Econometrica, Vol. 58, No. 5 (September, 1990), 1181–1194

ON THE NORMALIZATION OF STRUCTURAL EQUATIONS: PROPERTIES OF DIRECTION ESTIMATORS[1]

BY GRANT H. HILLIER

In the general structural equation model only the direction of the vector of coefficients of the endogenous variables is determined. The traditional normalization rule defines the coefficients that are of interest but should not be embodied in the estimation procedure: we show that the properties of the traditionally defined ordinary least squares and two stage least squares estimators are distorted by their dependence on the normalization rule. Symmetrically normalized analogues of these estimators are defined and are shown to have essentially similar properties to those of the limited information maximum likelihood estimator.

KEYWORDS: Structural equation, normalization, direction estimators, exact distributions.

1. INTRODUCTION

EXACT DISTRIBUTION THEORY for the classical structural equation model in econometrics is notoriously complex, and results for even the simplest cases have seemed too complicated to yield interesting analytical conclusions about the relative merits of different estimators. Thus, Anderson and Sawa (1973) and Anderson, Kunimoto, and Sawa (1982), for instance, have resorted to extensive numerical tabulations of the exact densities to extract such information. For the case of an equation with just two endogenous variables these tabulations suggest that, in several respects, the limited information maximum likelihood (LIML) estimator is superior to the ordinary least squares (OLS) and two-stage least squares (TSLS) estimators, among others, and these conclusions are supported by the higher-order asymptotic results in Anderson, Kunimoto, and Morimune (1986) and other work referenced there.

The basis for inference in this model is the joint distribution of the included endogenous variables (represented by the reduced form), together with the maintained hypothesis that a submatrix of the reduced form coefficient matrix has rank one less than its column dimension. This condition determines the direction of a vector in R^{n+1}—where $n+1$ is the total number of endogenous variables in the equation—but not its length. Thus, some normalization rule is needed to determine the coefficient vector uniquely, and the distribution theory referred to above has focused on results for the n coefficients remaining when the other is assumed to be unity.

In this paper we focus on the estimation of the direction of the $(n+1)$-dimensional coefficient vector in an unnormalized equation. Existing distribution results for the coefficients in a normalized equation are easily translated

[1] An earlier version of this paper bore the title "On the interpretation of exact results for structural equation estimators." My thanks to the referees and the Co-Editor for comments on earlier versions of the paper that helped to both clarify the message and broaden the scope of the paper.

into results for the corresponding direction estimators, and, as we shall see, this at once makes these complex formulae more intelligible. In fact, at least in the case $n = 1$, the results clearly demonstrate the superiority of the LIML estimator and hence explain analytically the results of the numerical studies referred to earlier. In the case $n > 1$ the results are less decisive, but they do, nevertheless, provide considerably greater insight into the properties of these estimators than their counterparts for the normalized case.

The results we obtain suggest that the normalization rule embodied in the OLS and TSLS estimators distorts their properties. This prompts the question of whether or not analogues of OLS and TSLS based on the same normalization rule as the LIML estimator would correct this distortion, and we show below that this is indeed the case.[2] Thus, our results show that the relatively poor performance of the OLS/TSLS estimators evident in the numerical studies referred to above can be attributed to their dependence on a particular normalization rule. This is not to deny that the coefficients of the right-hand-side endogenous variables may be of primary interest, but does imply that the estimator for those coefficients should be based on a procedure that initially ignores the normalization rule suggested by that interest.

2. MODEL AND ASSUMPTIONS

We consider a single structural equation, written without an explicit normalization rule,

$$(1) \qquad (y,Y)\beta_\Delta = Z_1\gamma + u,$$

with corresponding reduced form

$$(2) \qquad (y,Y) = (Z_1,Z_2)\begin{bmatrix} \pi_1 & \Pi_1 \\ \pi_2 & \Pi_2 \end{bmatrix} + (v,V),$$

where y is $T\times 1$, Y is $T\times n$, Z_1 is $T\times K_1$, Z_2 is $T\times K_2$, $K_2 \geqslant n$, and $Z=(Z_1,Z_2)$ is fixed and of full column rank $K=K_1+K_2$. To simplify the notation, but without loss of generality, we shall assume that the model is already in canonical form—see Phillips (1983) for details of the reduction to canonical form. Thus, we assume that the rows of (v,V) are independent normal vectors with mean zero and covariance matrix I_{n+1}, and also that $Z'Z=I_K$.

As is well known, (1) is compatible with (2) if and only if the relations $\gamma=(\pi_1,\Pi_1)\beta_\Delta$,

$$(3) \qquad (\pi_2,\Pi_2)\beta_\Delta = 0,$$

and $u=(v,V)\beta_\Delta$, hold for some $\beta_\Delta \neq 0$. Equation (3) implies $\operatorname{rank}(\pi_2,\Pi_2) \leqslant n$, and we shall assume that the identification condition $\operatorname{rank}(\pi_2,\Pi_2)=n$ also holds. Under this condition equation (3) uniquely determines the "direction" of β_Δ, but not its length.

[2] This question was suggested by a referee, who also pointed out that the analog of the TSLS estimator can be viewed as a generalized method of moments estimator, normalized to have unit length.

In practice equation (1) is usually normalized by singling out an endogenous variable (y, say) as the "dependent" variable, so that β_Δ takes the form

$$(4) \qquad \beta_\Delta = \begin{bmatrix} 1 \\ -\beta \end{bmatrix},$$

and (3) becomes

$$(5) \qquad \pi_2 = \Pi_2 \beta .$$

This normalization rule implies the compatibility of (1) with (2) (hence that $\operatorname{rank}(\pi_2, \Pi_2) \leqslant n$), and uniqueness here requires $\operatorname{rank}(\Pi_2) = n$. Equation (5) and the condition $\operatorname{rank}(\Pi_2) = n$ together imply $\operatorname{rank}(\pi_2, \Pi_2) = n$, but the converse is clearly not true. The traditional normalization rule therefore seems a stronger assertion about the model (i.e., the joint density of (y, Y)) than the condition $\operatorname{rank}(\pi_2, \Pi_2) = n$. However, equation (3) and the condition $\operatorname{rank}(\pi_2, \Pi_2) = n$ uniquely determine the "direction" of β_Δ, and the normalization (4) is available for β_Δ in any direction except those in which its first element is zero. Since the exceptional set here is of measure zero, equation (3) can be regarded as formally equivalent to equation (5), so that either β or (some suitable measure of) the "direction" of β_Δ can be used to parameterize the problem. Here we focus on the estimation of the "direction" of β_Δ in (3).

Let $Q_i = \beta_\Delta' S_i \beta_\Delta$, $i = 0, 1, 2$, with $S_0 = (y, Y)'(I - Z_1 Z_1')(y, Y)$, $S_1 = (y, Y)'(I - ZZ')(y, Y)$, and $S_2 = (y, Y)' Z_2 Z_2' (y, Y)$. The OLS and TSLS estimators for β in (5) are obtained by minimizing Q_0 and Q_2, respectively, with respect to β_Δ, subject to the normalization rule (4).The LIML procedure leads to the minimization of $\lambda(\beta_\Delta) = Q_2/Q_1$, and, since this ratio is invariant to the length of β_Δ, determines only an estimator for its direction. To define the direction of β_Δ we normalize it so that $\beta_\Delta' \beta_\Delta = 1$, and refer to the resulting estimator as the "symmetrically normalized" estimator.

In Section 4 below we consider, in addition to the estimators for the direction of β_Δ that are induced by the OLS and TSLS estimators for β, analogues of these two estimators defined by minimizing Q_0 and Q_2, respectively, subject to $\beta_\Delta' \beta_\Delta = 1$. These, of course, are simply the unit-length characteristic vectors corresponding to the smallest characteristic roots of S_0 and S_2 respectively.

3. DIRECTION ESTIMATORS

In the context of equation (3), the estimation problem can be thought of as one of determining just the *direction* of the vector β_Δ, and even that can only be determined up to sign. It is convenient to use points on the surface of the unit sphere in $n + 1$ dimensions,

$$S_{n+1}\colon \{\alpha; \alpha \in R^{n+1}, \alpha'\alpha = 1\},$$

to indicate direction, and we shall henceforth replace β_Δ in (3) by α, and treat the estimation problem as one of determining (up to sign) the position of the point α on S_{n+1}. An estimate of α will also be a point on S_{n+1}, and we denote this (random) point by h.

Now, to each point $h=(h_1,h_2')'\in S_{n+1}$, with h_2 $n\times 1$, and $h_1\neq 0$ (i.e., almost everywhere on S_{n+1}), there corresponds a point $r=-h_1^{-1}h_2\in R^n$. The two points h and $-h$ both yield the same r, but apart from this r is unique and ranges over all of R^n as h ranges over S_{n+1}. Conversely, each point $r\in R^n$ determines a pair of points $\pm h$ on S_{n+1}, with h given by

$$(6)\qquad h=\begin{bmatrix}1\\-r\end{bmatrix}(1+r'r)^{-1/2},$$

so that $r=-h_1^{-1}h_2$. If we impose the restriction $h_1>0$ there is thus a one-to-one correspondence between points in R^n and points on (one hemisphere of) S_{n+1}. Hence, a probability density (with respect to Lebesgue measure) on R^n can be thought of as inducing a measure on one hemisphere of S_{n+1} via (6), and we can extend the measure to all of S_{n+1} by requiring that the measures at $+h$ and $-h$ be equal. We evaluate densities on S_{n+1} with respect to the invariant measure, $(h'dh)$, on S_{n+1} and, with the above convention on sign, we have

$$(7)\qquad (h'dh)=2(1+r'r)^{-(n+1)/2}(dr)$$

where (dr) denotes Lebesgue measure on R^n (see Phillips (1984, 1985) and Hillier (1987); the invariant measure $(h'dh)$ on S_{n+1} is defined in Muirhead (1982, Chapter 2)). By construction, the density of h on S_{n+1} induced by that of r will exhibit antipodal symmetry, i.e. $\mathrm{pdf}(h)=\mathrm{pdf}(-h)$. Conversely, of course, any (antipodally symmetric) density on S_{n+1} for h induces a density on R^n for $r=-h_1^{-1}h_2$.

We shall interpret $r\in R^n$ as an estimate of β in (5), and $h\in S_{n+1}$ as either a direct estimate of the point $\alpha\in S_{n+1}$ that satisfies $(\pi_2,\Pi_2)\alpha=0$, or the estimate of α that is induced by the estimate r of β via (6). In the latter case it should be emphasized that, since r and h are essentially one-to-one functions of each other, their respective densities $\mathrm{pdf}(r)$ and $\mathrm{pdf}(h)$ are simply different, but equivalent, ways of summarizing the properties of either one. As we shall see shortly, however, $\mathrm{pdf}(h)$ is much easier to interpret than $\mathrm{pdf}(r)$.[3]

Corresponding to the relations (6) between the estimates r and h we have, for the parameters β and α,

$$(8)\qquad \alpha=\begin{bmatrix}1\\-\beta\end{bmatrix}(1+\beta'\beta)^{-1/2};\qquad \beta=-\alpha_1^{-1}\alpha_2.$$

Note that, in the canonical model, $\beta=0$ if and only if Y is independent of u in equation (1), and in that case α in (8) corresponds to e_1, the first coordinate axis ($e_1'=(1,0,\ldots,0)$, $1\times(n+1)$).

[3] The vector $r\in R^n$ is simply one of several alternative sets of (local) coordinates for the point $h\in S_{n+1}$, albeit the most interesting set (another set of coordinates is described in equation (17) below). Viewed this way, the properties of r are inherited from those of h, so that $\mathrm{pdf}(h)$ becomes the main object of concern and not merely a more convenient way of summarizing the properties of r. A similar point can be made in other models where the normalization rule is arbitrary, e.g., time-series models.

4. DENSITIES AND PROPERTIES: $n = 1$

The exact density functions of the OLS and TSLS estimators for β in the normalized equation were first derived for the case $n = 1$ by Richardson (1961) and Sawa (1968). The corresponding result for the LIML estimator for β was first given by Mariano and Sawa (1972), but the results below for LIML are based on Hillier (1987) and thus differ slightly from those in Mariano and Sawa (1972).

Defining h as in (6) and α as in (8), and using (7), we have, for the OLS and TSLS estimators (cf. Phillips (1983, equation (3.45)):

$$\text{(9)} \qquad \text{pdf}(h) = c_{11}\left[\cos^2\theta_1\right]^{(\nu-1)/2}\exp\{-d^2/2\} \times \sum_{j,k=0}^{\infty}\frac{c_\nu(j,k)}{j!k!}\left[d^2/2\right]^{j+k}\left[\cos^2\theta\right]^j\left[\sin^2\bar{\theta}_1\right]^k$$

where $c_{11} = [\Gamma((\nu+1)/2)/2\Gamma(1/2)\Gamma(\nu/2)]$, $d^2 = \text{tr}[(\pi_2, \Pi_2)'(\pi_2, \Pi_2)]$ is the single nonvanishing characteristic root of $(\pi_2, \Pi_2)'(\pi_2, \Pi_2)$ $(= \Pi_2'\Pi_2(1+\beta^2)$ in the normalized case), θ_1 and $\bar{\theta}_1$ are the angles between h and α, respectively, and the first coordinate axis, e_1, and θ is the angle between h and the true direction α. The coefficients $c_\nu(j,k)$ are given by

$$c_\nu(j,k) = ((\nu+1)/2)_j((\nu-1)/2)_k/(\nu/2)_{j+k},$$

where $\nu = T - K_1$ for OLS, $\nu = K_2$ for TSLS, and $(a)_t = a(a+1)\ldots(a+t-1)$.

For the LIML estimator we have (see equation (A.1) in Appendix A)

$$\text{(10)} \qquad \text{pdf}(h) = (2\pi)^{-1}\exp\{-d^2/2\}\sum_{j,k=0}^{\infty}\frac{a(j,k)}{j!k!}\left[d^2/2\right]^{j+k}\left[\cos^2\theta\right]^j,$$

with coefficients $a(j,k) = a_1(j,k)$ given in equation (A.4) in Appendix A.

The key difference between the results (9) for the OLS/TSLS estimators, and (10) for the LIML estimator, is that the latter depends upon h only through θ, the angle between h and α, while the former depends, in general, upon both θ and θ_1 the angle between h and the first coordinate axis. Only when $\beta = 0$, so that $\alpha \equiv e_1$, or $\nu = 1$ (the exactly identified case for the TSLS estimator), does (9) depend on θ alone. Notice too that (9) also involves $\bar{\theta}_1$, the angle between α and e_1. Evidently the normalization rule (4), which gives special emphasis to the first coordinate axis, plays an important role in determining the properties of the estimators that embody it.

In the totally unidentified case $d^2 = 0$ and (9) and (10) reduce to

$$\text{(11)} \qquad \text{pdf}(h) = c_{11}\left[\cos^2\theta_1\right]^{(\nu-1)/2},$$

$$\text{(12)} \qquad \text{pdf}(h) = (2\pi)^{-1},$$

respectively, the latter being, of course, the uniform distribution on the unit circle.

We now summarize the properties of the direction estimators that follow from (9) and (10).

Properties of the LIML direction estimator:

(i) The density is symmetric about the true points $\pm\alpha$, having modes at $\pm\alpha$ and antimodes at $\pm\bar{\alpha}$, where $\bar{\alpha}$ is orthogonal to α. These properties follow simply from the fact that (10) depends on h only through $\cos^2\theta$, and is an increasing function of $\cos^2\theta$.

(ii) The concentration of the density near $\pm\alpha$ depends only upon d^2, $T-K_1$, and K_2 (the last two through the coefficients $a_1(j,k)$). For the normalized case $d^2 = \Pi_2'\Pi_2(1+\beta^2)$, so that $|\beta|$ and the "concentration parameter" $\Pi_2'\Pi_2$ play essentially the same role.

(iii) In the unidentified case h is uniformly distributed on S_2, i.e., the LIML estimator is completely uninformative about the direction of β_Δ. This result persists asymptotically because (12) holds for all sample sizes (cf. Phillips (1987)).

These properties—in particular, the position of the modes, and the symmetry of the distribution about them—are clearly desirable properties of the LIML estimator. Indeed, for a distribution defined on the circle such symmetry is the natural analogue of unbiasedness for a distribution on R^2, and we shall thus say that, in the case $n=1$, the LIML direction estimator is *spherically unbiased*. As we shall now see, the traditional OLS and TSLS estimators do not, in general, have this property.

Properties of the OLS/TSLS direction estimators:

The density in (9) may be written in the form

$$\text{(13)} \qquad \text{pdf}(h) = ca(h)b(h)$$

with $a(h) = (\cos^2\theta_1)^{(\nu-1)/2}$, $b(h) = g(h)/g(\alpha)$, where

$$\text{(14)} \qquad g(h) = \sum_{j,k=0}^{\infty} \frac{c_\nu(j,k)}{j!k!}\left[d^2/2\right]^{j+k}\left[\cos^2\theta\right]^j\left[\sin^2\bar{\theta}_1\right]^k,$$

and c is a constant. In this decomposition we have $0 \leqslant a(h) \leqslant 1$, with $a(h)$ symmetric about $\pm e_1$, while $b_0 \leqslant b(h) \leqslant 1$, with $b_0 = g(\bar{\alpha})/g(\alpha)$, and $b(h)$ is symmetric about $\pm\alpha$. Note that $a(h)$ depends only upon ν, while $b(h)$ depends upon ν, d^2, and $\bar{\theta}_1$ (hence β^2). The following properties can be deduced fairly easily from this decomposition:

(i) Except when $\beta = 0$ or $\nu = 1$, the density is not symmetric about $\pm\alpha$. The modes are at points between $\pm\alpha$ and $\pm e_1$, and the density is "twisted" away from the true points $\pm\alpha$ towards the first coordinate axis, $\pm e_1$. The antimodes are at $\pm e_2$, $e_2 = (0,1)'$, where the density touches the circle (because $a(h) = 0$ at these points). The position of the modes depends upon ν, d^2, and $|\beta|$.

(ii) The density depends upon $|\beta|$ through the characteristic root d^2 and, independently, through the term $\sin^2\bar{\theta}_1 = \beta^2/(1+\beta^2)$. Thus, d^2 and $|\beta|$ have separate influences on the properties of the estimator.

(iii) In the unidentified case the density is symmetric about modes at $\pm e_1$, and has antimodes at $\pm e_2$. For the TSLS estimator this is also true asymptotically since (11) holds for all sample sizes.

Both the lack of symmetry about the true points $\pm\alpha$, and the fact that the modes do not occur at these points, are clearly undesirable properties of these estimators. As noted above, these properties reflect the dependence of the estimators on an explicit normalization rule.

The effect of the normalization rule is most dramatically brought out in the unidentified case, when the model clearly contains no information about the direction of β_{Δ}. This circumstance is accurately reflected by the properties of the LIML estimator, whatever the sample size. The properties of the OLS and TSLS estimators, on the other hand, are in this case determined *entirely* by the normalization rule, the densities being concentrated around the direction (that of the first coordinate axis) chosen by the normalization rule. For the TSLS estimator this effect is independent of the sample size but is exacerbated by the degree of overidentification $(K_2 - 1)$. For the OLS estimator the effect is exacerbated by increasing sample size.

The general shapes of these densities, supported on the circumference of the unit circle, are depicted in Figure 1. The position of α in relation to e_1 corresponds to $\beta = -1$ (so that α is in the center of the positive quadrant), but otherwise the figure is meant to indicate only the general shapes of the densities: no particular values of $T - K_1$, K_2, or d^2 are implied. Figure 1 makes quite clear, I think, the superiority of the LIML estimator over the OLS/TSLS estimators when these are based on the traditional normalization rule (4). Thus, we next address the question of whether the simple change in the normalization rule suggested in Section 2 is enough to correct these deficiencies.

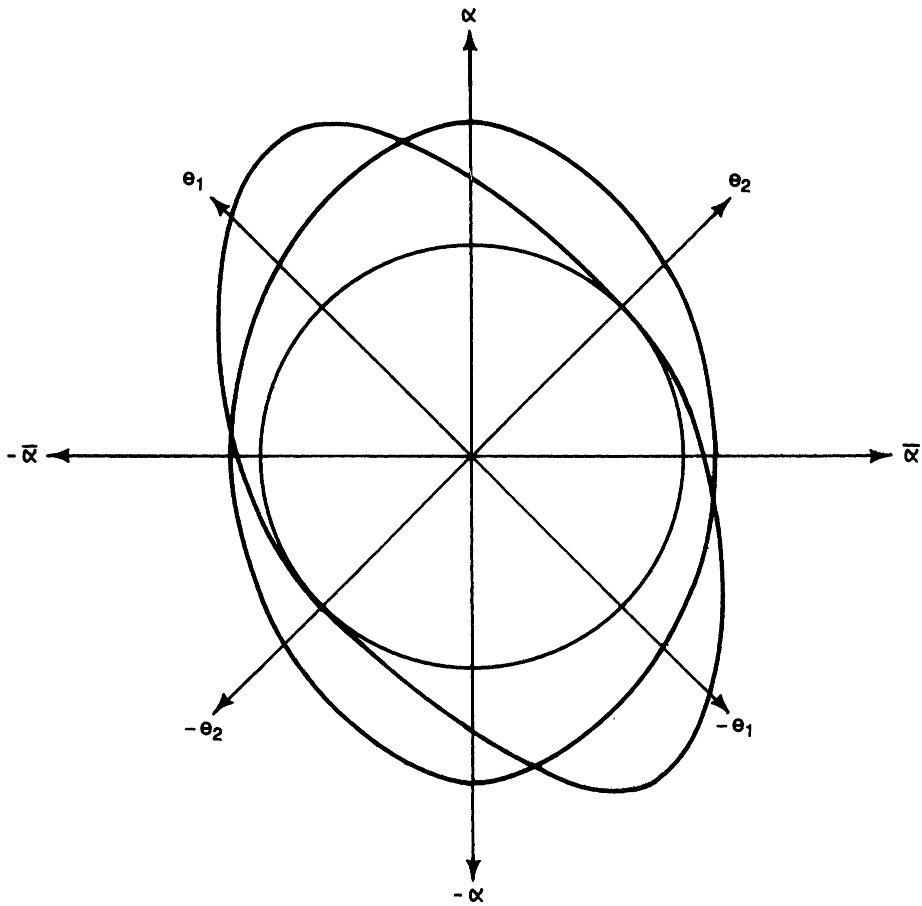
The matrices S_0 and S_2 that define the OLS and TSLS estimators are noncentral Wishart matrices with common covariance matrix I_{n+1}, noncentrality matrix $(\pi_2, \Pi_2)'(\pi_2, \Pi_2)$, and degrees of freedom $\nu = T - K_1$ (for S_0) and $\nu = K_2$ (for S_2). In Appendix B we derive the density (with respect to the invariant measure on S_{n+1}) of the unit-length characteristic vector corresponding to the smallest characteristic root of such a matrix.

It turns out (see equation (B.6) in Appendix B) that these densities are of precisely the same form as that of the LIML estimator in (10), but with different numerical coefficients $a(j,k) = a_2(j,k)$ (equation (B.9) in Appendix B). Hence, the properties of the symmetrically normalized analogues of the OLS/TSLS direction estimators are essentially the same as those of the LIML estimator. In particular, both are symmetrically distributed about the true points $\pm\alpha$, and, in the unidentified case, are uniformly distributed on S_2.

Not surprisingly, the distorted properties of the traditional OLS and TSLS estimators flow from the normalization rule used to define them, not from the statistics (the matrices S_0 and S_2) upon which they are based.

5. THE GENERAL CASE

Earlier results for the case $n = 1$ have, in recent years, been generalized to the case $n > 1$. Results for the OLS and TSLS estimators may be found in Phillips (1980) and Hillier (1985), while those for the LIML estimator may be found in Phillips (1984, 1985) and Hillier (1987). Using the decomposition



Note: The position of α corresponds to β = −1

FIGURE 1.—Densities of OLS/TSLS and LIML direction estimators: $n = 1$.

$(\pi_2, \Pi_2) = ADL'$ given in Appendix B, the totally unidentified case corresponds to $D = 0$, and equations (11) and (12) generalize easily in this case to give

$$\text{(15)} \qquad \text{pdf}(h) = c_{1n}\left[\cos^2\theta_1\right]^{(\nu-n)/2}$$

for the OLS and TSLS estimators with

$$c_{1n} = \left[\Gamma((\nu+1)/2)/2\pi^{n/2}\Gamma((\nu-n+1)/2)\right], \qquad \text{and}$$

$$\text{pdf}(h) = \Gamma((n+1)/2)/2\pi^{(n+1)/2}$$

(the uniform distribution on S_{n+1}) for the LIML estimator. The statements in Section 4 about the properties of these estimators in the totally unidentified case when $n = 1$ thus generalize completely when $n > 1$.

Since (15) is the leading term in the series expansion of the OLS/TSLS density when $n > 1$, the distortion of the properties of these estimators noted in

Section 4 for the case $n=1$ clearly persists when $n>1$. The interesting question, therefore, is whether the nice symmetry of the distributions of the LIML and symmetrically normalized OLS/TSLS estimators about the true point α generalizes to the case $n>1$. We shall now show that, in general, it does not, but that there is a less obvious sense in which the result for the case $n=1$ does generalize.

The densities of the LIML and symmetrically normalized OLS/TSLS direction estimators are of the form (generalizing equation (10)):

$$\text{(16)} \qquad \operatorname{pdf}(h)=\left[\Gamma((n+1)/2)/2\pi^{(n+1)/2}\right]\operatorname{etr}\{-D^2/2\} \times \sum_{\alpha,\kappa;\phi}\frac{a(\alpha,\kappa;\phi)}{j!k!}\theta_\phi^{\alpha,\kappa}C_\phi^{\alpha,\kappa}(DL'VV'LD/2, D^2/2)$$

with $a(\alpha,\kappa;\phi)=a_1(\alpha,\kappa;\phi)$ (equation (A.3) in Appendix A) for the LIML estimator and $a(\alpha,\kappa;\phi)=a_2(\alpha,\kappa;\phi)$ (equation (B.8) in Appendix B) for the OLS/TSLS estimators. Here $VV'=I-hh'$ and $LL'=I-\alpha\alpha'$, while $D=\operatorname{diag}\{d_1,d_2,\dots,d_n\}$, where $d_1^2\leqslant d_2^2\leqslant\cdots\leqslant d_n^2$ are the nonzero characteristic roots of $(\pi_2,\Pi_2)'(\pi_2,\Pi_2)$. The scalar powers $(d^2\cos^2\theta/2)^j(d^2/2)^k$ in (10) are, in the general case, replaced by the invariant polynomials $C_\phi^{\alpha,\kappa}(DL'VV'LD/2, D^2/2)$ (see Davis (1979) and Chikuse and Davis (1986) for background and notation).

The symmetry noted for the case $n=1$ arises from the fact that (10) depends upon h only through θ, where θ is the angle between h and α. Now, any point $h\in S_{n+1}$ other than $\pm\alpha$ may be represented (uniquely) in the form

$$\text{(17)} \qquad h=\alpha\cos\theta+L\omega\sin\theta$$

where α, L, and θ are as above, $0<\theta<\pi$, and $\omega\in S_n$. Here $L\omega$ is the unit vector lying along the orthogonal projection of h onto the space orthogonal to α spanned by the columns of L, and θ and ω are a set of coordinates for h. That is, any point $h\in S_{n+1}$ can be located by the angle (θ) it makes with an arbitrary fixed point on S_{n+1} (in this case, α) and its direction in the n-dimensional space orthogonal to α (indicated by ω). The density of h can be expressed as a function of θ and ω (we could, but do not, transform $h\to(\theta,\omega)$) and is constant on the surface $\theta=$ constant if and only if it does not depend on ω or, what is the same thing, it is invariant under $\omega\to H\omega$, $H\in O(n)$ ($O(n)$ denotes the group of all $n\times n$ orthogonal matrices).

Using (17), the matrix $L'VV'L$ that occurs in (16) becomes $L'VVL=L'(I-hh')L=I_n-\sin^2\theta\,\omega\omega'$. Now, consider a rotation of $\omega:\omega\to H\omega, H\in O(n)$. Since the polynomials $C_\phi^{\alpha,\kappa}(A,B)$ are invariant under $A\to H'AH$, $B\to H'BH$, $H\in O(n)$, we have

$$\begin{aligned} &C_\phi^{\alpha,\kappa}(D^2-\sin^2\theta D\omega\omega'D, D^2)\\ &\quad\to C_\phi^{\alpha,\kappa}(D^2-\sin^2\theta DH\omega\omega'H'D, D^2)\\ &\quad= C_\phi^{\alpha,\kappa}(H'D^2H-\sin^2\theta H'DH\omega\omega'H'DH, H'D^2H).\end{aligned}$$

Hence, $\text{pdf}(h)$ in (16) is invariant under rotations of ω if and only if D is invariant under $D \to H'DH$. If the diagonal elements of D (or, equivalently, the n nonzero characteristic roots of $(\pi_2, \Pi_2)'(\pi_2, \Pi_2)$) are distinct the only matrices $H \in O(n)$ with this property are the 2^n matrices $H = \text{diag}(\pm 1, \pm 1, \ldots, \pm 1)$.

Thus, if the characteristic roots of $(\pi_2, \Pi_2)'(\pi_2, \Pi_2)$ are distinct the density (16) is *not* constant on the surface $\theta = \text{constant}$, but for fixed θ there are 2^n directions in which the density is the same, namely, those generated by replacing ω in (17) by $H\omega$, with $H = \text{diag}\{\pm 1, \ldots, \pm 1\}$. Since $\text{pdf}(h) = \text{pdf}(-h)$, there are thus 2^{n+1} points on S_{n+1} at which the density is the same, and the symmetry noted for the case $n = 1$ is simply the degenerate case of this less compelling symmetry of the distribution (16).

The condition that D be invariant under $D \to H'DH$ for all $H \in O(n)$ is satisfied only when $D = \lambda I_n$, i.e., when the nonzero characteristic roots of $(\pi_2, \Pi_2)'(\pi_2, \Pi_2)$ are all equal. In this case the density (16) *is* constant on the surface $\theta = \text{constant}$ (the polynomial $C_\phi^{\alpha,\kappa}(\lambda^2(I_n - \sin^2 \theta\omega\omega'), \lambda^2 I_n)$ is a polynomial of degree j in $\cos^2 \theta$). Thus, in this special case the symmetry noted for the case $n = 1$ does generalize; in general it does not.

Evidently, the properties of the LIML and symmetrically normalized OLS/TSLS estimators depend on both the magnitude of the characteristic roots of $(\pi_2, \Pi_2)'(\pi_2, \Pi_2)$, and on the dispersion of those roots.

6. CONCLUDING REMARKS

An interesting aspect of the results in Section 4 is that the traditional normalization rule is innocuous when $\beta = 0$, i.e., when the right-hand-side endogenous variables are actually weakly exogenous. This result is consistent with both other evidence (see, e.g., Hillier (1985) and Phillips (1983)) and with the intuition that procedures analogous to ordinary regression methods should work well when the right-hand-side variables are not genuinely endogenous. In general, however, our results also confirm other evidence (Anderson et al. (1986, 1982, 1973)) that this analogy cannot be relied upon when the model really involves endogenous variables, and help to explain why this is so.

Since $\text{pdf}(h)$ and $\text{pdf}(r)$ are merely different ways of summarizing the properties of either h or r, the properties of (say) r can be deduced from $\text{pdf}(h)$. For instance, it is easy to show, using $\text{pdf}(h)$, that for $n = 1$ and all five of the estimators considered here, the median of r lies between β and the origin, a result consistent with the numerical results in Anderson et al. (1982, 1973). The densities (10) and (16) seem to be the most promising objects for further study in this model.

Department of Econometrics, Monash University, Clayton, Melbourne, Victoria 3168, Australia

Manuscript received March, 1988; final revision received August, 1989.

APPENDIX A

THE COEFFICIENTS IN THE LIML DENSITY

Equation (41) in Hillier (1985) may be used to obtain the following expression for the density of the LIML estimator, r, for β in (4):

$$\text{(A.1)} \quad \operatorname{pdf}(r) = \Gamma((n+1)/2)[\pi(1+r'r)]^{-(n+1)/2}\operatorname{etr}\{-D^2/2\}[C/\Gamma_n((n+1)/2)]$$

$$\times \int_{f_1>0}\int_{R>0} f_1^{m(n+1)/2-1}(1+f_1)^{-(m+K_2)(n+1)/2}\operatorname{etr}(-R)\,{}_2F_2((m+K_2)/2,$$

$$(n+3)/2;(m+n+2)/2,(n+1)/2;f_1R/(1+f_1))$$

$$\times \sum_{\alpha,\kappa;\phi} \frac{((m+K_2-n)/2)_\kappa}{j!k!(K_2/2)_\phi}\frac{C_\alpha(R)}{C_\alpha(I)}\theta_\phi^{\alpha,\kappa}C_\phi^{\alpha,\kappa}(DL'VV'LD/2, D^2/2)$$

$$\div (1+f_1)^{j+k}(dR)df_1$$

where $m = T - K$, D and L are as defined in Appendix B, $V' = (I + rr')^{-1/2}(r, I)$,

$$\text{(A.2)} \quad C = \frac{\Gamma_{n+1}((m+K_2)/2)\Gamma_n((n+3)/2)\Gamma_n((m-1)/2)\pi^{(n+1)/2}}{\Gamma_{n+1}(K_2/2)\Gamma_{n+1}(m/2)\Gamma_n((m+n+2)/2)\Gamma((n+1)/2)},$$

and the notation $R > 0$ indicates that the integral in (A.1) is over the space of real positive definite symmetric matrices. This can be converted to the density of h by simply noting that $VV' = I - hh'$ and using (7). Hence, the coefficients $a_1(\alpha, \kappa; \phi)$ in equation (16) in the text are given by

$$\text{(A.3)} \quad a_1(\alpha,\kappa;\phi) = C\big[((m+K_2-n)/2)_\kappa/(K_2/2)_\phi\Gamma_n((n+1)/2)\big]$$

$$\times \int_{f_1>0}\int_{R>0}\operatorname{etr}(-R)[C_\alpha(R)/C_\alpha(I_n)]f_1^{m(n+1)/2-1}$$

$$\times (1+f_1)^{-(f+(m+K_2)(n+1)/2)}\,{}_2F_2((m+K_2)/2,(n+3)/2;$$

$$(m+n+2)/2,(n+1)/2;f_1R/(1+f_1))df_1(dR)$$

$$= C\left[\frac{((m+K_2-n)/2)_\kappa\Gamma(m(n+1)/2)\Gamma(f+K_2(n+1)/2)}{(K_2/2)_\phi\Gamma(f+(m+K_2)(n+1)/2)}\right]$$

$$\times \sum_{l=0}^{\infty}\sum_{\lambda}\frac{((m+K_2)/2)_\lambda((n+3)/2)_\lambda(m(n+1)/2)_l}{l!((m+n+2)/2)_\lambda((n+1)/2)_\lambda(f+(m+K_2)(n+1)/2)_l}$$

$$\times \sum_{\rho\in\alpha.\lambda}((n+1)/2)_\rho\big(\theta_\rho^{\alpha,\lambda}\big)^2 C_\rho(I)/C_\alpha(I)$$

where $f = j + k$, λ is a partition of l, and ρ is a partition of $j + l$.

When $n = 1$ (A.3) can be simplified to

$$\text{(A.4)} \quad a_1(j,k) = \frac{\Gamma((m+1)/2)\Gamma((K_2+1)/2)\Gamma((m+K_2-1)/2)}{2\Gamma((K_2-1)/2)\Gamma((m+3)/2)\Gamma((m+K_2+1)/2)}$$

$$\times \frac{(1)_j((m+K_2-1)/2)_k(K_2)_f}{(K_2/2)_f(m+K_2)_f}\,{}_4F_3((m+K_2)/2,2,m,j+1;$$

$$(m+3)/2,1,f+m+K_2;1).$$

In particular, $a_1(0,0) = 1$.

APPENDIX B

DISTRIBUTION OF THE CHARACTERISTIC VECTORS CORRESPONDING TO THE SMALLEST CHARACTERISTIC ROOTS OF S_0 AND S_2

Let $S \sim W_{n+1}(\nu, I_{n+1}, (\pi_2, \Pi_2)'(\pi_2, \Pi_2))$. Since (π_2, Π_2) is of rank n we may write

$$(\text{B.1}) \qquad (\pi_2, \Pi_2) = ADL'$$

where A $(K_2 \times n)$ satisfies $A'A = I_n$, D is a diagonal matrix $D = \text{diag}\{d_1, d_2, \ldots, d_n\}$, where $0 < d_1^2 \leqslant d_2^2 \leqslant \cdots \leqslant d_n^2$ are the n nonzero characteristic roots of $(\pi_2, \Pi_2)'(\pi_2, \Pi_2)$, and L $(n+1 \times n)$ satisfies $L'L = I_n$. This decomposition is unique if the roots are distinct and the elements in, say, the first row of L are taken to be positive. The matrix L determines α in the equation $(\pi_2, \Pi_2)\alpha = 0$ uniquely up to sign, and in the normalized case we can take

$$(\text{B.2}) \qquad L = \begin{bmatrix} \beta' \\ I_n \end{bmatrix} (I + \beta\beta')^{-1/2}.$$

Using (B.1) we have (cf. Muirhead (1982, Section 10.3))

$$(\text{B.3}) \qquad \text{pdf}(S) = C_2 \,\text{etr}\{-D^2/2\}\,\text{etr}\{-S/2\}|S|^{(\nu-n-2)/2}\,{}_0F_1(\nu/2, DL'SLD/4)$$

with $C_2 = [2^{(n+1)\nu/2}\Gamma_{n+1}(\nu/2)]^{-1}$. We now transform from S to its characteristic roots and vectors: $S = HFH'$ (see Muirhead (1982, Theorem 3.2.17) for technical details and notation). We have

$$(dS) = (H'dH)\prod_{i<j}^{n}(f_i - f_j)\prod_{i=1}^{n}(f_i - f)\prod_{i=1}^{n} df_i\, df$$

where $F = \text{diag}\{f_1, f_2, \ldots, f_n, f\}$, $f_1 > f_2 > \cdots > f_n > f > 0$, $H \in O(n+1)$, and $(H'dH)$ denotes the unnormalized invariant measure on the orthogonal group $O(n+1)$.

Next, partition $H = (H_1, h)$, where $H_1(n+1 \times n)$ satisfies $H_1'H_1 = I_n$ and $H_1'h = 0$, and

$$F = \begin{pmatrix} F_1 & 0 \\ 0 & f \end{pmatrix},$$

with $F_1 = \text{diag}\{f_1, \ldots, f_n\}$. We need to integrate out H_1, F_1, and f. Now, for fixed h we may set $H_1 = VH_2$, where V is a fixed matrix satisfying $V'V = I_n$, $V'h = 0$, and $H_2 \in O(n)$. The set $\{H_1; H_1 = VH_2, H_2 \in O(n)\}$ is identical to the set $\{H_1; H_1'H_1 = I_n, H_1'h = 0\}$, the relationship is one-to-one, and we have (cf. Muirhead (1982, p. 397))

$$(H'dH) = (h'dh)(H_2'dH_2).$$

To facilitate the integration with respect to H_2 we write the Bessel function in (B.3) as an inverse Laplace transform:

$$(\text{B.4}) \qquad {}_0F_1(\nu/2, DL'SLD/4) = a_n \int_{\text{Re}(W)>0} \text{etr}(W)|W|^{-\nu/2}\,\text{etr}\{DL'SLDW^{-1}/4\}\,(dW)$$

where $a_n = \Gamma_n(\nu/2)2^{n(n-1)/2}/(2\pi i)^{n(n+1)/2}$. Writing

$$S = VH_2F_1H_2'V' + fhh' = VH_2(F_1 - fI_n)H_2'V' + fI_{n+1},$$

and integrating out H_2 gives,

$$(\text{B.5}) \qquad \text{pdf}(h, F) = C_2^* \,\text{etr}\{-D^2/2\}\,\text{etr}\{-F/2\}|F|^{(\nu-n-2)/2}$$
$$\times a_n \int_{\text{Re}(W)>0} \text{etr}(W)|W|^{-\nu/2}\,\text{etr}\{fD^2W^{-1}/4\}$$
$$\times {}_0F_0^{(n)}\big((F_1 - fI_n)/2, DL'VV'LDW^{-1}/2\big)(dW),$$

with $C_2^* = \pi^{n^2/2}C_2/2\Gamma_n(n/2)$ (for the other notation see Muirhead (1982, Chapter 7)).

The inverse Laplace transform in (B.5) may be evaluated by using results from Davis (1979), and we then find that the distribution of h has the form

$$\text{(B.6)} \qquad \operatorname{pdf}(h) = \left[\Gamma((n+1)/2)/2\pi^{(n+1)/2}\right] \operatorname{etr}\{-D^2/2\} \times \sum_{\alpha,\kappa;\phi} \frac{a_2(\alpha,\kappa;\phi)}{j!k!} \theta_\phi^{\alpha,\kappa} C_\phi^{\alpha,\kappa}(DL'VV'LD/2, D^2/2),$$

with coefficients $a_2(\alpha, \kappa; \phi)$ given by

$$\text{(B.7)} \qquad a_2(\alpha,\kappa;\phi) = \left[2\pi^{(n+1)/2} C_2^*/2^{j+k}\Gamma((n+1)/2)(\nu/2)_\phi\right] \times \iint_{f_1 > \cdots > f_n > f} \operatorname{etr}\{-F/2\}|F|^{(\nu-n-2)/2} f^k \left[C_\alpha(F_1 - fI_n)/C_\alpha(I_n)\right] \times \prod_{i<j}^{n}(f_i - f_j) \prod_{i=1}^{n}(f_i - f) \prod_{i=1}^{n} df_i\, df.$$

To evaluate the coefficients $a_2(\alpha, \kappa; \phi)$ explicitly, put $\bar{f}_1 = (f_i - f)/f$, $i = 1, \ldots, n$ ($\bar{f}_1 > \bar{f}_2 > \cdots > \bar{f}_n > 0$, $df_i = f d\bar{f}_i$ so that

$$\text{(B.8)} \qquad a_2(\alpha,\kappa;\phi) = \left[2\pi^{(n+1)/2} C_2^*/2^{j+k}\Gamma((n+1)/2)(\nu/2)_\phi\right] \times \iint_{\bar{f}_1 > \cdots > \bar{f}_n > 0} \int_{f>0} \operatorname{etr}\{-f(I+\bar{F}_1)/2\} \exp\{-f/2\} f^{j+k+(n+1)\nu/2-1} \times |\bar{F}_1|\,|I+\bar{F}_1|^{(\nu-n-2)/2}\left[C_\alpha(\bar{F}_1)/C_\alpha(I_n)\right] \prod_{i<j}^{n}(\bar{f}_i - \bar{f}_j) \prod_{i=1}^{n} d\bar{f}_i\, df$$

$$= \left[\pi^{(n+1)/2} C_2/2^{j+k}\Gamma((n+1)/2)(\nu/2)_\phi\right] \times \int_{f>0}\int_{R>0} \operatorname{etr}\{-f(I+R)/2\} \exp\{-f/2\} f^{j+k+(n+1)\nu/2-1} \times |R|\,|I+R|^{(\nu-n-2)/2}\left[C_\alpha(R)/C_\alpha(I_n)\right](dr)\,df$$

(compare equation (A.3) in Appendix A).

In the case $n = 1$ it is straightforward to evaluate the $a_2(j, k)$ in (B.8) and we find, for $n = 1$,

$$\text{(B.9)} \qquad a_2(j,k) = \left[\Gamma((\nu+1)/2)/2\Gamma((\nu+3)/2)\right] \times \left[(2)_j((\nu-1)/2)_k(\nu)_{j+k}/2^{j+k}(\nu/2)_{j+k}((\nu+3)/2)_{j+k}\right] {}_2F_1(j+2, j+k+\nu, j+k+(\nu+3)/2, 1/2).$$

In particular, $a_2(0,0) = 1$ since ${}_2F_1(2, \nu, (\nu+3)/2, 1/2) = 2\Gamma((\nu+3)/2)/\Gamma((\nu+1)/2)$.

For arbitrary n the coefficients $a_2(\alpha, \kappa; \phi)$ are much more complicated and we omit the details.

REFERENCES

Anderson, T. W., N. Kunimoto, and K. Morimune (1986): "Comparing Single Equation Estimators in a Simultaneous Equation System," *Econometric Theory*, 2, 1–32.

Anderson, T. W., N. Kunimoto, and T. Sawa (1982): "Evaluation of the Distribution Function of the Limited Information Maximum Likelihood Estimator," *Econometrica*, 50, 1009–1027.

Anderson, T. W., and T. Sawa (1973): "Evaluation of the Distribution Function of the Two-stage Least Squares Estimate," *Econometrica*, 25, 77–83.

Chikuse, Y., and A. W. Davis (1986): "A Survey on the Invariant Polynomials with Matrix Arguments in Relation to Econometric Distribution Theory," *Econometric Theory*, 2, 232–248.

Davis, A. W. (1979): "Invariant Polynomials with Two Matrix Arguments, Extending the Zonal Polynomials: Applications to Multivariate Distribution Theory," *Annals of the Institute of Statistical Mathematics*, A, 31, 465–485.

HILLIER, G. H. (1985): "On the Joint and Marginal Densities of Instrumental Variable Estimators in a General Structural Equation," *Econometric Theory*, 1, 53–72.

——— (1987): "Joint Distribution Theory for Some Statistics Based on LIML and TSLS," Cowles Foundation Discussion Paper No. 840, June, 1987.

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.

MUIRHEAD, R. J. (1982): *Aspects of Multivariate Statistical Theory*. New York: Wiley.

PHILLIPS, P. C. B. (1980): "The Exact Finite Sample Density of Instrumental Variable Estimators in an Equation with $n+1$ Endogenous Variables," *Econometrica*, 48, 861–878.

——— (1983): "Exact Small Sample Theory in the Simultaneous Equations Model," in *Handbook of Econometrics*, ed. by M. D. Intriligator and Z. Griliches. Amsterdam: North Holland.

——— (1984): "The Exact Distribution of LIML: I," *International Economic Review*, 25, 249–261.

——— (1985): "The Exact Distribution of LIML: II," *International Economic Review*, 26, 21–36.

——— (1987): "Partially Identified Econometric Models," Cowles Foundation Discussion Paper No. 845.

RICHARDSON, D. H. (1961): "The Exact Distribution of a Structural Coefficient Estimator," *Journal of the American Statistical Association*, 63, 1214–1226.

SAWA, T. (1968): "The Exact Sampling Distribution of Ordinary Least Squares and Two-stage Least Squares Estimators," *Journal of the American Statistical Association*, 64, 923–927.