MONASH

WP 5/89

MONASH UNIVERSITY

AUSTRALIA

ON THE NORMALISATION OF STRUCTURAL EQUATIONS: PROPERTIES OF DIRECTION ESTIMATORS

Grant H. Hillier

Working Paper No. 5/89

May 1989

DEPARTMENT OF ECONOMETRICS

ISSN 1032-3813

ISBN 0 86746 799 1

ON THE NORMALISATION OF STRUCTURAL EQUATIONS: PROPERTIES OF DIRECTION ESTIMATORS

Grant H. Hillier

Working Paper No. 5/89

May 1989

DEPARTMENT OF ECONOMETRICS, FACULTY OF ECONOMICS AND POLITICS

MONASH UNIVERSITY, CLAYTON, VICTORIA 3168, AUSTRALIA.

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

by

Grant H. Hillier

May 1989

---

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 Associate Editor for comments on the first version of the paper that helped to both clarify the message and broaden the scope of the paper.

Abstract

In the classical structural equation model only the direction of the vector of coefficients of the endogenous variables is determined. The traditional normalisation rule defines the coefficients that are of interest but should not be embodied in on 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 normalisation rule. Properly normalised analogues of these estimators are defined and are shown to have essentially similar properties to those of the Limited Information Maximum Likelihood estimator.

Keywords: Direction estimators; Exact Distributions; Induced densities; Normalisation rules.

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 [3] and Anderson, Kunimoto, and Sawa [2], 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 [1] 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 latter 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 normalisation 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 unnormalised equation. Existing distribution results for the coefficients in a normalised equation are easily translated 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 normalised case.

The results we obtain suggest that the normalisation 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 normalisation rule as the LIML estimator would correct this distortion, and we show below that this is indeed the case.[2] Thus, while the traditional normalisation rule can certainly serve to define what is of particular interest in the model - usually, the coefficients of the right-hand-side endogenous variables - it is neither logically necessary nor, as we shall show, wise, to impose this rule on the estimation procedure.

2. MODEL AND ASSUMPTIONS

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

$$(y, Y)\beta_{\Delta} = Z_1\gamma + u , \tag{1}$$

with corresponding reduced form

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

where y is Tx1, Y is Txn, Z_1 is TxK_1, Z_2 is TxK_2, $K_2 \geq n$, 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 [11] 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 compatable with (2) if and only if the relations $\gamma = (\pi_1,\Pi_1)\beta_\Delta$,

$$(\pi_2,\Pi_2)\beta_\Delta \;=\; 0 \;, \tag{3}$$

and $u = (v,V)\beta_\Delta$, hold for some $\beta_\Delta \neq 0$. Equation (3) implies $\text{rank}(\pi_2,\Pi_2) \leq n$, and to ensure an essentially unique solution requires the identification condition $\text{rank}(\pi_2,\Pi_2) = n$. This is the maintained hypothesis.

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

$$\beta_\Delta \;=\; \begin{bmatrix}1\\ -\beta\end{bmatrix}, \tag{4}$$

and (3) becomes

$$\pi_2 \;=\; \Pi_2\beta \;. \tag{5}$$

This normalisation rule implies the compatability of (1) with (2) (hence that $\text{rank}(\pi_2,\Pi_2) \leq n$), and uniqueness here requires $\text{rank}(\Pi_2) = n$.

Equation (5) and the condition $\text{rank}(\Pi_2) = n$ together imply $\text{rank}(\pi_2, \Pi_2) = n$, but the converse is clearly not true. The traditional normalisation rule therefore seems a stronger assertion about the model (i.e., the joint density of (y, Y)) than the condition $\text{rank}(\pi_2, \Pi_2) = n$. However, equation (3) determines only the 'direction' of the vector β_Δ, and the set of directions in which the traditional normalisation is impossible is of measure zero. Hence, the 'primitive' equation (3) can be regarded as formally equivalent to equation (5), and we shall focus on the estimation of the direction of the vector β_Δ in (3).

Let $Q_i = \beta_\Delta' S_i \beta_\Delta$, $i = 0, 1, 2$, with $S_0 = (y,Y)'(I - Z_1Z_1')(y,Y)$, $S_1 = (y,Y)'(I - ZZ')(y,Y)$, and $S_2 = (y,Y)'Z_2Z_2'(y,Y)$. The OLS and TSLS estimators for β in (4) are obtained by minimising Q_0 and Q_2, respectively, with respect to β_Δ, subject to the normalisation rule (4). The LIML procedure leads to the minimisation 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. If, as we shall in the next section, and is common in practice, 'direction' is indicated by position on the surface of the unit $(n+1)$-sphere, the LIML estimator for the direction of β_Δ minimises $\lambda(\beta_\Delta)$ subject to $\beta_\Delta'\beta_\Delta = 1$.

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 minimising 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

Equation (3) is clearly incapable of determining the *length* of β_Δ. Thus, in the context of the primitive 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,

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

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 estimator for α will also be a point, h say, on S_{n+1}.

Now, to each point $h = (h_1, h_2')' \in S_{n+1}$, with h_2 nx1, 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

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

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 S_{n+1} via (6), and we can drop the requirement that $h_1 > 0$ by allocating the mass at r equally to the two points $\pm h$. We evaluate measures 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

$$(h'dh) = 2(1 + r'r)^{-(n+1)/2} \bigwedge_{i=1}^{n} dr_i \tag{7}$$

(see Phillips [12], [13], and Hillier [7]; the invariant measure $(h'dh)$ on S_{n+1} is defined in Muirhead [9, Chapter 2], as is the exterior product notation 'Λ'). By construction, the measure on S_{n+1} for h induced by that of r will exhibit antipodal symmetry, i.e. pdf(h) = pdf(-h).

Given an estimator, r, for β in the normalised equation, we may thus construct an estimator, h, for the point $\alpha \in S_{n+1}$ satisfying $(\pi_2, \Pi_2)\alpha = 0$ via (6), and the density of r induces a probability measure for h on S_{n+1}. With a slight abuse of notation we shall denote both measures on R^n (w.r. to Lebesgue measure), and measures on S_{n+1} (w.r. to the invariant measure) by the same symbol, pdf($\cdot$), although the latter are not, strictly speaking, density functions.

Likewise, of course, a measure for the direction estimator h will induce a measure for the coefficient estimator r, and the properties of r will reflect those of h. Thus, although I would argue that it is the direction of β_Δ that is the proper focus of attention in this model, if interest is nevertheless centered on β in the normalised equation, the properties of the direction estimator are pertinent because they explain the behaviour of r. For instance, in the case n = 1, r is simply the slope of the line orthogonal to ±h, and, in the representation

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

for the parameters corresponding to (6), β is the slope of the line orthogonal to $\pm\alpha$. Loosely speaking, getting h right is a prerequisite for getting r right, and of course this heuristic argument can be both formalised and generalised to the case $n > 1$. We shall not pursue this formality here, though, because our main point is that interpreting the estimation problem as one of estimating a direction is both more natural and more productive.

Before proceeding we 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 ($e_1' = (1, 0, \ldots, 0)$, $1 \times (n+1)$).

4. DENSITIES AND PROPERTIES: n=1

The exact density functions of the OLS and TSLS estimators for β in the normalised equation were first derived for the case n=1 by Richardson [15] and Sawa [16]. The corresponding result for the LIML estimator for β was first given by Mariano and Sawa [8], but the results below for LIML are based on [7] and thus differ slightly from those in [8].

Defining h as in (6) and α as in (8), and using (7), we have, for the OLS and TSLS estimators (cf. Phillips [11], equation (3.45)):

$$\mathrm{pdf}(h) = C_{11}[\cos^2\theta_1]^{(\nu-1)/2}\exp\{-d^2/2\}$$

$$\sum_{j,k=0}^{\infty} \frac{c_\nu(j,k)}{j!\ k!}[d^2/2]^{j+k}[\cos^2\theta]^j[\sin^2\bar{\theta}_1]^k \qquad (9)$$

where $C_{11} = [\Gamma((\nu+1)/2)/2\Gamma(1/2)\Gamma(\nu/2)]$, $d^2 = \mathrm{tr}[(\pi_2,\Pi_2)'(\pi_2,\Pi_2)]$ is the single non-vanishing characteristic root of $(\pi_2,\Pi_2)'(\pi_2,\Pi_2)$ ($= \Pi_2'\Pi_2(1+\beta^2)$ in the normalised 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 (cf. [7], equation (52)):

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

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 normalisation 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

$$\mathrm{pdf}(h) = C_{11}[\cos^2\theta_1]^{(\nu-1)/2} , \tag{11}$$

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

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 normalised 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 [14]).

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

$$\mathrm{pdf}(h) \;=\; c\; a(h)\; b(h) \tag{13}$$

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

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

and c is a constant. In this decomposition we have $0 \le a(h) \le 1$, with $a(h)$ symmetric about $\pm e_1$, while $b_0 \le b(h) \le 1$, 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 normalisation rule.

The effect of the normalisation 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 normalisation rule, the densities being concentrated around the direction (that of the first coordinate axis) chosen by the normalisation 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 centre 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 2 shows the components a(h), b(h) in equation (13), and their product a(h)b(h). To simplify the picture these are graphed with θ, the angle (measured in a clockwise direction) between h and α, on the horizontal axis, and are again stylized rather than an exact representation for particular values of ν and d^2.

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 normalisation rule (4). Thus, we next address the question of whether the simple change in the normalisation 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 non-central Wishart matrices with common covariance matrix I_{n+1}, non-centrality 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 properly normalised 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 normalisation rule used to define them, not from the statistics (the matrices S_0 and S_2) upon which they are based.

The median of r

Any subset of R^n evidently has an image on S_{n+1}, and its probability content may thus be calculated from either the density of r, or the density of h, as convenient. For example, it is easy to see that, when n = 1,

$$\mathrm{pr}\{r < \beta\} = \mathrm{pr}\{-h_1^{-1}h_2 < \beta\} = 2\mathrm{pr}\{0 < \theta < \pi;\ -\pi/2 < \theta_1 < \pi/2\}$$

$$= 2\mathrm{pr}\{h \in B^+\}\ ,\ \text{say}\ ,$$

where

$$B^+ = \{h\ ;\ h \in S_{n+1};\ 0 < \theta < \pi;\ -\pi/2 < \theta_1 < \pi/2\} \tag{15}$$

(see Fig. 1).

In the case of the LIML and properly normalised OLS/TSLS estimators the symmetry of the distributions of h about $\pm\alpha$, together with the observation that B^+ occupies an arc of length less than (more than) $\pi/2$ when $\beta < 0$ ($\beta > 0$), implies that, in each case,

$$\left.\begin{array}{l} \mathrm{pr}\{r < \beta\} \gtrless 1/2 \quad \text{as} \quad \beta \gtrless 0 \\ \mathrm{pr}\{r < 0\} \gtrless 1/2 \quad \text{as} \quad \beta \lessgtr 0 \end{array}\right\} \tag{16}$$

and $\mathrm{pr}\{r < \beta\} = 1/2$ just if $\beta = 0$ (assuming $d^2 \neq 0$). The same conclusion can be obtained for the direction estimators induced by the traditional OLS/TSLS estimators,[3] so that, for all five estimators, *the median lies between the true value of* β *and the origin.*

This result is consistent with the numerical results in [2] and [3], but these studies also indicate that the median of the LIML estimator is much closer to β than it is for the (traditional) OLS/TSLS estimators. This suggests that the density of the LIML estimator is much more concentrated near α than is indicated in Fig. 1, and thus adds further weight to our earlier conclusions.

5. THE GENERAL CASE

Earlier results for the case n=1 have, in recent years, been generalised to the case $n > 1$. Results for the OLS and TSLS estmators may be found in [10] and [6], while those for the LIML estimator may be found in [12], [13], and [7]. Using the decomposition $(\pi_2, \Pi_2) = ADL'$ given in Appendix B, the totally unidentified case corresponds to $D = 0$, and equations (11) and (12) generalise easily in this case to give

$$\mathrm{pdf}(h) = C_{1n}[\cos^2\theta_1]^{(\nu-n)/2} \tag{17}$$

for the OLS and TSLS estimators, with $C_{1n} = [\Gamma((\nu+1)/2)/2\pi^{n/2}\Gamma((\nu-n+1)/2)]$, and $\mathrm{pdf}(h) = \Gamma((n+1)/2)/2\pi^{(n+1)/2}$ (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 generalise completely when $n > 1$.

Since (17) 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 properly normalised OLS/TSLS estimators about the true point α generalises 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 generalise.

The densities of the LIML and properly normalised OLS/TSLS direction estimators are of the form (generalising equation (10)):

$$\text{pdf}(h) = [\Gamma((n+1)/2)/2\pi^{(n+1)/2}] \,\text{etr}\{-D^2/2\}$$

$$\sum_{\alpha,\kappa;\phi} \frac{a(\alpha,\kappa;\phi)}{j!\ k!}\, \theta_{\phi}^{\alpha,\kappa}\, C_{\phi}^{\alpha,\kappa}(DL'VV'LD/2,\ D^2/2) \qquad (18)$$

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 $D = \text{diag}\{d_1, d_2, \ldots, d_n\}$, where $d_1^2 \le d_2^2 \le \ldots \le d_n^2$ are the non-zero 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 [5] and Chikuse and Davis [4] 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

$$h = \alpha\cos\theta + L\omega\sin\theta \qquad (19)$$

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 θ = 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 nxn orthogonal matrices).

Using (19), the matrix $L'VV'L$ that occurs in (18) becomes $L'VVL = L'(I-hh')L = I_n - \sin^2\theta\omega\omega'$. Now, consider a rotation of $\omega : \omega \to H\omega$, $H\epsilon O(n)$. Since the polynomials $C_\phi^{\alpha,\kappa}(A,B)$ are invariant under $A \to H'AH$, $B \to H'BH$, $H\epsilon O(n)$, we have

$$C_\phi^{\alpha,\kappa}(D^2 - \sin^2\theta D\omega\omega' D, D^2) \to C_\phi^{\alpha,\kappa}(D^2 - \sin^2\theta DH\omega\omega' H' D, D^2)$$
$$= C_\phi^{\alpha,\kappa}(H' D^2 H - \sin^2\theta H' DH\omega\omega' H' DH,\ H' D^2 H)$$

Hence, pdf(h) in (18) 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 non-zero 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 $H = \mathrm{diag}\{\pm1, \pm1, \ldots, \pm1\}$.

Thus, if the characteristic roots of $(\pi_2,\Pi_2)'(\pi_2,\Pi_2)$ are distinct the densities (18) are *not* constant on the surface θ = constant, but for fixed θ there are 2^n directions in which the density is the same, namely, those generated by replacing ω in (19) by $H\omega$, with $H = \mathrm{diag}\{\pm1, \ldots, \pm1\}$. Since pdf(h) = 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 (18).

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 non-zero characteristic roots of $(\pi_2,\Pi_2)'(\pi_2,\Pi_2)$ are all equal. In this case the density (18) *is* constant on the surface θ = 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 generalise; in general it does not.

Evidently, the properties of the LIML estimator (and the properly normalised 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

The OLS and TSLS estimators, as traditionally defined, embody a normalisation rule that is appropriate essentially only when the right-hand-side endogenous variables are weakly exogenous (i.e. $\beta = 0$). In general, this distorts the properties of these estimators. The LIML and properly normalised OLS/TSLS estimator are free of this distortion, whatever the number of endogenous variables in the equation.

This conclusion focuses attention on the densities (10) (for n = 1) and (18) (for n > 1), which differ for the different estimators only in respect of the numerical coefficients a(j,k) (or $a(\alpha, \kappa; \phi)$). To compare the different estimators one can use a measure of the information in pdf(h) about α, say

$$I(D;\ T-K_1,\ K_2) = -\int_{h'h=1} \ell n[\mathrm{pdf}(h)]\ \mathrm{pdf}(h)\ (h'dh),$$

or, alternatively, the risk function relative to the loss $\mathcal{L} = 1 - \cos^2\theta$,

$$R(D,\ T-K_1,\ K_2) = 1 - E(\cos^2\theta].$$

It is clear from (10) that, in the case n = 1 at least, these functions do not depend on the position of α, and that $I(d;\ T-K_1,\ K_2)$ is, in each case, an increasing function of d^2. I hope to report the results of further study of these functions in a later paper.

APPENDIX A

The coefficients in the LIML density

Equation (41) in [7] may be used to obtain the following expression for the density of the LIML estimator, r, for β in (4):

$$\mathrm{pdf}(r) = \Gamma((n+1)/2)[\pi(1+r'r)]^{-(n+1)/2}\mathrm{etr}\{-D^2/2\}$$

$$[C/\Gamma_n((n+1)/2)] \int_{f_1>0} \int_{R>0} f_1^{m(n+1)/2-1}(1+f_1)^{-(m+K_2)(n+1)/2}\mathrm{etr}(-R)$$

$${}_2F_2((m+K_2)/2,\ (n+3)/2;\ \ (m+n+2)/2,\ (n+1)/2;\ \ f_1R/(1+f_1)$$

$$\sum_{\alpha,\kappa;\phi} \frac{((m+K_2-n)/2)_\kappa}{j!k!\ (K_2/2)_\phi} \frac{C_\alpha(R)}{C_\alpha(I)}\, \theta_\phi^{\alpha,\kappa} C_\phi^{\alpha,\kappa}(DL'VV'LD/2,\ D^2/2)/(1+f_1)^{j+k}(dR)df_1 \quad (A.1)$$

where $m = T\text{-}K$, $V' = (I + rr')^{-1/2}(r,\ I_n)$,

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

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 (18) in the text are given by

$$a_1(\alpha,\kappa;\phi) = C\Big[((m+K_2-n)/2)_\kappa/(K_2/2)_\phi\Gamma_n((n+1)/2)\Big]$$

$$\int_{f_1>0} \int_{R>0} \mathrm{etr}(-R)[C_\alpha(R)/C_\alpha(I_n)]f_1^{m(n+1)/2-1}(1+f_1)^{-(f+(m+K_2)(n+1)/2)}$$

$$ {}_2F_2((m+K_2)/2,\ (n+3)/2;\ \ (m+n+2)/2,\ (n+1)/2;\ \ f_1R/(1+f_1))df_1(dR) $$

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

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

$$ \times \sum_{\rho\in\alpha.\lambda}((n+1)/2)_\rho(\theta_\rho^{\alpha,\lambda})^2C_\rho(I)/C_\alpha(I)\ . \qquad (A.3) $$

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

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

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

$$ \frac{(1)_j((m+K_2-1)/2)_k(K_2)_f}{(K_2/2)_f(m+K_2)_f}\ {}_4F_3((m+K_2)/2,2,m,\ j+1;(m+3)/2,1,f+m+K_2;\ 1) \qquad (A.4) $$

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

$$(\pi_2,\Pi_2) \quad = \quad ADL' \tag{B.1}$$

where A $(K_2 \times n)$ satisfies $A'A = I_n$, D is a diagonal matrix $D = \mathrm{diag}\{d_1,\ d_2,\ \ldots,\ d_n\}$, where $0 < d_1^2 \le d_2^2 \le \ldots \le d_n^2$ are the n non-zero 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 normalised case we can take

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

Using (B.1) we have (cf. [9], Section 10.3)

$$\begin{aligned} \mathrm{pdf}(S) \quad = \quad & C_2\ \mathrm{etr}\{-D^2/2\}\ \mathrm{etr}\{-S/2\}\ |S|^{(\nu-n-2)/2} \\ & {}_0F_1(\nu/2,\ DL'SLD/4) \end{aligned} \tag{B.3}$$

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 [9], Theorem 3.2.17 for technical details and notation). We have

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

where $F = \mathrm{diag}\{f_1,\ f_2,\ \ldots,\ f_n,\ f\}$, $f_1 > f_2 > \ldots > f_n > f > 0$, $H \in O(n+1)$, and $(H'dH)$ denotes the unnormalised 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 $F_1 = \mathrm{diag}\{f_1, \ldots, f_n\}$. 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. [9], 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:

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

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,

$$\mathrm{pdf}(h,F) = C_2^* \mathrm{etr}\{-D^2/2\}\mathrm{etr}\{-F/2\}|F|^{(\nu-n-2)/2}$$

$$a_n \int_{\mathrm{Re}(W)>0} \mathrm{etr}(W)|W|^{-\nu/2}\mathrm{etr}\{fD^2W^{-1}/4\}$$

$${}_0F_0^{(n)}((F_1 - fI_n)/2,\ DL'VV'LDW^{-1}/2)\ (dW)\ , \qquad \text{(B.5)}$$

with $C_2^* = \pi^{n^2/2}C_2/2\Gamma_n(n/2)$ (for the other notation see [9], Chapter 7).

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

$$\text{pdf}(h) = [\Gamma((n+1)/2)/2\pi^{(n+1)/2}]\text{etr}\{-D^2/2\}$$

$$\sum_{\alpha,\kappa;\phi} \frac{a_2(\alpha,\kappa;\phi)}{j!\ k!} \theta_{\phi}^{\alpha,\kappa} C_{\phi}^{\alpha,\kappa} (DL'VV'LD/2,\ D^2/2)\ , \qquad \text{(B.6)}$$

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

$$a_2(\alpha,\kappa;\phi) = [2\pi^{(n+1)/2}C_2^*/2^{j+k}\Gamma((n+1)/2)(\nu/2)_{\phi}]$$

$$\int_{f_1>\dots>f_n>f} \int \text{etr}\{-F/2\}|F|^{(\nu-n-2)/2} f^k [C_{\alpha}(F_1-fI_n)/C_{\alpha}(I_n)]$$

$$\prod_{i<j}^{n} (f_i-f_j) \prod_{i=1}^{n} (f_i-f) \prod_{i=1}^{n} df_i\ df \qquad \text{(B.7)}$$

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

$$a_2(\alpha,\kappa;\phi) = [2\pi^{(n+1)/2}C_2^*/2^{j+k}\Gamma((n+1)/2)(\nu/2)_{\phi}]$$

$$\int_{\bar{f}_1>\dots>\bar{f}_n>0} \int_{f>0} \int \text{etr}\{-f(I+\bar{F}_1)/2\} \exp\{-f/2\} f^{j+k+(n+1)\nu/2-1}$$

$$|\bar{F}_1|\ |I+\bar{F}_1|^{(\nu-n-2)/2}[C_{\alpha}(\bar{F}_1)/C_{\alpha}(I_n)] \prod_{i<j}^{n} (\bar{f}_i-\bar{f}_j) \prod_{i=1}^{n} d\bar{f}_i\ df$$

$$= [\pi^{(n+1)/2}C_2/2^{j+k}\Gamma((n+1)/2)(\nu/2)_{\phi}]$$

$$\int_{f>0} \int_{R>0} \text{etr}\{-f(I+R)/2\} \exp\{-f/2\} f^{j+k+(n+1)\nu/2-1}$$

$$|R|\ |I+R|^{(\nu-n-2)/2}[C_{\alpha}(R)/C_{\alpha}(I_n)](dR)\ df \qquad \text{(B.8)}$$

(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 $n = 1$,

$$a_2(j,k) = [\Gamma((\nu+1)/2)/2\Gamma((\nu+3)/2)]$$

$$[(2)_j((\nu-1)/2)_k(\nu)_{j+k}/2^{j+k}(\nu/2)_{j+k}((\nu+3)/2)_{j+k}]$$

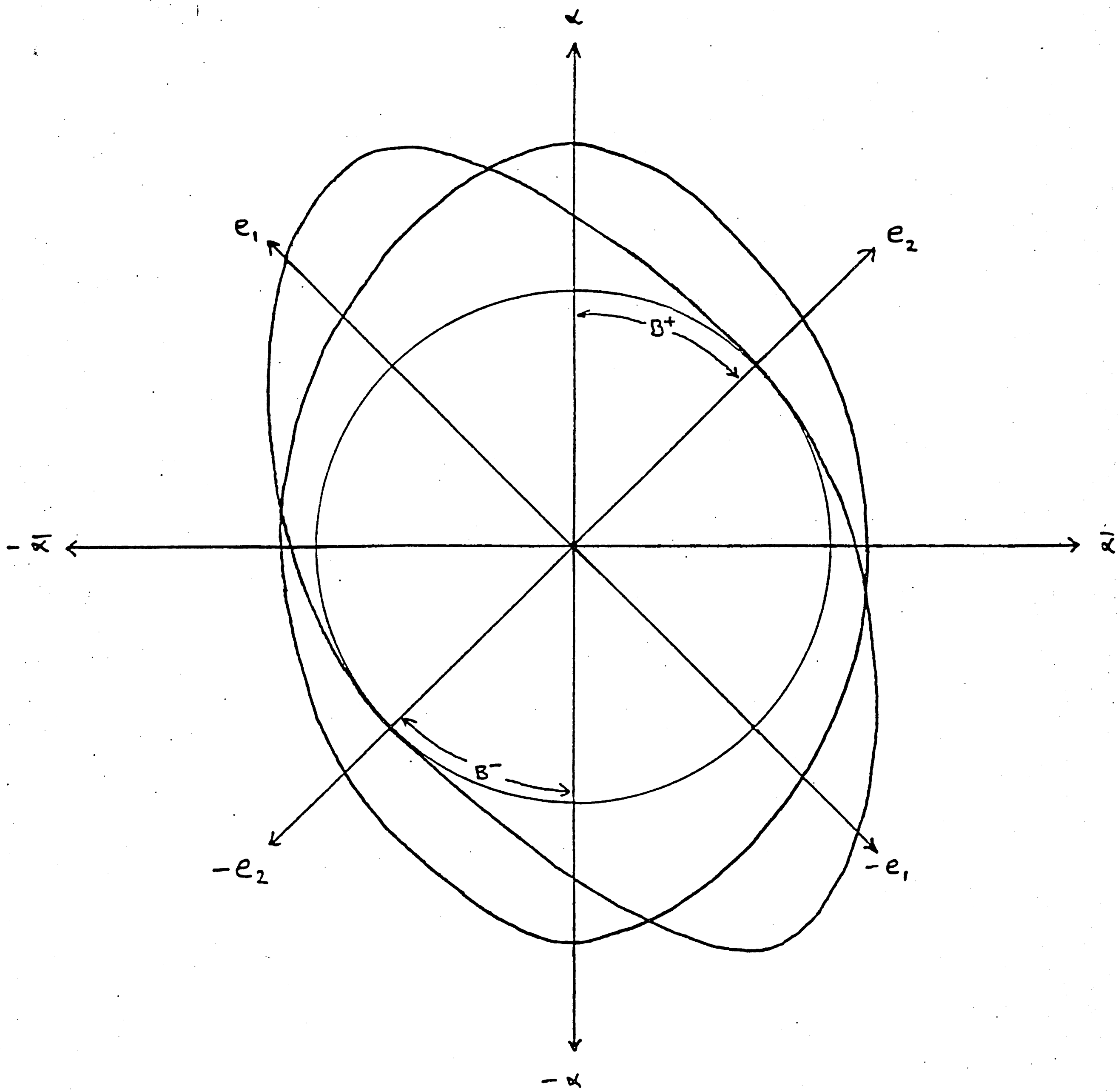
$$_2F_1(j+2,\ j+k+\nu,\ j+k+(\nu+3)/2,\ 1/2)\ . \qquad \text{(B.9)}$$

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.

FIGURE 1

Densities of OLS/TSLS and LIML direction estimators: n=1



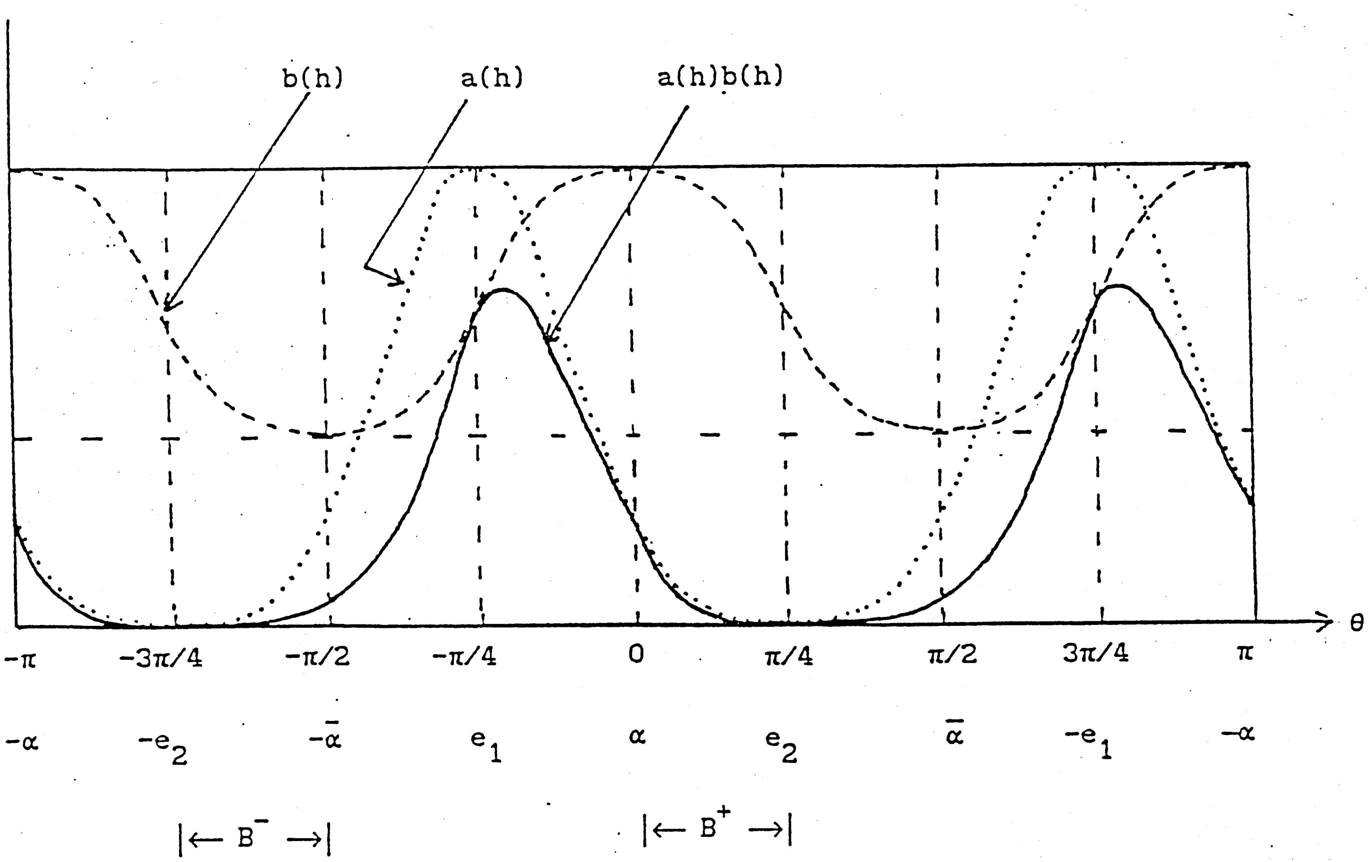
Note: The position of α corresponds to $\beta = -1$.

The sets of B^+ (see equation (15) in the text) and B^-, its mirror image, yield coefficient estimators for the normalised equation satisying $r < \beta$.

FIGURE 2

a(h), b(h), and the product a(h)b(h)

(β = -1)



Footnotes

2. This question was suggested by a referee, who also points out that the analogue of the TSLS estimator can be viewed as a generalised method of moments estimator, normalised to have unit length.

3. For the OLS/TSLS estimators (16) follows from the decomposition (13), the symmetry of b(h) about α, and the fact that $0 \leq a(h) \leq 1$.

References

[1] Anderson, T.W., N. Kunimoto, and K. Morimune "Comparing single equation estimators in a simultaneous equation system", Econometric Theory, 2 (1986), 1-32.

[2] Anderson, T.W., N. Kunimoto, and T. Sawa "Evaluation of the distribution function of the limited information maximum likelihood estimator", Econometrica, 50 (1982), 1009-1027.

[3] Anderson, T.W. and T. Sawa "Evaluation of the distribution function of the two-stage least squares estimate", Econometrica, 25 (1973), 77-83.

[4] Chikuse, Y. and A.W. Davis "A survey on the invariant polynomials with matrix arguments in relation to econometric distribution theory", Econometric Theory, 2 (1986), 232-248.

[5] Davis, A.W. "Invariant polynomials with two matrix arguments, extending the zonal polynomials: Applications to multivariate distribution theory", Annals of the Institute of Statistical Mathematics, A, 31 (1979), 465-485.

[6] Hillier, G.H. "On the joint and marginal densities of instrumental variable estimators in a general structural equation", Econometric Theory, 1 (1985), 53-72.

[7] Hillier, G.H. "Joint distribution theory for some statistics based on LIML and TSLS", Cowles Foundation Discussion Paper No. 840, June 1987.

[8] Mariano, R.S. and T. Sawa "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, (1972), 159-163.

[9] Muirhead, R.J. Aspects of Multivariate Statistical Theory, New York: Wiley, 1982.

[10] Phillips, P.C.B. "The exact finite sample density of instrumental variable estimators in an equation with n+1 endogenous variables", Econometrica, 48 (1980), 861-878.

[11] ———— "Exact small sample theory in the simultaneous equations model" In M.D. Intriligator and Z. Griliches, eds., Handbook of Econometrics, Amsterdam: North Holland, 1983.

[12] ———— "The exact distribution of LIML: I", International Economic Review, 25 (1984), 249-261.

[13] ———— "The exact distribution of LIML: II", International Economic Review, 26 (1985), 21-36.

[14] ———— "Partially identified econometric models", Cowles Foundation Discussion Paper No. 845, July, 1987.

[15] Richardson, D.H. "The exact distribution of a structural coefficient estimator", Journal of the American Statistical Association, 63 (1961), 1214-1226.

[16] Sawa, T. "The exact sampling distribution of ordinary least squares and two-stage least squares estimators", Journal of the American Statistical Association, 64 (1968), 923-927.

MONASH UNIVERSITY

DEPARTMENT OF ECONOMETRICS

WORKING PAPERS

1988

1/88 John Preston and Esme Preston, "Two Papers on Simple Word Processors for the PC".

2/88 Esme Preston, "Multilingual and Mathematical Text Processing".

3/88 Esme Preston and John Preston, "The Accountant's PC: Getting Started".

4/88 Maxwell L. King, "Towards a Theory of Point Optimal Testing".

5/88 Ralph D. Snyder, "Statistical Foundations of Exponential Smoothing".

6/88 Grant H. Hillier, "On the Interpretation of Exact Results for Structural Equation Estimators".

7/88 Kuldeep Kumar, "Some Recent Developments in Non-Linear Time Series Modelling.

8/88 Maxwell L. King, "Testing for Fourth-Order Autocorrelation in Regression Disturbances When First-Order Autocorrelation is Present".

9/88 Ralph D. Snyder, "Kalman Filtering with Partially Diffuse Initial Conditions".

10/88 Asraul Hoque, "Indirect Rational Expectations and Estimation in a Single Equation Macro Model".

11/88 Asraul Hoque, "Efficiency of OLS Relative to C-O for Trended x and Positive Autocorrelation Coefficient".

12/88 Russel J. Cooper & Keith R. McLaren, "Regular Alternatives to the Almost Ideal Demand System".

13/88 Maxwell L. King, "The Power of Student's t Test: Can a Non-Similar Test Do Better?".

14/88 John Hamilton, "On-Line Management of Time Series Databases: Database Retrieval Program DBR".