FINITE-SAMPLE OPTIMALITY OF TESTS IN A STRUCTURAL EQUATION

Randolph G. K. Tan, Nanyang Technological University, Singapore

Paper to be presented at the Eighth World Congress of the Econometric Society, to be held 11 to 16 August 2000 in Seattle, USA.

1. Introduction

In a linear regression model, the t test is uniformly most powerful and is equivalent to an exact likelihood ratio test. Unfortunately, these optimality results are not assured when there is endogeneity of some of the regressors. One approach which is rather tentative is to offer a large-sample justification for conducting inference in a fashion analogous to linear regression. This is inadequate because it ignores the fact that in the transition from a linear regression to a structural equation model, we also have to contend with the different roles that exogenous and endogenous regressors play. Indeed, the present state of econometric practice does not distinguish between these in the application of t-tests, whereas the logic of identifiability seems to suggests otherwise.

Except for the Anderson-Rubin F test whose exact power function was derived by Revankar and Mallela (1972), there are few known finite sample optimality results for coefficient tests in a structural equation. The few existing results relate to the testing of identifiability of the equation, which corresponds to tests of the coefficients of its exogenous variables. Among optimality results and small sample optimality results in this area are the maximal invariant for testing identifiability (see Muirhead, Theorem 10.2.1) and the finding that the likelihood ratio (LR) test statistic for identifiability is approximately ancillary (Hosoya et al, 1989). The two are connected in that a maximal invariant is also ancillary if the action of transformations induced by the invariance group on the parameter space is transitive (see Lehmann, 1986, pp.542-548 for full details). Such a condition is not generally satisfied by the curved exponential model that we are talking about. The extent to which it is not satisfied is in some sense the degree of curvature - a point that may be deduced from Hosoya et al's (1989) analysis. Thus, it is only in the case of exact identifiability that this maximal invariant is ancillary to the problem.

In conventional use, any approach by the F ratio must have been compelled by connections to linear regression. Unfortunately, when not all of the regressors (especially those that are endogenous) vanish under the null, the fundamentally nonlinear character of the hypothesis means that this lacks justification even from an asymptotic standpoint. In an environment clouded by uncertainty over the finite-sample approach and pressed by empirical demands, the asymptotic results of Kadane (1984) and Morimune and Tsukuda (1984) were very important in establishing the reliability of the LR test from limited information maximum likelihood (LIML) estimation.

To address continuing uncertainty over the properties of the t ratio, Morimune (1989) carries out both a Monte Carlo study of the size properties and asymptotic analysis of the power properties for variants of the t ratio approach constructed from ordinary least squares (OLS), two stage least squares (TSLS) and LIML as well as the LIML-based LR test. What is interesting to note is that the OLS-based t ratio leads to the most extreme divergence from the nominal size, and the problem was worse when testing the coefficients of the endogenous variables (refer to Morimune, 1989, tables II and III to see which column depicts the widest departure from the nominal size). In terms of adherence to size, the best performers in order appear to be the t ratio based on LIML, the t ratio based on TSLS and the LR test based on

LIML. In *all* these three cases, there is clearly a worsening in size-adherence when it is the endogenous variable whose coefficient is being tested. In this paper, we will assert that the information contained in the maximal invariants for the testing problem is most seriously distorted in the case of OLS, and preserved intact by LIML, leading quite naturally to the results seen in these simulation results.

If a problem is invariant under a group of transformations, then appealing to the principle of invariance suggests that the search for an optimal test need only be confined to the class of tests which share the same invariant properties. Since a necessary and sufficient condition for a test statistic to be invariant is that it depends on the sample space through the maximal invariant, the class of all invariant test procedures can be characterized if the maximal invariant is found. By maintaining a constant value on orbits and assigning different values to each orbit in the sample space, the maximal invariant serves as a functional representation of the sample space that obscures information irrelevant to the inference at hand. The practical interpretation of this concept for inference is that optimality criteria for maximizing power can be applied to the distribution of the maximal invariant. Given the invariance group of a testing problem, a maximal invariant is specific to that group but need not be unique.

The invariance properties of a problem are completely described by the action of a transformation group which leaves the problem invariant. The group and its corresponding action are referred to as the invariance group, and is all the information that is necessary to defining a maximal invariant.

The following notational conventions are used. Square brackets will be reserved exclusively for enclosing the arguments of a function, as in $f[\bullet]$ or $\mathbf{Q}[\bullet]$. This is a convenient means of tracing the relations between elements (scalar, matrix or group) as they are transformed using a (broadly defined) functional dependency on certain arguments. Other times, we omit the arguments in the interests of notational economy. Bold letters usually denote matrix statistics that are functions of the observations, while script lettering denotes arbitrary matrices which are usually, but not confined to, elements of the transformation groups. Further, in any font: $\mathbf{U}_\Omega$ would denote an upper-block triangular matrix from decomposing a symmetric matrix Ω; $\boldsymbol{H}$ would denote an element of an orthogonal group; and $\boldsymbol{T}_{\mathbf{W}}$ would denote an upper-triangular matrix from the Cholesky factorization of $\mathbf{W}$.

2. Model and assumptions

Modeling traditionally begins by postulating a behavioural equation of interest which links endogenous and exogenous variables. Such a relationship is written in unnormalized form as

(1) $\mathbf{Y}\beta=\mathbf{Z}_1\gamma_1+\mathbf{Z}_2\gamma_2+\mathbf{e}$

This is a structural equation because it is based on behavioural mechanisms in which there is strong theoretical foundation. Underlying (1) is a process which is thought to generate the data, represented as

(2) $\mathbf{Y}=\mathbf{Z}\Pi+\mathbf{V}$

The sample comprises T time-series or cross-sectional observations on endogenous and exogenous variables. Each column of $\mathbf{Y}=(\mathbf{Y}_1 \;\; \mathbf{Y}_2)$ represents a $T\times1$ vector of observations on an endogenous variable, of which there are (n+1) in the model. Each column of $\mathbf{Z}=(\mathbf{Z}_1 \;\; \mathbf{Z}_2 \;\; \mathbf{Z}_3)$ is a $T\times1$ vector of observations on an exogenous variable, of which there are K in total. Of these, (K_1+K_2) variables are included in the structural equation with the first K_1 contained in $\mathbf{Z}_1$ and the next K_2 in $\mathbf{Z}_2$. The remaining K_3 (=K-K_1-K_2) columns in $\mathbf{Z}_3$ represent what are usually referred to as the excluded exogenous variables, often used as instrumental variables in estimation. The remaining terms in the model are random disturbances, in the T 1 vector **e**, and T (n+1) matrix **V**.

Requiring that the behavioural specification in (1) be consistent with (2) implies the parametric restrictions

(3) $\Pi_1 \text{b} = \text{g}_1, \qquad \Pi_2 \text{b} = \text{g}_2,$

(4) $\Pi_3 \text{b} = 0$

and **e**=**V**β. Of these, (4) is crucial and implies that a necessary and sufficient condition for a structural form unique up to normalization to exist is the rank condition

(5) $rank[\Pi_3] = n$.

A test of identifiability is one of $H_0^{\Pi_3}: rank[\Pi_3] = n$ versus $H_1^{\Pi_3}: rank[\Pi_3] = n+1$. It will be convenient to refer to various submatrices in Π by partitioning it as

(6) $$\Pi = \begin{pmatrix} \Pi_1 \\ \Pi_2 \\ \Pi_3 \end{pmatrix} = \begin{pmatrix} \Pi_{11} & \Pi_{12} \\ \Pi_{21} & \Pi_{22} \\ \Pi_{31} & \Pi_{32} \end{pmatrix} \begin{matrix} \kappa_1 \\ \kappa_2 \\ \kappa_3 \end{matrix}$$

$$\qquad n+1 \qquad\qquad n_1+1 \quad n_2$$

This paper focuses on those hypotheses which specify that some of the variables in (1) may have zero coefficients. These tests of variable exclusions are often described as tests of significance when the judgment is made on the basis of how large an estimate is in absolute value compared to a measure of distributional spread such as the estimated standard deviation. In all of the remaining cases of testing, the effect of the variable exclusions is to require a modified compatibility condition in the spirit of (5) to be satisfied.

For the purposes of parametric inference, we need the assumption that rows of $\mathbf{V}$ be a independently distributed as multivariate normal vectors, mean $\mathbf{0}$ and covariance matrix Ω. Given that our main concern is with the slope coefficients in (1) and (2), we can always transform the model so that Ω=$\mathbf{I}_{n+1}$ without any loss of generality. To complete the specification of the model in canonical form, we also assume that the columns of $\mathbf{Z}$ are orthonormal vectors. If the model is not yet in canonical form, it can be converted to one by post-multiplying Y by and postmultiplying Z by to give (see Phillips (1983) details),

The representation makes clear that the ranks of various submatrices of Π that are involved in testing should be preserved, given that invariance is our main concern in this paper. This can be achieved with non-singular blocks in the diagonals of

$$\mathbf{U}_{\Omega}^{-1} = \begin{pmatrix} \mathbf{\Omega}_{11}^{-1/2} & -\mathbf{\Omega}_{11}^{-1/2}\mathbf{\Omega}_{12}\mathbf{\Omega}_{22\cdot 1}^{-1/2} \\ \mathbf{0} & \mathbf{\Omega}_{22\cdot 1}^{-1/2} \end{pmatrix} \begin{matrix} n_1+1 \\ n_2 \end{matrix}$$

$$\begin{matrix} n_1+1 & \qquad n_2 \end{matrix}$$

and

$$\mathbf{U}_{\mathbf{Z}'\mathbf{Z}}^{-1} = \begin{pmatrix} (\mathbf{Z}_1'\mathbf{Z}_1)^{-1/2} & -(\mathbf{Z}_1'\mathbf{Z}_1)^{-1}\mathbf{Z}_1'\mathbf{Z}_2(\mathbf{Z}_2'\mathbf{P}_{\mathbf{Z}_1}\mathbf{Z}_2)^{-1/2} & (\mathbf{Z}_1'\mathbf{Z}_1)^{1/2} \\ \mathbf{0} & (\mathbf{Z}_2'\mathbf{P}_{\mathbf{Z}_1}\mathbf{Z}_2)^{-1/2} & -(\mathbf{Z}_2'\mathbf{P}_{\mathbf{Z}_1}\mathbf{Z}_2)^{-1}\mathbf{Z}_2'\mathbf{P}_{\mathbf{Z}_1}\mathbf{Z}_3(\mathbf{Z}_2'\mathbf{P}_{\mathbf{Z}_1}\mathbf{Z}_2)^{-1/2} \\ \mathbf{0} & \mathbf{0} & (\mathbf{Z}_3'\mathbf{P}_{\mathbf{Z}_1,\mathbf{Z}_2}\mathbf{Z}_3)^{-1/2} \end{pmatrix}$$

Since the process generating the data takes precedence over any relationship between $\mathbf{Y}$ and $\mathbf{Z}_1$ in (1), (2) may be regarded as the working hypothesis providing the basis on which investigations about (1) are to be carried out.

It is well known that the maximum likelihood estimates $\mathbf{X}=\mathbf{Z}'\mathbf{Y}$ and $\mathbf{W}=\mathbf{Y}'(\mathbf{I}-\mathbf{Z}\mathbf{Z}')\mathbf{Y}$ are independent and constitute sufficient statistics for inference.

Note that with the model in canonical form, $\mathbf{X}$ is matrix normal with the mean $E[\mathbf{X}] = \mathbf{Z}'\mathbf{Z}\Pi = \Pi$ and covariance matrix $\mathbf{I}_{K_2(n+1)}$. The partitioned form of these are

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \mathbf{X}_3 \end{pmatrix} = \begin{pmatrix} \mathbf{X}_{11} & \mathbf{X}_{12} \\ \mathbf{X}_{21} & \mathbf{X}_{22} \\ \mathbf{X}_{31} & \mathbf{X}_{33} \end{pmatrix} \begin{matrix} K_1 \\ K_2 \\ K_3 \end{matrix}$$
$$\qquad\qquad n_1+1 \quad n_2$$

and

$$W = \begin{pmatrix} \mathbf{W}_{11} & \mathbf{W}_{12} \\ \mathbf{W}_{21} & \mathbf{W}_{22} \end{pmatrix} \begin{matrix} n_1+1 \\ n_2 \end{matrix} .$$
$$n_1+1 \quad n_2$$

First, we define certain functions of the sufficient statistics which play important roles in the sequel. Let $\mathbf{S}_{(0)} = \mathbf{X}_2'\mathbf{X}_2$, $\mathbf{S} = \mathbf{X}_3'\mathbf{X}_3$, $\mathbf{S}_{11} = \mathbf{X}_{31}'\mathbf{X}_{31}$ and denote $\mathbf{W}_{22\cdot 1} = \mathbf{W}_{22} - \mathbf{W}_{21}(\mathbf{W}_{11}'\mathbf{W}_{11})^{-1}\mathbf{W}_{12}$. Based on our assumptions, $\mathbf{S}_{(0)}$, $\mathbf{S}$ and $\mathbf{W}$ are independent. A block-triangular decomposition $\mathbf{W} = \mathbf{U}_\mathbf{W}'\mathbf{U}_\mathbf{W}$ is specified by having

(7) $$\mathbf{U}_\mathbf{W} = \begin{pmatrix} \mathbf{W}_{11}^{1/2} & \mathbf{W}_{11}^{-1/2}\mathbf{W}_{12} \\ \mathbf{0} & \mathbf{W}_{22\cdot 1}^{1/2} \end{pmatrix}$$

Then define a statistic

(8) $$\mathbf{R}[\mathbf{X},\mathbf{W}] = \mathbf{U}_\mathbf{W}'^{-1}\mathbf{S}\mathbf{U}_\mathbf{W}^{-1} .$$

Though not explicitly used in Hillier (1987), its construction is clearly suggested. The matrix statistic $\mathbf{R}$ is itself not a maximal invariant for any of the testing problems, but its intrinsic appeal is already evident from the compact way in which it can be used to define the likelihood ratio test for $H_0^{\beta_2}$. Essentially, R is a symmetrized version of $\mathbf{SW}^{-1}$ and either form can be represented in the likelihood function. It turns out to have a fundamental role in the maximal invariants that we formulate in each of the three cases.

3. Invariance of the Testing Problems

Let $b = (b_1', b_2')'$ be a partition into column vectors of (n_1+1) and n_2 components. Then, apart from identifiability, the remaining three types of tests are of

(9) $H_0^{\gamma_2} : \gamma_2 = 0$ against $H_1^{\gamma_2} : \gamma_2 \neq 0$,

(10) $H_0^{\beta_2} : \beta_2 = 0$ against $H_1^{\beta_2} : \beta_2 \neq 0$, and

(11) $H_0^{\beta_2,\gamma_2} : \beta_2 = 0, \gamma_2 = 0$ against $H_1^{\beta_2,\gamma_2} : \beta_2 \neq 0, \gamma_2 \neq 0$.

These can also be written as rank conditions, with all three taking the general form $H_0 : rank[\Pi_*] = n_*$ against $H_1 : rank[\Pi_*] = n_* + 1$, where Π_* has dimensions $K_* \times (n_* + 1)$ and comprises submatrices from the partition of Π (see (6)). Implicit in (9)-(11) is the idea that the tests are to be carried out for a model that is already identified (even possibly over-identified). This is an explicit requirement that (5) be the maintained hypothesis throughout. The preferred tests would therefore compare a new identifiability condition with the one held under the maintained hypothesis. If no comparison to (5) were made, then the new rank restrictions would merely be variants of the identifiability hypothesis with a rearrangement of the variables that are excluded from the structural equation.

The likelihood ratio tests from limited information maximum likelihood (LIML) offers a well-known example where comparison to the maintained hypothesis is the key. Here, the test statistics for (9)-(11) turn out to have the general form $\{(1+\lambda_1^*)/(1+\lambda_1^{LIML})\}^{-T/2}$, where λ_1^{LIML} is the largest latent root ... (see Hosoya *et al* (1989)). λ_1^* is just its counterpart from the model modified to fit the variable exclusions postulated by the null in question.

The action of a group of transformations $\boldsymbol{T}$ on the sample space $(\mathbf{Z},\mathbf{Y})$ can be described completely in terms of maps of the sufficient statistics. The resulting maps of the moments of the sufficient statistics that are implied has the effect of inducing transformations in parameter space.

Definition 1. A testing problem is said to be invariant with respect to a group $\boldsymbol{T}$ if the hypotheses being tested remain unchanged under the action of $\boldsymbol{T}$. We call $\boldsymbol{T}$ the invariance group for the problem.

It will be convenient to use $\boldsymbol{T}_{\mathrm{b}_2}$ to represent the transformation group which leaves the problem of testing the presence of β_2 in (1) unchanged. The natures of $\boldsymbol{T}_{\mathrm{b}_2}$, (and using the same notation) $\boldsymbol{T}_{\mathrm{g}_2}$ and $\boldsymbol{T}_{\mathrm{g}_2,\mathrm{b}_2}$ have been extensively discussed in Hillier (1987) and hinge on the fact that in addition to (3), another compatibility condition implied by coefficient restrictions on (1) must be preserved. Thus, while (2) looks outwardly like a general linear model (2), there is a change in its behaviour due to the possibility that a linear combination, $\mathbf{Y}\beta$, of the endogenous variables may be represented by a linear regression on $\mathbf{Z}_1$, a selection from the full regressor set. This resulting model incorporating (1) and (2) has been characterized as a curved exponential model (Hosoya et al (1989)).

Definition 2. A function $h[\mathbf{X},\mathbf{W}]$ on $\boldsymbol{X}$ is said to be invariant under the group $\boldsymbol{T}_{\mathrm{b}_2}$ if

(12) $h[\boldsymbol{t}_{\mathrm{b}_2}(\mathbf{X},\mathbf{W})] = h[\mathbf{X},\mathbf{W}]$

for all $(\mathbf{X},\mathbf{W})\in\boldsymbol{X}$ and $\boldsymbol{t}_{\mathrm{b}_2} \in \boldsymbol{T}_{\mathrm{b}_2}$. If $h[\mathbf{X},\mathbf{W}]$ satisfies the additional property where $h[\mathbf{X}^\circ,\mathbf{W}^\circ] = h[\mathbf{X},\mathbf{W}]$ implies that $(\mathbf{X}^\circ,\mathbf{W}^\circ) = \boldsymbol{t}_{\mathrm{b}_2}(\mathbf{X},\mathbf{W})$, then it is said to be maximal invariant under $\boldsymbol{T}_{\mathrm{b}_2}$.

Thus, proving maximal invariance just involves showing that such a t_{β_2} exists and can be found from information on $(\mathbf{X},\mathbf{W})$ and $(\mathbf{X}°,\mathbf{W}°)$.

The aim is to prove that a function denoted by $h_{b_2}[\mathbf{X},\mathbf{W}]$ satisfies the conditions of being a maximal invariant for the testing problems (9)-(11). In the following, we make use of the fact that elements of the group of symmetric positive definite matrices can be related (though not always uniquely) to elements of the triangular, block triangular or general linear group.

Since the actions of each of the invariance groups $\boldsymbol{T}_{b_2}$, $\boldsymbol{T}_{g_2}$ and $\boldsymbol{T}_{g_2,b_2}$ is transitive on that part of the sample space relating to $\mathbf{X}_1$ (theorems 4, 5 and 6 of Hillier(1987)), an invariance approach to the problem of testing for (endogenous *and* exogenous) variable exclusions reduces the relevant sample space *further*, from $(\mathbf{X},\mathbf{W})$ to $(\mathbf{X}_2,\mathbf{X}_3,\mathbf{W})$. We can go even further. Aside from $\mathbf{S}$ and $\mathbf{W}$ which have already been defined, the only other function of the sufficient statistics which plays a role later is $\mathbf{S}_{(1)}=\mathbf{X}_2'\mathbf{X}_2$. This constitutes a further reduction in dimensionality, from $((K_2+K_3)(n+1) + (n+1)(n+2)/2)$ to *at least*

(13a) $(K_2(K_2+1)/2 + (n+1)(n+2))$ when $K_2<n+1$, or

(13b) $3(n+1)(n+2)/2$ when $K_2\geq n+1$.

The actual reductions which an invariance allows in each case will be obtained from the specification of the maximal invariants.

4. Maximal Invariants

4.1. Identifiability and further exclusion of exogenous variables

Given what is said above about the requirement for compatibility between the structural equation and the process which has given rise to the data, the requisite condition is one of identifiability in the sense of being essential to the existence of the behavioural equation. The problem of testing identifiability is invariant under parameter transformations induced by the action of the group of transformations

$$\boldsymbol{T}_{\Pi_3} = \{\boldsymbol{t}_{\Pi_3} = (\boldsymbol{H}\ ,\boldsymbol{Q}):\quad \boldsymbol{H}\ \in \mathbf{O}[K_3], \boldsymbol{Q} \in \mathbf{GL}[n+1]\}$$

on the sufficient statistic defined by

$$\boldsymbol{t}_{\Pi_3} : \mathbf{X}_3 \to \boldsymbol{H}\,\mathbf{X}_3\boldsymbol{Q},\quad \mathbf{W} \to \boldsymbol{Q}'\mathbf{W}\boldsymbol{Q}\ .$$

Theorem 1 (*Identifiability*)

For the problem of testing $H_0^{\Pi_3}$ vs $H_1^{\Pi_3}$, the latent roots of $\mathbf{R}$ constitute a maximal invariant function of the sufficient statistics.

PROOF.

We denote the latent roots of $\mathbf{R}$ by $h_{\Pi_3}[\mathbf{X}_3, \mathbf{W}]$. The latent roots of $\mathbf{R} = \mathbf{U}_S'^{-1}\mathbf{W}\mathbf{U}_S^{-1}$ are equal to those of $\mathbf{W}\mathbf{U}_S^{-1}\mathbf{U}_S'^{-1} = \mathbf{W}\mathbf{S}^{-1}$. Under the action of the group, we get the map $\mathbf{W}\mathbf{S}^{-1} \to \boldsymbol{Q}'\mathbf{W}\mathbf{S}^{-1}\boldsymbol{Q}'^{-1}$, thereby proving invariance. To prove the stronger result of maximal invariance, let $h_{\Pi_3}[\mathbf{X}_3, \mathbf{W}] = h_{\Pi_3}[\overline{\mathbf{X}}_3, \overline{\mathbf{W}}]$. Here and in the sequel, the bar notation denotes the statistics from another sample of observations which yields the same value of a maximal invariant. Then, there exists $\mathrm{H}_1 \in \mathbf{O}[n+1]$ such that

$$\mathrm{H}_1'\mathbf{U}_S'^{-1}\mathbf{W}\mathbf{U}_S^{-1}\mathrm{H}_1 = \overline{\mathbf{U}}_S'^{-1}\overline{\mathbf{W}}\overline{\mathbf{U}}_S^{-1}$$

(Q.E.D.)

It is interesting to note that Constantine (1963) first solved this problem and derived the exact form of the density function of the maximal invariant. The treatment varies only in the simplifying transformation which is applied to the sufficient statistic. Constantine's (1963) choice maps Y to two subspaces, one spanned by the columns of X and the other its orthogonal complement. The latter gives rise to the centrally distributed term.

The exact probability density function of $h_{\Pi_3}[\mathbf{X}_3, \mathbf{W}]$ is easily derived from an existing result in Muirhead (1982, Theorem 10.4.2) and takes the form

$$(14) \quad \begin{gathered} etr[-\tfrac{1}{2}\mathrm{M}]\,{}_1F_1^{(n+1)}[\tfrac{1}{2}(\nu + K_3); \tfrac{1}{2}K_3; \tfrac{1}{2}\mathrm{M}, \Lambda(\mathbf{I}+\Lambda)^{-1}] \\ \frac{\Gamma_{n+1}[(\nu+K_3)/2]}{\Gamma_{n+1}[K_3/2]\Gamma_{n+1}[\nu/2]} \frac{\pi^{(n+1)^2/2}}{\Gamma_{n+1}[(n+1)/2]} \prod_{i=1}^{n+1} \frac{\lambda_i^{(K_3-n-2)/2}}{(1+\lambda_i)^{(\nu+K_3)/2}} \prod_{i<j}^{n+1} (\lambda_i - \lambda_j) \end{gathered}$$

As the source emphasizes, this depends only on the latent roots of M, making it convenient to specify one single parameter targeted by the test. Since M is clearly non-negative definite, the null hypothesis of identifiability says that the smallest of these roots should be zero. Under the alternative hypothesis, M would have full rank and therefore only positive latent roots. Thus, the power function is to be analyzed based on its dependence on the smallest root of M. The remaining n larger latent roots are 'nuisance' parameters. It will be difficult to isolate the key parameter being tested, in view of their entanglement in a zonal polynomial term.

Given in this form, the maintained hypothesis can be easily incorporated into the testing problems. The aim should be to ensure that the maintained hypothesis is satisfied even under the alternative hypotheses considered below.

Testing the exclusion of exogenous variables is the most important because of its relationship to identifiability. Since exogenous variables excluded from an equation determine its identifiability. accepting the null $H_0^{\gamma_2}$ would render an equation that is already identified (by virtue the maintained hypothesis) overidentified. Should that occur, the compatibility condition (3) would become obsolete in a sense and should be updated to reflect the stronger claim to identifiability.

In the case of testing $H_0^{\gamma_2}$, the problem remains unchanged under maps of the form $\mathbf{Y} \to \boldsymbol{H}_0 \mathbf{Y} \boldsymbol{Q}$, $\mathbf{Z}_1 \to \boldsymbol{H}_0 \mathbf{Z}_1 \boldsymbol{H}_1'$. The maintained hypothesis also remains unchanged under the additional maps $\mathbf{Z}_2 \to \boldsymbol{H}_0 \mathbf{Z}_2 \boldsymbol{H}_2'$. These imply an invariance group specified by

$$\boldsymbol{T}_{\gamma_2} = \{\boldsymbol{t}_{\gamma_2} = (\boldsymbol{H}_1, \boldsymbol{H}_2, \boldsymbol{Q}) : \boldsymbol{H}_1 \in \mathbf{O}[K_2], \boldsymbol{H}_2 \in \mathbf{O}[K_3], \boldsymbol{Q} \in \mathbf{GL}[n+1]\}$$

Its action is defined by

$$\boldsymbol{t}_{\gamma_2} : (\mathbf{X}_{21}, \mathbf{X}_3, \mathbf{W}) \to (\boldsymbol{H}_1 \mathbf{X}_{21} \boldsymbol{Q}, \boldsymbol{H}_2 \mathbf{X}_3 \boldsymbol{Q}, \boldsymbol{Q}' \mathbf{W} \boldsymbol{Q})$$

The problems of testing identifiability and exogenous variable exclusions can also be compared in terms of their invariance groups. Thus, $\boldsymbol{T}_{\mathrm{P}_3} \subseteq \boldsymbol{T}_{\gamma_2}$ since every member of the former is a member of the latter. A larger group of invariance transformations imposes stronger restrictions on the feasible class of procedures (which are based on the sufficient statistics).

<u>Theorem 2.</u> (Test of $\gamma_2 = \mathbf{0}$)

For the problem of testing (10) in the structural equation model (1) and (2), a maximal invariant is given by the function of the sufficient statistics defined by

$$\text{(15)} \qquad \begin{pmatrix} \boldsymbol{H}_1' & \mathbf{0} \\ \mathbf{0} & \boldsymbol{H}_2' \end{pmatrix} \begin{pmatrix} \mathbf{X}_2 \\ \mathbf{X}_3 \end{pmatrix} \mathbf{W}^{-1} (\mathbf{X}_2' \;\; \mathbf{X}_3') \begin{pmatrix} \boldsymbol{H}_1 & \mathbf{0} \\ \mathbf{0} & \boldsymbol{H}_2 \end{pmatrix}$$

where $\boldsymbol{H}_1$ and $\boldsymbol{H}_2$ are arbitrary orthogonal matrices, representing the fact that the latent roots of the two main diagonal blocks and the roots of the off-diagonal block are to be preserved. Equivalently, a maximal invariant is given by the combined set of latent roots of $\mathbf{WS}^{-1}$, $\mathbf{S}_{(0)}\mathbf{W}^{-1}$ and $(\mathbf{S}+\mathbf{S}_{(0)})\mathbf{W}^{-1}$. This set, which we denote by $h_{g_2} = h_{g_2}[\mathbf{X}_2, \mathbf{X}_3, \mathbf{W}]$, has dimensions ($K_2$+2$n$+2), or (3$n$+2), depending on whether $K_2<n$, or $K_2 \geq n$ respectively.

PROOF

It is clear that $h_{g_2}[\mathbf{X}_2, \mathbf{X}_3, \mathbf{W}] = h_{g_2}[\overline{\mathbf{X}}_2, \overline{\mathbf{X}}_3, \overline{\mathbf{W}}]$, using the arbitrary orthogonal matrices as the basic device by which an member of the orthogonal group can be found according to the definitions of the invariance group.

It should be noted that the latent roots both of $\mathbf{WS}^{-1}$, and $\mathbf{S}_{(0)}\mathbf{S}^{-1}$ are invariant, but that they do not form a maximal invariant set. This set has dimensions (K_2+n+1), or (2n+1), depending on whether $K_2<n$, or $K_2 \geq n$ respectively. Under the action of any element $\boldsymbol{t}_{g_2}$ of the invariance group, the following maps occur: $\mathbf{W} \rightarrow \boldsymbol{Q}'\mathbf{W}\boldsymbol{Q}$, $\mathbf{S} \rightarrow \boldsymbol{Q}'\mathbf{S}\boldsymbol{Q}$ and $\mathbf{S}_{(0)} \rightarrow \boldsymbol{Q}'\mathbf{S}_{(0)}\boldsymbol{Q}$. Under this $\mathbf{WS}^{-1} \rightarrow \boldsymbol{Q}'\mathbf{WS}^{-1}\boldsymbol{Q}'^{-1}$, and $\mathbf{S}_{(0)}\mathbf{S}^{-1} \rightarrow \boldsymbol{Q}'\mathbf{S}_{(0)}\mathbf{S}^{-1}\boldsymbol{Q}'^{-1}$, thus preserving the (ordered) eigenvalues in each case. To see if these invariant statistics are maximal invariant, let there be two samples of observations which produce the same value of the latent roots of the these matrices. Denote these samples by $(\mathbf{X}_2, \mathbf{X}_3, \mathbf{W})$ and $(\overline{\mathbf{X}}_2, \overline{\mathbf{X}}_3, \overline{\mathbf{W}})$. If $\mathbf{WS}^{-1}$, and $\mathbf{S}_{(0)}\mathbf{S}^{-1}$ are maximal invariant, then it should be possible to relate the two samples of sufficient statistics by a member of the transformation group. The transformation matrices will be determined by these given samples. Now, $H_1'\mathbf{U}'^{-1}\mathbf{WU}^{-1}H_1 = \overline{\mathbf{U}}'^{-1}\overline{\mathbf{WU}}^{-1}$ and $\overline{\mathbf{W}} = \overline{\mathbf{U}}'H_1'\mathbf{U}'^{-1}\mathbf{WU}^{-1}H_1\overline{\mathbf{U}} = \overline{\boldsymbol{Q}}'\mathbf{W}\overline{\boldsymbol{Q}}$ by defining $\overline{\boldsymbol{Q}} = \mathbf{U}^{-1}H_1\overline{\mathbf{U}}$. This also gives $\overline{\mathbf{U}} = H_1'\mathbf{U}\overline{\boldsymbol{Q}}$, implying that $\overline{\mathbf{S}} = \overline{\boldsymbol{Q}}'\mathbf{S}\overline{\boldsymbol{Q}}$, from which we get $\overline{\mathbf{X}}_3 = \boldsymbol{H}_1\mathbf{X}_3\overline{\boldsymbol{Q}}$, for some

arbitrary $\boldsymbol{H}_1 \varepsilon \mathbf{O}[K_2]$. If the latent roots are equal, then there exists some (possibly non-unique) matrix $H_2 \varepsilon \mathbf{O}[n+1]$ such that $H_2' \mathbf{U}'^{-1} \mathbf{S}_{(0)} \mathbf{U}^{-1} H_2 = \overline{\mathbf{U}}'^{-1} \overline{\mathbf{S}}_{(0)} \overline{\mathbf{U}}^{-1}$. This gives $\overline{\mathbf{S}}_{(0)} = \overline{\mathbf{U}}' H_2' \mathbf{U}'^{-1} \mathbf{S}_{(0)} \mathbf{U}^{-1} H_2 \overline{\mathbf{U}}$, where the r.h.s. is of the form $\overline{\boldsymbol{Q}_1}' \mathbf{S}_{(0)} \overline{\boldsymbol{Q}_1}$ with some $\overline{\boldsymbol{Q}_1} = \mathbf{U}^{-1} H_2 \overline{\mathbf{U}}$. Working through, we have $\overline{\mathbf{U}} = H_2' \mathbf{U} \overline{\boldsymbol{Q}_1} \Rightarrow \overline{\mathbf{S}} = \overline{\boldsymbol{Q}_1}' \mathbf{S} \overline{\boldsymbol{Q}_1} \Rightarrow \overline{\mathbf{X}}_3 = \boldsymbol{H}_2 \mathbf{X}_3 \overline{\boldsymbol{Q}_1}$ for some arbitrary $\boldsymbol{H}_2 \varepsilon \mathbf{O}[K_2]$ and note that $\overline{\boldsymbol{Q}_1} \in \mathbf{GL}[n+1]$ is also uniquely determined. (Q.E.D.)

When $K_2 \geq n+1$, the $\mathbf{S}_{(0)}$ becomes non-singular, in which case the second matrix in theorem 2 may equivalently be replaced by its inverse. This case is simpler than dealing $H_0^{\mathrm{b}_2}$, as intuition would lead us to expect when only the exogenous variables are involved. This is a reflection of the fact that the null in this case is implied by the maintained hypothesis. Since the rank of an augmented matrix $(\Pi_2' \quad \Pi_3')$ cannot be less than the rank of the original Π_3 anyway, the relationship between the null and the maintained hypotheses in (9)-(11) is partly an algebraic tautology and quite independent of the dimensions involved.

4.2. Testing endogenous variables

By virtue of its role as the data generating process, (2) captures all the sample information as well as any inference these may supply for the structural coefficients (β,γ). This is the reason for re-stating the above in terms of its consequences for the parameters of (2). Provided the maintained hypothesis applies, the above may be re-formulated as a rank testing problem of

$H_0^{\mathrm{b}_2}: rank[\Pi_{31}] = n_1$ against $H_1^{\mathrm{b}_2}: rank[\Pi_{31}] = n_1 + 1$.

That these take on the form of rank restrictions is the reason why invariant aspects confined to the linear aspects of (2), though well known (see Muirhead (1982), chap.10), actually throw very little light on the present problem.

The space of the sufficient statistics ($\mathbf{X}$,$\mathbf{W}$) may be regarded as the sample space re-defined, where $\mathbf{X}$ has the multivariate normal $N(\Pi, \mathbf{I}_{K(n+1)})$, $\mathbf{W}$ has the Wishart distribution $W_{n+1}(T\text{-}K, \mathbf{I}_{n+1})$, and they are independent.

Re-formulating the hypotheses of interest as rank tests reveals the asymmetric that the maintained hypothesis plays. The rank of a matrix is the maximum number of linearly independent rows or columns, or alternatively, the order of its largest non-vanishing minor. Thus, in the case where exclusion of exogenous variables is under test, the maintained hypothesis clearly implies the null.

The transformations groups used to define the maximal invariants for (9), (10) and (11) are specified as follows (Hillier, 1987). By observing that the problem of testing $H_0^{b_2}$ is invariant under the map $\mathbf{Y}\rightarrow \boldsymbol{H}\mathbf{Y}\boldsymbol{Q}_{11}$, $\mathbf{Z}_1 \rightarrow \boldsymbol{H}_0\mathbf{Z}_1\boldsymbol{H}'$, $\mathbf{Z}_2 \rightarrow \boldsymbol{H}_0\mathbf{Z}_2\boldsymbol{H}'$where $\boldsymbol{Q}_{11}\in\mathbf{GL}[n_1+1]$. At the same time, the condition (4) must also be maintained, which would occur under additional maps of the form $\mathbf{Y}\rightarrow\boldsymbol{H}_0\mathbf{Y}\boldsymbol{Q}$,. Together with the preceding, this means that we must have $\boldsymbol{Q} = \begin{pmatrix} \boldsymbol{Q}_{11} & \boldsymbol{Q}_{12} \\ \mathbf{0} & \boldsymbol{Q}_{22} \end{pmatrix}$. Thus, the required invariance group is

$$\boldsymbol{T}_{b_2} = \{\boldsymbol{t}_{b_2} = (\boldsymbol{H}, \boldsymbol{Q}): \boldsymbol{H} \in \mathbf{O}(K_3), \boldsymbol{Q}_{11} \in \mathbf{GL}(n_1+1),$$
$$\boldsymbol{Q}_{12} \in \mathbf{M}(n_1+1, n_2), \boldsymbol{Q}_{22} \in \mathbf{GL}(n_2)\}$$

The group acts on the space of the sufficient statistic ($\mathbf{X}$,$\mathbf{W}$) in

$$\boldsymbol{t}_{b_2} : (\mathbf{X}_{31}, \mathbf{X}_{32}, \mathbf{W}) \rightarrow (\boldsymbol{H}\mathbf{X}_{31}\boldsymbol{Q}_{11}, \boldsymbol{H}\mathbf{X}_{31}\boldsymbol{Q}_{12} + \boldsymbol{H}\mathbf{X}_{32}\boldsymbol{Q}_{22}, \boldsymbol{Q}'\mathbf{W}\boldsymbol{Q})$$

Note that this implies that $\boldsymbol{Q}$ belongs to a subgroup of **GL**[*n*+1] whose elements are upper block-triangular matrices. In the specification of $\boldsymbol{\mathcal{T}}_{\mathrm{b}_2}$, the four groups which appear are also proper subgroups. That this is a more difficult testing problem than identifiability can now be made clear in an invariance sense. Although every element of the group can be uniquely decomposed as a product of elements of the subgroups, not *all* of the actions of the subgroups commute. Whereas $\boldsymbol{\mathcal{T}}_{\mathrm{P}_3}$ can be written as the direct product of its subgroups, $\boldsymbol{\mathcal{T}}_{\mathrm{b}_2}$ cannot. The reason is that

$$\mathbf{GL}[n+1] \neq \mathbf{GL}[n_1+1] \times \mathbf{M}[n_1+1, n_2] \times \mathbf{GL}[n_2]$$

Under the group action, any two elements in the space $\boldsymbol{X}$ which differ only in $(\mathbf{X}_1 \quad \mathbf{X}_2)$ are always related by some member of the group. Thus the group acts transitively on the subspace of the sufficient statistic spanned by $(\mathbf{X}_1 \quad \mathbf{X}_2)$. On the other hand, the action of the group on the remaining elements of the sufficient statistic does not give an injection in general, so that the subspace spanned by these elements may be divided in non-trivial orbits.

The tests are based on identifiability first having been established. This need not always occur by explicitly including the original compatibility condition (5), especially if the null and alternative hypotheses contain updated information about identification which supersedes it.

Whenever transformations of a symmetric positive definite matrix to latent roots are made using a spectral decomposition in the following, it will be assumed that this produces a square matrix with the roots arranged in ascending order. The main result is stated as the following theorem.

Theorem 3 (Test of β_2=**0**)

For the problem of testing (10) in the structural equation model (1) and (2), the maximal invariant is given by

$$h_{b_2}[\mathbf{X}_2, \mathbf{W}] = H'\mathbf{R}H$$

where $H = H[\mathbf{X}_2, \mathbf{W}]$ has the block diagonal form

(16) $H = \begin{pmatrix} H_1 & \mathbf{0} \\ \mathbf{0} & H_2 \end{pmatrix}$, $H_1 \in \mathbf{O}[n_1+1]$, $H_2 \in \mathbf{O}[n_2]$,

and is chosen to transform $\mathbf{R}$ to a matrix where the latent roots of $\mathbf{R}_{11}$ and $\mathbf{R}_{22}$ appear simultaneously along the principal diagonal instead.

PROOF

First, note that based on the specification of the invariance group, we need to preserve the rank of the submatrix $\mathbf{W}_{11}\mathbf{S}_{11}^{-1}$ - but not necessarily that of $\mathbf{W}_{22}\mathbf{S}_{22}^{-1}$ - in addition to the rank of $\mathbf{W}\mathbf{S}^{-1}$. Therefore, we must first write the latter in a form from which the latent roots of $\mathbf{W}_{11}\mathbf{S}_{11}^{-1}$ can be recovered. Also, $\mathbf{X}_1$ has no bearing on the problem because the action of □ $_{b_2}$ on the subspace spanned by $\mathbf{X}_1$ is transitive. The action of □ $_{b_2}$ via a specific group element □$_{b_2}$ transform ($\mathbf{X}_2$,$\mathbf{W}$) to (H$\mathbf{X}_2$Q,Q'$\mathbf{W}$Q). Now, $\mathbf{S}=\mathbf{X}'_2\mathbf{X}_2$ is transformed to $\mathbf{S}^*$=Q'$\mathbf{S}$Q=Q'$\mathbf{U}$'$\mathbf{U}$Q. Since the product of $\mathbf{U}$ and Q have the same block-triangular structure, their product must have the same. Comparing the block-triangular decompositions of $\mathbf{S}^*$=($\mathbf{U}$Q)'($\mathbf{U}$Q) with that of $\mathbf{S}$ demonstrates that

(17) □$_{b_2}$:$\mathbf{U} \rightarrow H_1\mathbf{U}$Q

where H_1 will be block triangular of the form (9) because it must also preserve block equivalence. Let $\mathbf{U}'^{-1}\mathbf{W}\mathbf{U}^{-1} = \overline{\mathbf{U}}'^{-1}\overline{\mathbf{W}\mathbf{U}}^{-1}$. Then $\overline{\mathbf{W}} = \overline{□}'\mathbf{W}\overline{□}$ where $\overline{□} = \mathbf{U}^{-1}\overline{\mathbf{U}} \in \mathbf{GL}[n+1]$. Obviously, $\overline{□}$ has the required upper block triangular structure. This also immediately yields $\overline{\mathbf{U}}[\overline{\mathbf{X}}_2, \overline{\mathbf{W}}] = \mathbf{U}[\mathbf{X}_2, \mathbf{W}]\overline{□}$. Now $\mathbf{U}$ is a uniquely defined by $\mathbf{S}$ through a block-triangular decomposition. At the same time, the fact that $\mathbf{U}$ and $\overline{□}$ have the same partitioned structure means that they belong to the same subgroup of $\boldsymbol{GL}[n+1]$, implying that $\overline{\mathbf{U}}$ also uniquely determines a block triangular. To complete the proof, we must show that this leads to $\overline{\mathbf{X}}_2 = \overline{□}\mathbf{X}_2\overline{□}$ for some $\overline{□} \in \mathbf{O}[K_2]$. This is clearly possible

Note: There may be concerns about whether $\mathbf{U}$ needs to come from a full Cholesky decomposition or whether block triangularity would suffice. This has a huge bearing on the result: mere block triangularity means that the matrix itself is a maximal invariant, while the former produces a stronger result in the sense that the sense that the maximal invariant is the set of eigenvalues and therefore of lower dimension. To see that the less stringent form is called for, note that in order for (10) to be possible, we need to define those playing the role of "square root matrices" or $\mathbf{S}^{1/2}$, such as $\mathbf{U}$ in (17), to be non-symmetric. With $\mathbf{X}'_{21}\mathbf{X}_{21}$ real, symmetric and non-singular by assumption, this can occur *either* by way of a Cholesky decomposition or by defining the non-symmetric equivalent from a spectral decomposition. The latter gives $\mathbf{S}_{11}^{1/2} = (\mathbf{X}'_{21}\mathbf{X}_{21})^{1/2} = \mathrm{L}_{\mathrm{S}}^{1/2} H_{\mathrm{S}} \in \boldsymbol{GL}(n_1+1)$. The Gram-Schmidt orthogonalization process shows that for any dimension m, $\boldsymbol{GL}(m)$ is isomorphic to the product space $\mathbf{O}(m)$ $\mathbf{T}(m)$, enabling us to write $\mathrm{Q} = \mathrm{H_Q T_Q}$ with $\mathrm{H_Q} \in \mathbf{O}(m)$ and $\mathrm{T}^* \in \mathbf{T}(m)$ are unique to Q. Since the inverse of Q must also satisfy a similar relationship, we have that $\mathrm{Q} = □^{-1}_{□^{-1}} □^{-1}_{□^{-1}} = □^{*}_{□} □^{*}_{□}$. In all cases, the triangular matrices involve are the upper triangular factors. This process of factorization of any element of the general linear group into a product of orthogonal and triangular matrices can also be reversed. (Q.E.D.)

First, note that the matrix is dependent on a more complicated function of the parameter matrix than just its latent roots. In fact, we can write $h_{\beta_2}[\mathbf{X}_3, \mathbf{W}] = \{H_1'\mathbf{R}_{11}H_1, H_1'\mathbf{R}_{12}H_2, H_2'\mathbf{R}_{22}H_2\}$ where H_1 and H_2 depend of $\mathbf{R}_{11}$ and $\mathbf{R}_{22}$ respectively because they must be selected to produce to produce a specific ordering of their latent roots. Clearly the latent roots of $\mathbf{R}_{11}=\mathbf{W}_{11}\mathbf{S}_{11}^{-1}$ and $\mathbf{R}_{22}$ are invariant but not maximal invariant. The same is true of the latent roots alone of the entire matrix $\mathbf{WS}^{-1}$. Separately, they capture only part of the information content of the maximal invariant, while combined they involve superfluous components. The definition of $\mathbf{R}$ is justified by the fact that it captures crucial information about the ranks of the whole matrix $\mathbf{WS}^{-1}$ and the relevant sub-matrix $\mathbf{W}_{11}\mathbf{S}_{11}^{-1}$.

In general, $\mathbf{U}$ belongs to a subgroup of Γ_Q, whereas H belongs a subgroup of this subgroup. This is because H should, in addition, be orthogonal. This makes it block-diagonal. In the extreme, if Q and therefore $\mathbf{U}$ were fully triangular, H would be both diagonal and orthogonal, of which the only possibility is the identity matrix or its negative

COROLLARY

Let the group of transformations $\square_{b_2 6}$ be such that its factor group $\boldsymbol{G_Q} \in \mathbf{GL}[n+1]$, identified with the element $\boldsymbol{Q}$, has a specific (non-symmetric) partitioned structure. Then for a function given by $h_{b_2}[\mathbf{X}, \mathbf{W}]$ to be maximal invariant, the element $\mathbf{U}$ in its specification must be a member of the group $\boldsymbol{G_Q}$.

In other words, the specification of a maximal invariant under $\boldsymbol{T}_{b_2}$ depends critically on the fact that the $\mathbf{U}$ must have exactly the same (non-symmetric) partitioned structure as $\boldsymbol{Q}$.

4.3 Joint test

Finally, we have the following:

Theorem 4. (Test of b_2=**0** and g_2=**0**)

For the problem of testing (10) in the structural equation model (1) and (2), a maximal invariant is given by the function of the sufficient statistics defined by

$$(18) \qquad \begin{pmatrix} \boldsymbol{H}_3' & \mathbf{0} \\ \mathbf{0} & \boldsymbol{H}_4' \end{pmatrix} \begin{pmatrix} \mathbf{X}_{21} \\ \mathbf{X}_{31} \end{pmatrix} \mathbf{W}_{11}^{-1} (\mathbf{X}_{21}' \quad \mathbf{X}_{31}') \begin{pmatrix} \boldsymbol{H}_3 & \mathbf{0} \\ \mathbf{0} & \boldsymbol{H}_4 \end{pmatrix}$$

where $\boldsymbol{H}_3$ and $\boldsymbol{H}_4$ are arbitrary orthogonal matrices, representing the fact that the latent roots of the two main diagonal blocks and the roots of the off-diagonal block are to be preserved. Equivalently, a maximal invariant is given by the combined set of latent roots of $\mathbf{W}_{11}\mathbf{S}_{11}^{-1}$, $\mathbf{S}_{(0),11}\mathbf{W}_{11}^{-1}$ and $(\mathbf{S}+\mathbf{S}_{(0)})_{11}\mathbf{W}_{11}^{-1}$. This set, which we denote by $h_{b_2,g_2} = h_{b_2,g_2}[\mathbf{X}_{21}, \mathbf{X}_{31}, \mathbf{W}_{11}]$, has dimensions ($3n_1$+3).

5. Discussion

It is now possible to clarify the precise difference between identifiability and variable exclusions in simultaneous equations modeling in terms of inference. Both tests of identifiability and variable exclusions may be naturally viewed as rank tests in the context of general linear model that serves as the DGP. Employing concrete representations of the invariance approach provides a useful perspective from which to view these notions.

First of all, consider the three categories of hypotheses about variable exclusions which may be tested. The first is also the most important, dealing with exclusions of exogenous variables. Exogenous variables which are excluded from an equation determine its identifiability. The importance of this is amplified when an equation is poised on the brink of becoming unidentifiable at $K_3=n$. An equation that is just identified depends critically on including one more exogenous variable.

The compatibility requirement $rank[\Pi_3]=n$ cannot be maintained when testing $H_0^{g_2}: g_2 = 0$. That is because if the null turns out to be true, the maintained will be superseded, raising the danger of logical inconsistencies later on. Indeed, these show up as a model whose maximal invariant has no real distribution.

In some sense, study of the maximal invariants reinforces views held in the literature about the logical order in which to approach the issue of identifiability and variable exclusions.

In practice, these issues have always been dealt with along the lines of borrowed logic. By appealing to the notion of apparent identifiability, means have been found to proceed more expeditiously to the actual modeling and estimation. It is usually after estimates have been obtained that attention to tests about possible variable exclusions arise. In the general linear models where this approach is used, all the explanatory variables play a symmetric role in terms of model determination. In a curved exponential model like (2) when (1) holds, however, this is not true. Thus, while overfitting (with regressors) may be safe in linear models, and indeed present less peril than underfitting, this is clearly not feasible with exogenous variables in a structural equation because of the tendency that would have of violating identifiability.

Overfitting with endogenous variables present a problem, not for the same reason but for the other side of it.

In terms of the invariance groups, note that $\boldsymbol{T}_{P_3} \subseteq \boldsymbol{T}_{g_2}$ by virtue of the fact that every member of the former is a member of the latter. A larger group of invariance transformations imposes more stringent restrictions on the feasible class of procedures (which are based on the sufficient statistics), thus making identifiability the natural prerquisite.

Concluding Remarks

Our analysis shows that tests of identifiability and variable exclusions are closely linked. Indeed, the latter can be specified as extensions of the rank condition for identifiability in cases where further exclusions of the variables are being considered. Basing our choice on the class of invariant procedures which are functions of maximal invariants, we showed above that the only other hypothesis distinct from identifiability that requires testing is that of endogenous variable exclusions. Thus, the invariance approach confirms the general intuition about asymmetry in the roles which are ascribed to the endogenous and exogenous variables. Borrowing the terminology of linear regression, this means that one should always test that the exogenous regressors have significant coefficients before attempting to do the same for regressors which are endogenous.

Other implications also derive from the matricvariate nature of the inference problem. The most obvious difference from the linear model is that a different maximal invariant exists for each of the three possible class of coefficient restrictions. Although curious from the point of view of the seemingly innocuous restrictions placed on (1), this situation again is one that is inherent in the curved exponential nature of the model (2). It actually allows for more focused inference by capturing the essentially separate burdens placed on the endogenous and exogenous components by the model's simultaneous structure.

As a result of the maximal invariants being of matrix form, the inverted matrix is also maximal invariant in each case. This would be helpful in situations where it is sometimes easier to deal with the distributions of the inverse such as when, for instance, the non-centrality parameter can be conditioned out (Tan,1995). Evidently, this would be true only of transformations of the maximal invariant which preserve its dimensionality. The results demonstrate that dimension-reducing transformations relate to closely related problems which are nonetheless distinct.

A prominently related problem is that of testing a sequence of hypotheses, usually in the interests of progressively refining the model. If sequential testing were to be necessary, then more stringent conditions must be placed on the transformation groups which leave the testing procedures invariant. Specifically, by iterating the arguments put forward in the proof of theorem 3 about the conditions on Q_{11}, it follows that $\mathbf{Q}_S$ has to be fully triangular. A correspondingly stronger result is then produced for which the dimension of the maximal invariant is substantially reduced. In the leading case of a model with only two endogenous variables, both methods merge. The limited permutation of available alternative models in this case offers another way in which to view how much simplification is involved in the leading case.

These results offer an explanation for the phenomenon found in Morimune's (1989) simulation results. The matrix variates defined in theorems (2), (3) and (4) are pivot elements preserved in the LIML and LR procedures, but clearly distorted in the case of OLS and TSLS. In terms of *k*-class estimation, the criterion for optimization must treat the *all* endogenous variables symmetrically - distinguishing. all endogenous variables (dependent or independent) from exogenous variables, We can go even further and conjecture that Morimune's (1989) results for the size extend in the same way to test power performance, a fact already supported by existing asymptotic analysis.

Finally, we note that there is an analogy with Hillier's (1991) findings that OLS/TSLS suffer the most distortion from arbitrary normalization while LIML suffers the least.

Acknowlegment

I am indebted to Grant Hillier for having allowed me access to his unpublished paper during my dissertation work at Monash. Responsibility for any mistakes in this paper or in interpreting his results is mine alone. I also thank Shahidur Rahman for helpful comments.

References

Constantine, A.G. (1963). “Some non-central distribution problems in multivariate analysis,” *Annals of Mathematical Statistics* 34.

Hillier, G. H. (1987). “Hypothesis testing in a structural equation: Part I. Reduced form equivalences and invariant test procedures,” unpublished manuscript, Monash University.

Hillier, G. H. (1990). "On the Normalization of Structural Equations: Properties of Direction Estimators," *Econometrica* 58, 1181-1194.

Hosoya, Y., Y. Tsukuda and N. Terui (1989). “Ancillarity and the limited information maximum-likelihood estimation of a structural equation in a simultaneous equation system,” *Econometric Theory*, 5, 385-404.

Kadane, J. B. (1974), "Testing a subset of the overidentifying restrictions," *Econometrica* 42, 853-867.

Lehmann, E. L. (1986). Testing Statistical Hypotheses. Second edn. Wiley, New York.

Morimune, K. (1989). "*t* test in a structural equation," *Econometrica* 57, 1341-1360.

Morimune, K. and Y. Tsukuda (1984), "Testing a subset of coefficients in a structural equation," *Econometrica* 52, 427-448.

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

Phillips, P.C.B. (1983), "Exact small sample theory in the simultaneous equations model," in *Handbook of Econometrics*, ed. by M.D. Intriligator and Z. Griliches. North-Holland, Amsterdam.

Revanker, N.S. and P. Mallela (1972). "The power function of an F-test in the context of a structural equation," *Econometrica*, 40, 913-916.

Tan, R.G.K. (1995). "Likelihood Ratio Test for the Coefficient of an Endogenous Variable in a Structural Equation." SABRE Centre working paper no. 19/95, Nanyang Business School, Nanyang Technological University..