:\hat{\beta}_2$ decreases by 20% in specification A, and 8% in specification B. The second step Aitken estimator is also affected similarly.

Figures 3 and 6 show the significant impact when a scaled Student-t distribution is used in generating the disturbances. LIML is still median unbiased but its cdf deviates from the standard normal distribution dramatically in

both specifications. The computed standard deviation of LIML estimates across replications increases from 1.11 in Figure 1 to 1.58 in Figure 4, and from 1.20 in Figure 4 to 1.69 in Figure 6. Compared to cases in Figures 1 and 4 with normal disturbances, the median bias of 2SLS increases by 44% in specification A, and by 34% in specification B, with their values exceeding the prediction based on (16). Meanwhile, the median of the difference σ_{β}^2 increases by 35% in specification A, and by 32% in specification B. A close look at the components of σ_{β}^2 in (15) reveals that, on “average”, $u^{*0}P_X u^* = \frac{3}{4}\sigma_u^2$ remains close to $(K_2 \text{ } G_1)$, but the modified sample concentration parameter $\hat{\sigma}_{\beta}^2$ becomes smaller due to excess kurtosis; in $\hat{\sigma}_{\beta}^2 = Y^0(P_X \text{ } P_{X_1})Y = \frac{1}{n}\frac{3}{4}\sigma_u^2$, both the numerator and the denominator increase, but the estimated standard errors $\frac{3}{4}\sigma_u$ and $\frac{1}{n}$ increase proportionately more than the numerator. One implication of this result is that in the presence of excess kurtosis in the disturbances, the observed difference between LIML and 2SLS will be more than what we expect under normality through its impact on $\hat{\sigma}_{\beta}^2$.

7. Concluding Remarks

Following an approach originally suggested by Pagan (1979), we have explored some additional consequences of viewing LIML as an iterated Aitken estimator. A curious result we derive in this context is that the Aitken estimator (without correcting for endogeneity) based on 2SLS residuals produces 2SLS estimate of the structural equation parameters. In the limited information specification, since the structural equation can be the only overidentified equation of the system, the feasible Aitken estimator is also 3SLS of the entire model. We also derive a simple formula for the difference between 2SLS and LIML in finite samples which turns out to be very similar to an expres-

sion for the approximate bias of 2SLS in small samples, cf. Nagar(1959), Buse(1992) and Bound et al. (1995). Since the first iterate of the iterated SUR is 2SLS, this expression also pinpoints the gain from iteration. Our analytical formula yields many of the results on the difference between 2SLS and LIML that Anderson and Sawa (1979) and Anderson et al. (1982) have obtained through extensive numerical simulations in a model with two endogenous variables. It reveals that the difference between 2SLS and LIML is $O_p(1/T)$. Some Monte Carlo experiments show that the formula is robust to skewness in the disturbances, but excess kurtosis will reduce the modified sample concentration parameter and result in greater bias in 2SLS.

The key conclusion worth re-emphasizing is that under many reasonable conditions 2SLS can be badly biased, particularly if the degree of overidentification is large. Given that LIML is known to be median unbiased, the expression for the difference between 2SLS and LIML derived in the paper can help to identify empirical situations where the finite sample bias in 2SLS is expected to be substantial.

References

- Amemiya, T., 1985. *Advanced Econometrics* (Basil Blackwell, Oxford).
- Anderson, T.W. and T. Sawa, 1979. Evaluation of the distribution function of the two-stage least squares estimate. *Econometrica* 47, 163-182.
- Anderson, T.W., N. Kunitomo and T. Sawa, 1982. Evaluation of the distribution function of the limited information maximum likelihood estimator. *Econometrica* 50, 1009-1027.
- Angrist, J.D. and A. Krueger, 1991. Does compulsory school attendance affect schooling and earnings. *Quarterly Journal of Economics* 106, 979-1014.
- Angrist, J.D., G.W. Imbens and A. Krueger, 1999. Jackknife instrumental variables estimation. *Journal of Applied Econometrics* 14, 57-67.
- Blomquist, S. and M. Dahlberg, 1999. Small sample properties of LIML and jackknife IV estimators: experiments with weak instruments. *Journal of Applied Econometrics* 14, 69-88.
- Bound, J., D.A. Jaeger, and R.M. Baker, 1995. Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak. *Journal of the American Statistical Association* 90, 443-450.
- Blundell, R. and S. Bond, 1998. Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics* 87, 115-143.

- Buse, A., 1992. The bias of instrumental variable estimators. *Econometrica* 60, 173-180.
- Court, R.H., 1973. Efficient estimation of the reduced form from incomplete models. *Review of Economic Studies* 40, 411-418.
- Geweke, J., 1993. Bayesian treatment of the independent student-t linear model. *Journal of Applied Econometrics* 8, S19-S40.
- Greene, W., 1997. *Econometric Analysis*, 3rd edition (Prentice-Hall, New Jersey, NJ).
- Hausman, J., 1983. Specification and estimation of simultaneous equation models. In: Z. Griliches, and M. Intriligator, eds., *Handbook of Econometrics* (North Holland, Amsterdam).
- Lahiri, K. and P. Schmidt, 1978. On the estimation of triangular structural systems. *Econometrica* 46, 1217-1222.
- Lahiri, K. and C. Teigland, 1987. On the normality of probability distributions of inflation and GNP forecasts. *International Journal of Forecasting* 3, 269-279.
- Magnus, J.R., and H. Neudecker, 1988. *Matrix Differential Calculus with Applications in Statistics and Econometrics* (John Wiley & Sons: Chichester).
- Mariano, R.S., 1982. Analytical small sample distribution theory in econometrics: The simultaneous-equations case. *International Economic Review* 23, 503-533.

- Mariano, R.S. and T. Sawa, 1972. The exact finite sample distribution of the limited information maximum likelihood estimator in the case of two included endogenous variables. *Journal of the American Statistical Association* 67, 159-163.
- 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.
- Narayanan, R., 1969. Computation of Zellner-Theil's three stage least squares estimates. *Econometrica* 37, 298-306.
- Pagan, A., 1979. Some consequences of viewing LIML as an iterated Aitken estimator. *Economics Letters* 3, 369-372.
- Phillips, P.C.B., 1983. Exact small sample theory in the simultaneous equations model. In: Z. Griliches, and M. Intriligator, eds., *Handbook of Econometrics* (North Holland, Amsterdam).
- Sawa, T., 1969. The exact sampling distribution of ordinary least squares and two-stage least squares estimators. *Journal of the American Statistical Association* 64, 923-937.