90-F-1

Tests of Overidentification and Exogeneity in Simultaneous Equation Models

by

T. W. Anderson
Department of Statistics
Stanford University

and

Naoto Kunitomo
Faculty of Economics
University of Tokyo

January 1990

1. Introduction

Two important underlying assumptions of the traditional simultaneous equation approach in econometrics are the identifying restrictions and predeterminedness (or exogeneity in some sense) of several variables in the system of structural equations. Although these assumptions are often made on a priori ground, in practice it may be advisable to examine these two conditions from a statistical point of view. In this respect a number of statistical testing procedures for these restrictions have been proposed by econometricians. For instance, the test procedures of Anderson and Rubin (1949), Koopmans and Hood (1953), Basmann (1960), Byron (1972), Wu (1973), Revankar and Hartley (1973), Revankar (1978), Hausman (1978), Kariya and Hodoshima (1980), Hwang (1980a), Hillier (1987), and Revankar and Yoshino (1989) among many others have drawn attention and have been applied in empirical works. However, since many testing procedures have been introduced based on intuitive reasoning, it may be difficult to understand the meaning of the statistics proposed.

The main purposes of this paper are to derive systematically several test procedures for each condition and to obtain the relationships among the different test statistics. For these intentions we consider a subsystem of structural equations and regard the single equation method as a special case of our formulation. Then we shall derive three types of test procedures, namely, the likelihood ratio (LR) test, Lagrange Multiplier (LM) test, and the Wald test for the block identifiability restrictions and the predeterminedness restrictions in the subsystem of structural equations. In this framework the test statistics we shall derive include most of the test statistics mentioned above as special cases and give new interpretations to some of them. These interpretations also apply to some test statistics commonly known in multivariate statistical analysis.

In a subsequent paper we shall derive the asymptotic distributions of these test criteria under very general conditions, based on a new central limit theorem using a Lindeberg-type condition for martingale differences [Anderson and Kunitomo (1989a,b)].

In Section 2 we formulate a subsystem of structural equations. In Section 3 we derive several statistics for testing identifying restrictions. We also relate these statistics to the statistics in multivariate statistical analysis yielding new interpretations of statistics commonly known among statisticians. Subsequently, in Section 4 we derive a number of test procedures for testing econometric predeterminedness restrictions. Finally, in Section 5 some concluding remarks are given. Useful lemmas are given in the appendices.

2. Two Hypotheses in a Subsystem of Structural Equations

2.1. The model.

We consider a subsystem of G_0 structural equations

$$YB = Z_1\Gamma + U, \tag{2.1}$$

where Y is a $T \times G$ matrix of observations on the endogenous variables appearing in the first G_0 structural equations, Z_1 is a $T \times K_1$ matrix of observations on the K_1 exogenous variables, B and Γ are $G \times G_0$ and $K_1 \times G_0$ matrices of (unknown) parameters, respectively, and U is a $T \times G_0$ matrix of unobservable disturbances. When $G_0 = 1$, (2.1) is the usual single structural equation. We require the columns of B to be linearly independent; that is, the rank of B is G_0.

The reduced form equation for the endogenous variables Y appearing in the first G_0 structural equations (2.1) with K ($K = K_1 + K_2$) predetermined variables is

$$Y = Z\Pi + V, \tag{2.2}$$

where $Z = (Z_1, Z_2)$ is a $T \times K$ matrix of predetermined variables ($T > K$) of rank K, and Z_2 is a $T \times K_2$ matrix of predetermined variables that are not included in (2.1). The predetermined variables may include lagged endogenous variables. V is a $T \times G$ matrix of disturbances whose t–th row is denoted by v_t'. We assume that

$$E(v_t) = 0, \tag{2.3}$$

$$E(v_t v_t') = \Omega, \tag{2.4}$$

where Ω is a $G \times G$ positive definite matrix.

In this paper we shall consider two hypotheses. One is that the set of G_0 equations (2.1) is identified as a block. That is, any matrix B such that $Z\Pi B = Z_1\Gamma$ for some Γ is obtained from any other by multiplication on the right by a nonsingular $G_0 \times G_0$ matrix. The other hypothesis that we consider is that a subset of the endogenous variables is uncorrelated with the disturbances in the block of equations.

2.2. Block identification.

The relationship between the reduced form and the structural equations involves

$$\Gamma = \Pi_{1\cdot}B, \tag{2.5}$$

$$U = VB. \tag{2.6}$$

where Π has been partitioned as

$$\Pi = \begin{pmatrix} \Pi_{1\cdot} \\ \Pi_{2\cdot} \end{pmatrix}. \tag{2.7}$$

Let u_t' be the t–th row of U. From (2.3), (2.4), and (2.6) we obtain

$$E(u_t) = 0, \tag{2.8}$$

$$E(u_t u_t') = B'\Omega B = \Sigma, \tag{2.9}$$

where Σ is a $G_0 \times G_0$ positive definite matrix. The block identifiability conditions are expressed as

$$H_\xi : \xi = 0, \tag{2.10}$$

where

$$\xi = \Pi_{2\cdot}B. \tag{2.11}$$

From (2.11) we obtain the rank condition for the identifiability of (2.1),

$$\mathrm{rank}\Pi_{2\cdot} = G - G_0. \tag{2.12}$$

The order condition is

$$L = K_2 - (G - G_0) \geq 0. \tag{2.13}$$

In the above notation, L is called the degree of overidentification.

Let $\nu_G \geq \cdots \geq \nu_1 \geq 0$ be the roots of

$$\left| \frac{1}{T}\Theta_T - \nu\Omega \right| = 0, \tag{2.14}$$

where

$$\Theta_T = \Pi_{2\cdot}' A_{22\cdot 1} \Pi_{2\cdot}, \tag{2.15}$$

$$A_{22\cdot 1} = Z_2' Z_2 - Z_2' Z_1 (Z_1' Z_1)^{-1} Z_1' Z_2. \tag{2.16}$$

The block identifiability conditions are equivalent to the hypothesis $H_\nu : \nu_1 = \cdots = \nu_{G_0} = 0$ and $\nu_{G_0+1} > 0$. The existence of a matrix B such that $\xi = 0$ is equivalent to (2.12), which, in turn, is equivalent to H_ν.

Note that the model (2.1) and the hypothesis H_ξ are invariant with respect to linear transformations on the right; that is, (2.1) and (2.10) can be multiplied on the right by an arbitrary nonsingular matrix A to yield another set of structural parameters

$$\widetilde{B} = BA, \; \widetilde{\Gamma} = \Gamma A, \; \widetilde{\Sigma} = A'\Sigma A. \tag{2.16$'$}$$

It may be convenient for some purpose to select a particular triple (B, Γ, Σ) by a suitable normalization of B such as requiring a submatrix of B to be I_{G_0}. The test procedures are invariant with respect to the group of transformations and hence do not depend on the normalization.

2.3. Exogeneity.

An essential difference between a system of structural equations and regression models in the multivariate analysis is that in the former correlation may exist between the endogenous variables y_t, which is the t–th row of Y, that is, v_t and the corresponding disturbance term u_t', but in the latter some components of y_t' and u_t' may be uncorrelated. In order to state this hypothesis we partition $Y = (Y_1, Y_2)$ into G_1 and G_2 columns $(G = G_1 + G_2)$, $V = (V_1, V_2)$, and

$$\Omega = \begin{pmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{pmatrix}. \tag{2.17}$$

From (2.9) the covariance matrix of v_{2t}' and u_t' is

$$\begin{aligned} \eta &= \mathrm{Cov}(v_{2t}, u_t) \\ &= (\Omega_{21}, \Omega_{22})B. \end{aligned} \tag{2.18}$$

We define the econometric predeterminedness restriction considered in this paper to be the hypothesis $H_\eta : \eta = 0$. The two hypotheses H_ξ and H_η imply the hypothesis $H_{\xi,\eta} : \xi = 0,\ \eta = 0$. When the disturbance terms follow the multivariate normal distribution, the uncorrelatedness implies an independence between a subset of regressors Y_2 and disturbance terms in (2.1). This testing problem has been sometimes called the test of independence. The hypothesis of predeterminedness in this paper has also been called weak exogeneity in econometrics. There are several different concepts of econometric exogeneity in simultaneous equation systems. Engle, Hendry, and Richard (1983) surveyed this issue in a systematic way. See Holly (1987), also.

The hypothesis of exogeneity is also invariant with respect to linear transformations on the right; that is, η in (2.18) can be multiplied on the right by a nonsingular matrix A. The test criteria for exogeneity are also invariant with respect to such transformations and hence with respect to normalization of B.

In Section 3 we obtain the likelihood ratio, the Lagrange multiplier, and Wald-type tests of over-identification. In Section 4 we find test criteria for predeterminedness.

3. Tests of Block Identifiability

In order to derive test statistics we assume that the disturbance terms $\{v_t\}$ are independently and normally distributed. The derivation here is considerably simpler than alternatives already known.

3.1. Likelihood ratio test.

Under the assumption of normal disturbances, the log likelihood function for $Y' = (y_1, \ldots, y_T)$ is

$$\log L_1 = c_1 - \frac{1}{2}T\log|\Omega| - \frac{1}{2}\mathrm{tr}(Y - Z\Pi)'(Y - Z\Pi)\Omega^{-1}, \tag{3.1}$$

where $c_1 = -\frac{1}{2}GT\log(2\pi)$. To maximize L_1 with respect to the covariance matrix Ω, we use Lemma A.1. The concentrated likelihood function is

$$\log L_2 = c_2 - \frac{1}{2}T\log|S|, \tag{3.2}$$

where $c_2 = c_1 + \frac{1}{2}GT\log T - \frac{1}{2}GT$ and $S = (Y - Z\Pi)'(Y - Z\Pi)$. The log likelihood maximized under the alternative hypothesis $H_A : \xi \neq 0$ (that is, no restriction) is

$$\log L_3 = c_2 - \frac{1}{2}T\log|Y'\overline{P}_Z Y|, \tag{3.3}$$

where $P_F = F(F'F)^{-1}F'$ denotes the projection operator onto the space spanned by the columns of F and $\overline{P}_F = I - P_F$ for any matrix F of full column rank.

To maximize the likelihood under the null hypothesis $H_\xi : \xi = 0$ we define the $G \times G$ matrix

$$H = \begin{pmatrix} B & \begin{matrix} 0 \\ I_{G*} \end{matrix} \end{pmatrix}. \tag{3.4}$$

where $G_* = G - G_0$. Since B is of rank G_0, there is a $G_0 \times G_0$ submatrix that is nonsingular. When the numbering of components of y_t is such that this matrix consists of the first G_0 rows of B, then H is nonsingular. The concentrated likelihood is rewritten as

$$\log L_2 = c_2 + \frac{1}{2}T\log|H'H| - \frac{1}{2}T\log|S^*|, \tag{3.5}$$

where $S^* = H'SH = (W - Z\Pi^*)'(W - Z\Pi^*), W = (W_0, W_*) = YH = (YB, Y_*)$, and

$$\Pi^* = (\Pi^*_{\cdot 0}, \Pi^*_{\cdot *}) = \Pi H = \begin{pmatrix} \Gamma & \Pi^*_{\cdot *} \\ \xi & \end{pmatrix}. \tag{3.6}$$

The unrestricted least squares estimate of Π^* is

$$\widehat{\Pi}^* = (\widehat{\Pi}^*_{\cdot 0}, \widehat{\Pi}^*_{\cdot *}) = (Z'Z)^{-1}Z'W. \tag{3.7}$$

Then

$$(3.8)\quad \begin{aligned} S^* &= (W - Z\Pi^*)'(W - Z\Pi^*) \\ &= W'\overline{P}_Z W + (\widehat{\Pi}^* - \Pi^*)Z'Z(\widehat{\Pi}^* - \Pi^*) \\ &= W'\overline{P}_Z W + \left[Z(\widehat{\Pi}^*_{\cdot 0} - \Pi^*_{\cdot 0}), Z(\widehat{\Pi}^*_{\cdot *} - \Pi^*_{\cdot *})\right]'\left[Z(\widehat{\Pi}^*_{\cdot 0} - \Pi^*_{\cdot 0}), Z(\widehat{\Pi}^*_{\cdot *} - \Pi^*_{\cdot *})\right]. \end{aligned}$$

By Lemma A.2 the minimum of $|S^*|$ with respect to $\Pi_{\cdot *} = \Pi^*_{\cdot *}$ is

$$(3.9)\qquad \frac{|W'\overline{P}_Z W| \cdot |W_0'\overline{P}_Z W_0 + (\widehat{\Pi}^*_{\cdot 0} - \Pi^*_{\cdot 0})'Z'Z(\widehat{\Pi}^*_{\cdot 0} - \Pi^*_{\cdot 0})|}{|W_0'\overline{P}_Z W_0|}.$$

The second determinant in the numerator of (3.9) is $|(W_0 - Z\Pi^*_{\cdot 0})'(W_0 - Z\Pi^*_{\cdot 0})|$, which is $|(W_0 - Z_1\Gamma)'(W_0 - Z_1\Gamma)|$ if $\xi = 0$. That determinant is minimized with respect to Γ at $\widehat{\Gamma} = (Z_1'Z_1)^{-1}Z_1'YB$. The log likelihood ratio criterion is the maximum with respect to B of

$$(3.10)\qquad \frac{1}{2}T\log\frac{|S| \cdot |H'H| \cdot |W_0'\overline{P}_Z W_0|}{|W'\overline{P}_Z W| \cdot |W_0'\overline{P}_{Z_1} W_0|} = \frac{1}{2}T\log\frac{|B'Y'\overline{P}_Z YB|}{|B'Y'\overline{P}_{Z_1} YB|}.$$

Lemma A.3 implies that the maximum of (3.10) is $T/2$ times the sum of the G_0 smallest characteristic roots of $Y'\overline{P}_{Z_1}Y(Y'\overline{P}_Z Y)^{-1}$. The log likelihood ratio times -2 is

$$(3.11)\qquad LR_1 = T\sum_{i=1}^{G_0}\log(1 + \lambda_i),$$

where $\lambda_G \geq \cdots \geq \lambda_1 \geq 0$ are the roots of

$$(3.12)\qquad |Y'(P_Z - P_{Z_1})Y - \lambda Y'\overline{P}_Z Y| = 0.$$

The above equation is a sample analogue of (2.14).

The likelihood ratio statistic (3.11) for $G_0 = 1$ was derived by Anderson and Rubin (1949); LR_1 corresponds to the smallest root in the limited information maximum likelihood (LIML) estimation method. When $G_0 = 2$, LR_1 is identical to the statistic proposed by Koopmans and Hood (1953) as the non–identification test. Anderson (1951) was the first to obtain the likelihood ratio criterion (3.11), which he did by differentiation.

3.2. Lagrange multiplier or score test.

The Lagrange Multiplier statistic, which is identical to the Rao score statistic, has been developed as a test statistic to test a hypothesis H about a vector parameter θ in a likelihood L. In these general terms the criterion is

$$LM = \left(\frac{\partial \log L}{\partial \theta} \bigg|_H \right)' \left(- \frac{\partial^2 \log L}{\partial \theta \partial \theta'} \bigg|_H \right)^{-1} \left(\frac{\partial \log L}{\partial \theta} \bigg|_H \right) \tag{3.13}$$

where H denotes the null hypothesis and the value of the parameter in (3.13) maximizes the likelihood under the null hypothesis. If the null hypothesis is

$$H : h(\theta) = 0 \tag{3.14}$$

and λ is vector of Lagrange multipliers

$$\frac{\partial \log L}{\partial \theta} \bigg|_H = -(\lambda|_H)' \frac{\partial h}{\partial \theta'} \bigg|_H . \tag{3.15}$$

Then

$$LM = (\lambda|_H)'(\text{est. asymp cov. of } \lambda)^{-1}(\lambda|_H). \tag{3.16}$$

In our problem $h(\theta) = \operatorname{vec} \Pi_{2 \cdot} B$, $\lambda = \operatorname{vec} \Lambda$, where Λ is $K_2 \times G_0$, and the Lagrange form is

$$\log L_4 = c_1 - \frac{T}{2} \log |\Omega| - \frac{1}{2} \operatorname{tr} \Omega^{-1} (Y - Z\Pi)'(Y - Z\Pi) + \operatorname{tr} \Lambda' \Pi_{2 \cdot} B. \tag{3.17}$$

Setting to 0 the derivative of $\log L_4$ with respect to each element of Π, we obtain

$$Z'(Y - Z\Pi)\Omega^{-1} + \begin{pmatrix} 0 \\ \lambda \end{pmatrix} B' = 0. \tag{3.18}$$

The upper half part of (3.18) gives $Z_1'(Y - Z\widehat{\Pi}) = 0$, and we have

$$\widehat{\Pi}_{1 \cdot} = (Z_1' Z_1)^{-1} Z_1' (Y - Z_2 \widehat{\Pi}_{2 \cdot}). \tag{3.19}$$

Then

$$\begin{aligned} Y - Z\widehat{\Pi} &= Y - Z_1 \widehat{\Pi}_{1 \cdot} - Z_2 \widehat{\Pi}_{2 \cdot} \\ &= \overline{P}_{Z_1} (Y - Z_2 \widehat{\Pi}_{2 \cdot}). \end{aligned} \tag{3.20}$$

Multiplying (3.18) on the right by ΩB, we obtain

$$\begin{pmatrix} 0 \\ \Lambda \end{pmatrix} = -Z'\overline{P}_{Z_1} YB\Sigma^{-1}. \tag{3.21}$$

Using Lemma A.4, we find the first and second derivatives of the log-likelihood function as

$$\frac{\partial \log L_4}{\partial \mathrm{vec}\,\Pi} = \frac{\partial \log L_1}{\partial \mathrm{vec}\,\Pi} + \mathrm{vec}\left[\begin{pmatrix} 0 \\ \Lambda \end{pmatrix} B'\right], \quad \frac{\partial^2 \log L_1}{\partial \mathrm{vec}\,\Pi \partial (\mathrm{vec}\,\Pi)'} = -\Omega^{-1} \otimes Z'Z. \tag{3.22}$$

Then we define an LM statistic by

$$LM_1 = \left(\left.\frac{\partial \log L_1}{\partial \mathrm{vec}\,\Pi}\right|_{\Pi=\widehat{\Pi},\Omega=\widehat{\Omega}}\right)' \left(-\left.\frac{\partial^2 \log L_1}{\partial \mathrm{vec}\,\Pi \partial (\mathrm{vec}\,\Pi)'}\right|_{\Pi=\widehat{\Pi},\Omega=\widehat{\Omega}}\right)^{-1} \left(\left.\frac{\partial \log L_1}{\partial \mathrm{vec}\,\Pi}\right|_{\Pi=\widehat{\Pi},\Omega=\widehat{\Omega}}\right); \tag{3.23}$$

Ω and Π are evaluated at their maximum likelihood estimators under the null hypothesis. Using Lemma A.5, $\Sigma = B'\Omega B$, and

$$\left.\frac{\partial \log L_1}{\partial \mathrm{vec}\,\Pi}\right|_{\Pi=\widehat{\Pi},\Omega=\widehat{\Omega}} = 0, \tag{3.24}$$

we have

$$\begin{aligned} LM_1 &= \left\{\mathrm{vec}\left[\begin{pmatrix} 0 \\ \Lambda \end{pmatrix} B'\right]\right\}' [\Omega \otimes (Z'Z)^{-1}] \mathrm{vec}\left[\begin{pmatrix} 0 \\ \Lambda \end{pmatrix} B'\right] \\ &= \mathrm{tr}\Sigma(0\Lambda')(Z'Z)^{-1}\begin{pmatrix} 0 \\ \Lambda \end{pmatrix}, \end{aligned} \tag{3.25}$$

where the unknown parameters in (3.25) are evaluated at their maximum likelihood estimators under the null hypothesis. (See Engle (1984), for instance.) The LM statistic in the form (3.23) is known as Rao's Score test statistic among statisticians. From (3.18) and $\overline{P}_{Z_1} P_Z \overline{P}_{Z_1} = P_Z - P_{Z_1}$, we further simplify LM_1 as

$$LM_1 = \mathrm{tr}\widehat{B}_H' Y'(P_Z - P_{Z_1}) Y \widehat{B}_H \widehat{\Sigma}_H^{-1}, \tag{3.26}$$

where $\widehat{B}_H$ and $\widehat{\Sigma}_H = (1/T)\widehat{B}_H' Y' \overline{P}_{Z_1} Y \widehat{B}_H$ are the maximum likelihood estimators of B and Σ under the null hypothesis. Let c_i satisfy

$$Y'(P_Z - P_{Z_1})Yc = \lambda_i Y' \bar{P}_Z Y c, \tag{3.27}$$

$$c'Y'\bar{P}_Z Y c = T, \tag{3.27$'$}$$

where λ_i satisfies (3.12), $i = 1, \ldots, G_0$, and let

$$C = (c_1, \ldots, c_{G_0}). \tag{3.27$''$}$$

Then $\widehat{B}_H = CA$ for arbitrary nonsingular A and $\widehat{\Sigma}_H = (1/T)\widehat{B}_H Y'\bar{P}_{Z_1} Y \widehat{B}_H$. (See Appendix B). When we use the roots of (3.12), this statistic is expressed as

$$LM_1 = T \sum_{i=1}^{G_0} \frac{\lambda_i}{1+\lambda_i}. \tag{3.28}$$

When $G_0 = 1$, this statistic LM_1 is the LM statistic proposed by Byron (1972). However, his derivation of the statistic is different from ours. It should be also noted that (3.28) is an analogue of the Bartlett-Nanda-Pillai trace criterion, which is well known in multivariate statistical analysis. (See Anderson (1984), Chapter 8.) Our derivation yields a new interpretation of the Bartlett–Nanda–Pillai test.

3.3. Wald test.

In general terms the Wald test is based on the statistic

$$h(\widehat{\theta})'[\text{asymp. cov. of } h(\widehat{\theta})]^{-1} h(\widehat{\theta}), \tag{3.29}$$

where $\widehat{\theta}$ is the maximum likelihood estimator of the parameter vector θ under the alternative hypothesis. In our problem the null hypothesis is that the rank of $\Pi_{2\cdot}$ is $G - G_0 = G_*$. To express this in the form of $h(\theta) = 0$ we partition $\Pi_{2\cdot}$ into $L = K_2 - G_*$ and G_* rows and G_0 and G_* columns:

$$\Pi_{2\cdot} = \begin{pmatrix} \Pi_{\ell 0} & \Pi_{\ell *} \\ \Pi_{m0} & \Pi_{m*} \end{pmatrix}. \tag{3.30}$$

Since $\Pi_{2\cdot}$ is of rank G_*, there is at least one square matrix Π_{m*} of order G_* that is nonsingular. There exists a $G \times G_0$ matrix

$$B = \begin{pmatrix} B_0 \\ -B_* \end{pmatrix}, \tag{3.31}$$

where B_0 is $G_0 \times G_0$, such that $\xi = \Pi_{2\cdot} B = 0$.

This equation can be partitioned as

(3.32) $$\xi_\ell = \Pi_{\ell 0} B_0 - \Pi_{\ell *} B_* = 0,$$

(3.33) $$\xi_m = \Pi_{m0} B_0 - \Pi_{m*} B_* = 0.$$

The second equation yields

(3.34) $$B_* = \Pi_{m*}^{-1} \Pi_{m0} B_0.$$

Substitution into (3.32) yields

(3.35) $$\xi_\ell = \left(\Pi_{\ell 0} - \Pi_{\ell *} \Pi_{m*}^{-1} \Pi_{m0} \right) B_0 = 0.$$

Since B is of rank G_0, B_0 is also of rank G_0. (If B_0 were of lower rank, there would exist a vector c such that $B_0 c = 0$; then by (3.34) $Bc = 0$.) Hence, if $\Pi_{2\cdot}$ is of rank G_*, $\Pi_{\ell 0} - \Pi_{\ell *} \Pi_{m*}^{-1} \Pi_{m0} = 0$. We take

(3.36) $$h(\theta) = \operatorname{vec} \left(\Pi_{\ell 0} - \Pi_{\ell *} \Pi_{m*}^{-1} \Pi_{m0} \right).$$

Let $Y = (Y_0, Y_*)$, $V = (V_0, V_*)$, $\Pi_{1\cdot} = (\Pi_{10}, \Pi_{1*})$, and $Z_2 = (Z_\ell, Z_m)$. The reduced form is

(3.37) $$Y_0 = Z_1 \Pi_{10} + Z_\ell \Pi_{\ell 0} + Z_m \Pi_{m0} + V_0,$$

(3.38) $$Y_* = Z_1 \Pi_{1*} + Z_\ell \Pi_{\ell *} + Z_m \Pi_{m*} + V_*.$$

A just-identified set of structural equations is

(3.39) $$Y_0 = Y_* B_* + Z_1 \Gamma + Z_\ell \xi_\ell + U,$$

where B_* is given by (3.34) and $B_0 = I$. Then a set of parameters is $B_*, \Gamma, \xi_\ell, \Pi_{1*}, \Pi_{\ell *}, \Pi_{m*}$, and Ω; the null hypothesis is defined by $\xi_\ell = 0$. The maximum likelihood estimator (under the alternative) of B_*, Γ, and ξ_ℓ is the indirect least squares estimator, which is the solution to

$$(3.40) \qquad \begin{pmatrix} \widehat{Y}'_* \\ Z'_1 \\ Z'_\ell \end{pmatrix} (\widehat{Y}_*, Z_1, Z_\ell) \begin{pmatrix} \widehat{B}_{*I} \\ \widehat{\Gamma} \\ \widehat{\xi}_\ell \end{pmatrix} = \begin{pmatrix} \widehat{Y}'_* \\ Z'_1 \\ Z'_\ell \end{pmatrix} Y_0,$$

where $\widehat{Y}_* = Z(Z'Z)^{-1}Z'Y_*$. Then solving the above equation with respect to $\widehat{\xi}_\ell$ we have

$$(3.41) \qquad Z'_\ell \overline{P}_{\widehat{Y}_*, Z_1} Z_\ell \widehat{\xi}_\ell = Z'_\ell \overline{P}_{\widehat{Y}_*, Z_1} Y_0.$$

Applying Lemma A.6 for $\overline{P}_{\widehat{Y}_*, Z_1}$ and noting that $Z_\ell = P_Z Z_\ell$, we write the estimator of ξ_ℓ as

$$(3.42) \qquad \widehat{\xi}_\ell = (Z'_\ell N Z_\ell)^{-1} Z'_\ell N Y_0,$$

where

$$N = \overline{P}_{Z_1} - \overline{P}_Z - (\overline{P}_{Z_1} - \overline{P}_Z) Y_* \left[Y'_* (\overline{P}_{Z_1} - \overline{P}_Z) Y_* \right]^{-1} Y'_* (\overline{P}_{Z_1} - \overline{P}_Z).$$

In the above we have utilized the fact that $Z'_\ell N Z_\ell$ is nonsingular because the matrix (Y_*, Z_1, Z_ℓ) is of full rank (a.s.) and rank $(N) =$ rank $(Z_\ell) = L$ (a.s.). Then the covariance matrix of the limiting normal distribution of $\sqrt{T}\,\mathrm{vec}\,(\widehat{\xi}_\ell - \xi_\ell)$ is the probability limit of

$$(3.43) \qquad \begin{aligned} &T\left[I_{G_0} \otimes (Z'_\ell N Z_\ell)^{-1} Z'_\ell N\right] (\Sigma \otimes I_T) \left[I_{G_0} \otimes (Z'_\ell N Z_\ell)^{-1} Z'_\ell N\right]' \\ &= \Sigma \otimes \left(\frac{1}{T} Z'_\ell N Z_\ell \right)^{-1}. \end{aligned}$$

We now define a Wald-type statistic by

$$(3.44) \qquad W_1 = \left(\mathrm{vec}\,\widehat{\xi}_\ell\right)' \left[\widehat{\Sigma}_I \otimes (Z'_\ell N Z_\ell)^{-1}\right]^{-1} (\mathrm{vec}\,\widehat{\xi}_\ell),$$

where $\widehat{\Sigma}_I = \widehat{B}'_I \widehat{\Omega} \widehat{B}_I$, $\widehat{B}'_I = (I_{G_0}, -\widehat{B}'_{*I})$, and $\widehat{\Omega} = (1/T) Y' \overline{P}_Z Y$. Since we have normalized $B_1 = I_{G_1}$, the two-stage least squares estimator of B under $H_\xi : \xi = 0$ is $\widehat{B}'_{TS} = (I_{G_0}, -\tilde{B}'_*)$

and $\tilde{B}_*$ satisfies $[Y_*'(P_Z - P_{Z_1})Y_*]\tilde{B}_* = Y_*'(P_Z - P_{Z_1})Y_0$. Thus using Lemma A.6 again, we have

$$
\begin{aligned}
(3.45)\; W_1 &= \{\text{vec}\,[(Z_\ell' N Z_\ell)^{-1} Z_\ell' N Y_0]\}'\,[\widehat{\Sigma} \otimes (Z_\ell' N Z_\ell)^{-1}]^{-1} \text{vec}\,[(Z_\ell' N Z_\ell)^{-1} Z_\ell' N Y_0] \\
&= [\text{vec}(Z_\ell' N Y_0)]'\,[I \otimes (Z_\ell' N Z_\ell)^{-1}]\,[\widehat{\Sigma}^{-1} \otimes (Z_\ell' N Z_\ell)]\,[I \otimes (Z_\ell' N Z_\ell)^{-1}]\text{vec}(Z_\ell' N Y_0) \\
&= [\text{vec}(Z_\ell' N Y_0)]'\,[\widehat{\Sigma}^{-1} \otimes (Z_\ell' N Z_\ell)^{-1}]\text{vec}(Z_\ell' N Y_0) \\
&= \text{tr}\{\widehat{\Sigma}^{-1} Y_0' N Z_\ell (Z_\ell' N Z_\ell)^{-1} Z_\ell' N Y_0\}.
\end{aligned}
$$

Since $N^2 = N$, there exists a $T \times L$ matrix X such that $N = X(X'X)^{-1}X'$ and

$$
\begin{aligned}
(3.46) \qquad & N Z_\ell (Z_\ell' N Z_\ell)^{-1} Z_\ell' N \\
&= X(X'X)^{-1} X' Z_\ell [Z_\ell' X (X'X)^{-1} X' Z_\ell]^{-1} Z_\ell' X (X'X)^{-1} X' \\
&= X(X'X)^{-1}X' = N.
\end{aligned}
$$

Then we have

$$
\begin{aligned}
(3.47) \qquad W_1 &= \text{tr}\,\{\widehat{\Sigma}^{-1} Y_0' N Y_0\} \\
&= \text{tr}\,\{\widehat{\Sigma}^{-1}(Y_0 - Y_*\widetilde{B}_*)'(P_Z - P_{Z_1})(Y_0 - Y_*\widetilde{B}_*)\} \\
&= \text{tr}\,\{\widehat{\Sigma}^{-1}\widetilde{B}_{TS}' Y'(P_Z - P_{Z_1}) Y \widetilde{B}_{TS}\}.
\end{aligned}
$$

The above derivation is an extension of Hwang (1980b). The last expression in (3.47) shows that except for $\widehat{\Sigma}$ the criterion does not depend on the selection of variables to define ξ_ℓ and ξ_m, but it does depend on the selection of variables to define Y_0 and Y_1. The criterion (3.47) can be modified by defining $\widehat{\Sigma}$ as $\widehat{B}'\widehat{\Omega}\widehat{B}$ with $\widehat{B}$ as some other estimate of B. The estimate of B obtained when $\xi_m = 0$ (that is, the alternative hypothesis holds) is $(I, -\widehat{B}_*')'$, where $\widehat{B}_*$ is defined by (3.40). If $\widehat{B} = \widehat{B}_{TS}$, then (3.47) is completely independent of the selection of variables in ξ_ℓ and ξ_m. This is the statistic derived by Wegge (1978) for $G_0 = 1$. Hwang (1980b) has shown that it is identical to the Wald statistic proposed by Byron (1974). If we use the maximum likelihood estimator of Σ under the null hypothesis,

$$
(3.48) \qquad \widehat{\Sigma} = \frac{1}{T}\widehat{B}' Y' \overline{P}_{Z_1} Y \widehat{B},
$$

the resulting statistic reduces to the statistic proposed by Basmann (1960) for the case of $G_0 = 1$. It can be interpreted as a Wald–type statistic in the present context.

If $\widehat{B} = \widehat{B}_{LI}$, the limited information maximum likelihood estimator under $H_\xi : \xi = 0$, and $\widehat{B}_{TS}$ in (3.47) is replaced by $\widehat{B}_{LI}$, the statistic is

$$
\begin{aligned}
(3.49) \qquad W_1' &= T\mathrm{tr}(\widehat{B}_{LI}' Y' \overline{P}_{Z_1} Y \widehat{B}_{LI})^{-1} \widehat{B}_{LI} Y' (P_Z - P_{Z_1}) Y \widehat{B}_{LI} \\
&= T \sum_{i=1}^{G_0} \lambda_i,
\end{aligned}
$$

where $\lambda_1, \ldots, \lambda_{G_0}$ are the G_0 smallest roots of (3.12).

It should be also noted that W_1' is an analogue (or generalization) of the Lawley–Hotelling Trace Criterion, which is well–known in multivariate statistical analysis. (See Anderson (1984), Chapter 8.) Thus our derivation also gives a new interpretation to the Lawley–Hotelling type statistic.

3.4 An Inequality Among Statistics.

We have derived three types of statistics for the block identifying restriction in a subsystem of structural equations. There is a simple inequality among the statistics we have derived. Using Lemma A.7, we have

$$
(3.50) \qquad 0 \leq LM_1 \leq LR_1 \leq W_1'.
$$

This type of inequality among three different types of statistics has been well–known for testing linear restrictions in the multivariate regression model (Anderson (1984), Chapter 8, for instance.) If we use the same significance point (based on the asymptotic χ^2 distribution), the Wald–type statistic tends to reject the hypothesis more often than the other statistics while the likelihood ratio statistic tends to reject the hypothesis more frequently than the LM statistics.

4. Tests of Predeterminedness

In this section we shall derive several tests of the null hypothesis of econometric predeterminedness $H_{\xi,\eta} : \xi = 0, \eta = 0$. We suppose that $G_0 \leq G_1$.

4.1. The likelihood ratio test.

We first find the likelihood function maximized under $H_{\xi,\eta}$. The hypothesis H_η is

$$
\begin{aligned}
(4.1)\qquad 0 &= (\Omega_{21}, \Omega_{22})\begin{pmatrix} B_1 \\ -B_2 \end{pmatrix} \\
&= \Omega_{21}B_1 - \Omega_{22}B_2 \\
&= \Omega_{22}(\rho B_1 - B_2),
\end{aligned}
$$

where $\rho = \Omega_{22}^{-1}\Omega_{21}$ and B' has been partitioned as $B' = (B_1', -B_2')$ with B_1 having G_1 rows. This fact suggests a re–parametrization since $B_2 = \rho B_1$ under H_η and then H_ξ is

$$
(4.2)\qquad 0 = \Pi_{21}B_1 - \Pi_{22}B_2 = (\Pi_{21} - \Pi_{22}\rho)B_1,
$$

where Π_{ij} denotes the i,j-th submatrix of Π partitioned into K_1 and K_2 rows and G_1 and G_2 columns. Let

$$
(4.3)\qquad Y_1^* = Y_1 - Y_2\rho, \quad V_1^* = V_1 - V_2\rho,
$$

$$
\begin{aligned}
(4.4)\qquad \Pi^{**} = (\Pi_{\cdot 1}^{**}, \Pi_{\cdot 2}^{**}) = \begin{pmatrix} \Pi_{11}^{**} & \Pi_{12}^{**} \\ \Pi_{21}^{**} & \Pi_{22}^{**} \end{pmatrix} &= \begin{pmatrix} \Pi_{11} & \Pi_{12} \\ \Pi_{21} & \Pi_{22} \end{pmatrix}\begin{pmatrix} I & 0 \\ -\rho & I \end{pmatrix} \\
&= \begin{pmatrix} \Pi_{11} - \Pi_{12}\rho & \Pi_{12} \\ \Pi_{21} - \Pi_{22}\rho & \Pi_{22} \end{pmatrix}.
\end{aligned}
$$

Then H_ξ is $\Pi_{21}^{**}B_1 = 0$. The reduced form for (Y_1^*, Y_2) is

$$
(4.5)\qquad (Y_1^*, Y_2) = Z\Pi^{**} + (V_1^*, V_2).
$$

The covariance matrix of each row of (V_1^*, V_2) is

$$
(4.6)\qquad \begin{pmatrix} I & -\rho' \\ 0 & I \end{pmatrix}\begin{pmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{pmatrix}\begin{pmatrix} I & 0 \\ -\rho & I \end{pmatrix} = \begin{pmatrix} \Omega_{11\cdot 2} & 0 \\ 0 & \Omega_{22} \end{pmatrix},
$$

where $\Omega_{11\cdot 2} = \Omega_{11} - \Omega_{12}\Omega_{22}^{-1}\Omega_{21}$. The likelihood function of the parameters $\Omega_{22}, \Omega_{11\cdot 2}, \Pi^{**}$, and ρ (subject to (4.2) or alternatively $\Pi_{21}^{**}B_1 = 0$) is

$$
\begin{aligned}
(4.7)\qquad L_5 = c_1|\Omega_{11\cdot 2}|^{-T/2} &\exp\left[-\frac{1}{2}\mathrm{tr}\Omega_{11\cdot 2}^{-1}(Y_1^* - Z\Pi_{\cdot 1}^{**})'(Y_1^* - Z\Pi_{\cdot 1}^{**})\right] \\
\times|\Omega_{22}|^{-T/2} &\exp\left[-\frac{1}{2}\mathrm{tr}\Omega_{22}^{-1}(Y_2 - Z\Pi_{\cdot 2})'(Y_2 - Z\Pi_{\cdot 2})\right].
\end{aligned}
$$

The function L_5 is maximized with respect to $\Pi_{\cdot 2}, \Omega_{22}$, and $\Omega_{11\cdot 2}$ at

$$\widehat{\Pi}_{\cdot 2} = (Z'Z)^{-1}Z'Y_2, \quad T\widehat{\Omega}_{22} = Y_2'\overline{P}_Z Y_2, \tag{4.8}$$

$$\begin{aligned} T\widehat{\Omega}_{11\cdot 2} &= (Y_1^* - Z\Pi_{\cdot 1}^{**})'(Y_1^* - Z\Pi_{\cdot 1}^{**}) \\ &= \left[Y_1 - (Y_2, Z_1)\begin{pmatrix}\rho \\ \Pi_{11}^{**}\end{pmatrix} - Z_2\Pi_{21}^{**}\right]' \left[Y_1 - (Y_2, Z_1)\begin{pmatrix}\rho \\ \Pi_{11}^{**}\end{pmatrix} - Z_2\Pi_{21}^{**}\right], \end{aligned} \tag{4.9}$$

and the concentrated likelihood is

$$L_6 = c_2|T\widehat{\Omega}_{22}|^{-T/2}|T\widehat{\Omega}_{11\cdot 2}|^{-T/2}. \tag{4.10}$$

Since $T\widehat{\Omega}_{22}$ is a sample quantity, we want to minimize $|T\widehat{\Omega}_{11\cdot 2}|$ with respect to ρ, Π_{11}^{**}, and Π_{21}^{**} subject to $\Pi_{21}^{**}B_1 = 0$.

The general alternative to $H_{\xi,\eta}$ is that Π and Ω are unrestricted; we term this as H_A. The problem of testing $H_{\xi,\eta}$ vs H_A has been reduced to testing (4.2) $\Pi_{21}^{**}B_1 = 0$ vs $\Pi_{21}^{**}B_1 \neq 0$. The likelihood ratio criterion by the algebra of Section 3.1 is

$$LR_2 = T\sum_{i=1}^{G_0} \log(1 + \lambda_i^*), \tag{4.11}$$

where $\lambda_{G_1}^* \geq \cdots \geq \lambda_1^* \geq 0$ are the roots of

$$|Y_1'(P_{Y_2,Z} - P_{Y_2,Z_1})Y_1 - \lambda^* Y_1'\overline{P}_{Y_2,Z}Y_1| = 0. \tag{4.12}$$

Another possible alternative to $H_{\xi,\eta}$ may be H_ξ, which defines the structural equations with the block identifiability restrictions. Because $H_{\xi,\eta}$ is nested within H_ξ, the log-likelihood ratio criterion for $H_{\xi,\eta}$ vs H_ξ is the difference between the statistic for $H_{\xi,\eta}$ vs H_A and the statistic for H_ξ vs H_A, namely

$$\begin{aligned} LR_3 &= T\sum_{i=1}^{G_0} \log(1 + \lambda_i^*) - T\sum_{i=1}^{G_0} \log(1 + \lambda_i) \\ &= T\sum_{i=1}^{G_0} \log\frac{1 + \lambda_i^*}{1 + \lambda_i}. \end{aligned} \tag{4.13}$$

When $G_0 = 1$ and $G_1 \geq 1, LR_3$ reduces to the statistic obtained by Hwang (1980a). Furthermore, LR_3 reduces to the statistic obtained by Kariya and Hodoshima (1980) when $G_0 = G_1 = 1$.

4.2. Lagrange multiplier test.

The Lagrange multiplier statistic for testing $H_{\xi,\eta} : \xi = 0, \eta = 0$ vs. H_A is the Lagrange multiplier statistic for testing $H_\zeta : \zeta = \Pi_{21}^{**} B_1 = 0$ vs. H_A. It is

$$LM_2 = T \sum_{i=1}^{G_0} \frac{\lambda_i^*}{1+\lambda_i^*}. \tag{4.14}$$

where $\lambda_1^*, \ldots, \lambda_{G_0}^*$ are the G_0 smallest roots of (4.12). This statistic LM_2 does not seem to have been derived previously.

Now consider testing $H_{\xi,\eta}$ vs. H_ξ. Let Λ and Λ_0 be $K_2 \times G_0$ and $G_2 \times G_0$ matrices of Lagrange multiplier parameters for $H_{\xi,\eta} : \xi = 0,\ \eta = 0$, respectively. The Lagrange form in this case is written as

$$\begin{aligned}\log L_7 = \log c_1 - \frac{T}{2} \log |\Omega| - \frac{1}{2} \mathrm{tr}\, \Omega^{-1} (Y - Z\Pi)'(Y - Z\Pi) + \mathrm{tr}\, \Lambda'(\Pi_{21}, \Pi_{22}) B \\ + \mathrm{tr}\, \Lambda_0'(\rho, I_{G_2}) B.\end{aligned} \tag{4.15}$$

Setting to 0 the derivative of log L_7 with respect to the components of B_2, we have

$$\Pi_{22}' \Lambda + \Lambda_0 = 0. \tag{4.16}$$

Substituting this relation into log L_7 and ignoring a constant term we obtain

$$\begin{aligned}\log L_8 = -\frac{T}{2} \log |\Omega_{11\cdot 2}||\Omega_{22}| - \frac{1}{2} \mathrm{tr}\, \Omega^{-1} (Y - Z\Pi)'(Y - Z\Pi) \\ + \mathrm{tr}\, \Lambda' \big(\Pi_{21} - \Pi_{22} \rho \big) B_1,\end{aligned} \tag{4.17}$$

and H_ξ is $(\Pi_{21} - \Pi_{22}\rho) B_1 = 0$ under $H\eta$.

The derivatives with respect to the elements of Π are

$$Z'Y\Omega^{-1} - Z'Z\Pi\Omega^{-1} + \begin{pmatrix} 0 & 0 \\ \Lambda B_1' & -\Lambda B_1' \rho' \end{pmatrix}. \tag{4.18}$$

Setting this matrix to 0 yields

$$Z'Z\Pi = Z'Y + \begin{pmatrix} 0 & 0 \\ \Lambda B_1'\Omega_{11\cdot 2} & 0 \end{pmatrix}. \tag{4.19}$$

We can write

$$\Omega^{-1} = \begin{pmatrix} \Omega_{11\cdot 2}^{-1} & -\Omega_{11\cdot 2}^{-1}\rho' \\ -\rho\Omega_{11\cdot 2}^{-1} & \rho\Omega_{11\cdot 2}^{-1}\rho' + \Omega_{22}^{-1} \end{pmatrix}. \tag{4.20}$$

The derivatives of $\log L_9$ with respect to the elements of ρ are

$$(Y_2 - Z_1\Pi_{12} - Z_2\Pi_{22})'(Y - Z\Pi)\begin{pmatrix} I_{G_1} \\ -\rho \end{pmatrix}\Omega_{11\cdot 2}^{-1} - \Pi_{22}'\Lambda B_1'. \tag{4.21}$$

When (4.21) is set to 0 and multiplied by $\Omega_{11\cdot 2}$ we obtain

$$Y_2'\overline{P}_Z Y_1 - Y_2'\overline{P}_Z Y_2\rho = \Pi_{22}'\Lambda B_1'\Omega_{11\cdot 2} \tag{4.22}$$

by use of (4.19). From the first set of columns of (4.19) we obtain

$$\Lambda B_1'\Omega_{11\cdot 2} = Z_2'\overline{P}_{Z_1}(Z_2\Pi_{21} - Y_1). \tag{4.23}$$

When we use this in (4.22), we obtain

$$Y_2'\overline{P}_Z Y\begin{pmatrix} I_{G_1} \\ -\widehat{\rho} \end{pmatrix} = \Pi_{22}' Z_2'\overline{P}_{Z_1}(Z_2\Pi_{21} - Y_1). \tag{4.24}$$

Multiply (4.24) on the right by B_1 and replace $\Pi_{21}B_1$ by $\Pi_{22}\rho B_1$. Then solve (4.24) multiplied by B_1 with respect to ρB_1 using (4.19) to obtain

$$\rho B_1 = (Y_2'\overline{P}_{Z_1}Y_2)^{-1} Y_2'\overline{P}_{Z_1}Y_1 B_1. \tag{4.25}$$

Lemma A.6 has been used. Multiply (4.23) on the right by B_1 using (4.25) to obtain

$$\begin{aligned} \lambda\Sigma &= Z_2'\overline{P}_{Z_1}(Z_2\Pi_{22}\rho - Y_1)B_1 \\ &= -Z_2'\overline{P}_{Y_2,Z_1}Y_1 B_1, \end{aligned} \tag{4.26}$$

where $\Sigma = B_1'\Omega_{11\cdot 2}B_1$. In this derivation we have used Lemma A.6 for (Y_2, Z_1).

From (4.26) we find that the Lagrange multiplier matrix in the sample for $H_\eta : \eta = 0$ is

$$
\begin{aligned}
\Lambda_0 &= -\Pi_{22}'\Lambda \\
&= -Y_2'\overline{P}_{Z_1}Z_2\left(Z_2'\overline{P}_{Z_1}Z_2\right)^{-1}Z_2'\overline{P}_{Y_2,Z_1}Y_1\widehat{B}_1\widehat{\Sigma}^{-1} \\
&= -Y_2'(\overline{P}_{Z_1} - \overline{P}_Z)\overline{P}_{Y_2,Z_1}Y_1\widehat{B}_1\widehat{\Sigma}^{-1} \\
&= -Y_2'P_Z\overline{P}_{Y_2,Z_1}Y_1\widehat{B}_1\widehat{\Sigma}^{-1},
\end{aligned}
\tag{4.27}
$$

where $\widehat{\Sigma}$ and $\widehat{B}_1$ are the maximum likelihood estimators of Σ and B_1. Since $\widehat{\Sigma}$ is a consistent estimator of Σ and $(1/T)\Lambda_0 \xrightarrow{\text{p}} 0$, we consider the quantity

$$
\begin{aligned}
\Lambda_0^* &= Y_2'P_Z\overline{P}_{Y_2,Z_1}Y_1\widehat{B}_1\Sigma^{-1} \\
&= Y_2'P_Z\overline{P}_{Y_2,Z_1}Y_1B_1\Sigma^{-1} + Y_2'P_Z\overline{P}_{Y_2,Z_1}Y_1(\widehat{B}_1 - B_1)\Sigma^{-1}
\end{aligned}
\tag{4.28}
$$

which is asymptotically equivalent to Λ_0. We note that $\overline{P}_{Y_2,Z_1}Y_1B_1 = \overline{P}_{Y_2,Z_1}U$ under $H_\eta : \eta = 0$. Now we apply the method to derive the asymptotic distribution in Lemma 3 in Anderson and Kunitomo (1989b) and substitute (Y_2, Z) and (Y_2, Z_1) for Z and Z_1, respectively. Let

$$
\begin{aligned}
R^* &= [(Z\Pi_{\cdot 1} + V_2\rho)J_1, Y_2, Z_1] \\
&= \left[(Y_2, Z)\begin{pmatrix} \rho \\ \Pi_{\cdot 1} - \Pi_{\cdot 2}\rho \end{pmatrix} J_1, Y_2, Z_1\right] \\
&= RF,
\end{aligned}
\tag{4.29}
$$

where $R = (Y_2, Z)$, $J_1' = (0, I_{G_1 - G_0})$ is a $(G_1 - G_0) \times G_1$ matrix and F is a $(G_2 + K) \times (G_* + K_1)$ matrix defined by

$$
F = \left[\begin{pmatrix} \rho \\ \Pi_{\cdot 1} - \Pi_{\cdot 2}\rho \end{pmatrix}\begin{pmatrix} 0 \\ I_{G_1 - G_0} \end{pmatrix}, \begin{pmatrix} I_{G_2 + K_1} \\ 0 \end{pmatrix}\right].
\tag{4.30}
$$

Let normalize $B_0 = I_{G_0}$ and partition $Y_1 = (Y_0, Y_{11})$ into $T \times (G_0 + (G_1 - G_0))$ submatrices. Then from (4.29) we write

$$
(Y_{11}, Y_2, Z_1) = R^* + (V_1^* J_1, 0).
\tag{4.31}
$$

Under the hypothesis $H_\eta : \eta = 0$ we have

$$U = V_1 B_1 - V_2 B_2 \tag{4.32}$$
$$= (V_1 - V_2\rho)B_1 = V_1^* B_1.$$

Since each row of R^* and R in (4.29) is uncorrleated with each row of $V_1^* = V_1 - V_2\rho$, R^* is uncorrelated with U. Then by using the same argument as in the proof of Theorem 5 in Anderson and Kunitomo (1989b), the asymptotic distribution of $\mathrm{vec}(\Lambda_0^*)$ is equivalent to the asymptotic distribution of

$$\mathrm{vec}\big(Y_2' P_Z \overline{P}_{R^*} U \Sigma^{-1}\big). \tag{4.33}$$

We write (4.33) as

$$\mathrm{vec}\big(Y_2' P_Z \overline{P}_{RF} U \Sigma^{-1}\big) = \big(\Sigma^{-1} \otimes Y_2' P_Z \overline{P}_{RF}\big)\mathrm{vec}(U). \tag{4.34}$$

Then conditional on Y_2 the covariance matrix of $\mathrm{vec}(\Lambda_0^*)$ is

$$(\Sigma^{-1} \otimes Y_2' P_Z \overline{P}_{RF})(\Sigma \otimes I_T)(\Sigma^{-1} \otimes \overline{P}_{RF} P_Z Y_2) = \Sigma^{-1} \otimes Y_2' P_Z \overline{P}_{RF} P_Z Y_2. \tag{4.35}$$

We now define an LM statistic by

$$LM_3 = (\mathrm{vec}\Lambda_0)'(\widehat{\Sigma}^{-1} \otimes Y_2' P_Z \overline{P}_{R\widehat{F}} P_Z Y_2)^{-1}(\mathrm{vec}\Lambda_0). \tag{4.36}$$

Then by the use of Lemma A.5, we rewrite (4.36) as,

$$LM_3 = \mathrm{tr}\big\{\widehat{B}_1' Y_1' \overline{P}_{Y_2,Z_1} \overline{P}_Z Y_2 (Y_2' \overline{P}_Z \overline{P}_{R\widehat{F}} \overline{P}_Z Y_2)^{-1} Y_2' \overline{P}_Z \overline{P}_{Y_2,Z_1} Y_1 \widehat{B}_1 \widehat{\Sigma}^{-1}\big\}, \tag{4.37}$$

where we have used the relation $\overline{P}_{Y_2,Z_1}\overline{P}_Z Y_2 = -\overline{P}_{Y_2,Z_1} P_Z Y_2$ and F is evaluated at its maximum likelihood estimator.

When $G_1 = G_0$, we have $\overline{P}_{RF} = \overline{P}_{Y_2,Z_1}$. Using Lemma A.6, we obtain the expression

$$LM_3 = \mathrm{tr}\big\{Y_1'(\overline{P}_{Y_2,Z_1} - \overline{P}_X) Y_1 \widehat{\Sigma}^{-1}\big\} \tag{4.38}$$
$$= T \sum_{i=1}^{G_0} \frac{\lambda_i^{**}}{1 + \lambda_i^{**}},$$

where λ_i^{**} are the characteristic roots of

$$|Y_1'(P_X - P_{Y_2,Z_1})Y_1 - \lambda^{**}Y_1'\overline{P}_X Y_1| = 0, \tag{4.39}$$

and $X = (Y_2, Z_1, \overline{P}_Z Y_2)$ is a $T \times (G_2 + K_1 + G_2)$ matrix. The second line of (4.38) implies that $\widehat{\Sigma} = (1/T)\widehat{B}_1' Y_1' \overline{P}_X Y_1 \widehat{B}_1$. In the present formulation of the LM test, $\widehat{\Sigma}$ should be based on the maximum likelihood estimator of Σ under the null hypothesis:

$$\widehat{\Sigma} = \widehat{\Omega}_{11\cdot 2} = \frac{1}{T} Y_1' \overline{P}_{Y_2,Z_1} Y_1. \tag{4.40}$$

However, in practice, several estimators of Σ could be used. For instance, instead of (4.40), we may use

$$\widehat{\Sigma} = \frac{1}{T - 2G_2 - K_1} Y_1' \overline{P}_X Y_1. \tag{4.41}$$

In particular, LM_3 with (4.41) reduces to the statistic proposed by Wu (1973) and Wu (1974) when $G_0 = G_1 = 1$.

On the other hand, consider a testing problem for

$$Y_1 B_1 = Y_2 B_2 + Z_1 \Gamma + E_3 B_3 + U, \tag{4.42}$$

where B_3 is a $G_2 \times G_0$ vector of unknown parameters, and E_3 is the least squares residuals $E_3 = Y_2 - \widehat{Y}_2 = \overline{P}_Z Y_2$. Hausman (1978) proposed the usual F test for $H_0 : B_3 = 0$ against $H_1 : B_3 \neq 0$ as a specification test when $G_0 = G_1 = 1$ and $B_0 = 1$. From (4.37) it is clear that LM_3 is proportional to Hausman's statistic in this case. In fact, Nakamura and Nakamura (1980) has shown this equivalence between Wu's test and Hausman's test for $G_1 = 1$. They also pointed out that a statistic proposed by Durbin (1954) is similar to them. Our derivation of statistics shows that these statistics can be interpreted as LM test procedures. Hwang (1985) also has shown the equivalence of Hausman's test and an LM test for $G_0 = 1$ and $G_1 \geq 1$ by a different method.

Another possibility of an estimator of Σ may be

$$\widehat{\Sigma} = \frac{1}{T - K - G_2} Y_1' \overline{P}_{Y_2,Z} Y_1 \tag{4.43}$$

because it is an unrestricted sum of squares from the regression residuals. Then, the statistic LM_3 with (4.43) reduces to the one proposed by Revankar (1978) when $G_0 = G_1 = 1$. Therefore, we can also reinterpret Revankar's test as an LM test procedure.

4.3. Wald test.

Now we consider Wald-type statistics for the present testing problems. For this purpose we first consider the null hypothesis $H_{\xi,\eta} : \xi = 0,\ \eta = 0$ vs the alternative hypothesis $H_A : \xi \neq 0$. In this case, our derivation of a Wald test is similar to Section 3.3. Thus a Wald-type statistic is

$$W_2 = T \sum_{i=1}^{G_0} \lambda_i^*, \tag{4.44}$$

where λ_i^* are the characteristic roots of (4.44).

When $G_0 = G_1 = 1$, W_2 reduces to the statistic proposed by Revankar and Hartley (1973). Although their derivation was different from ours, we can interpret their statistic as a Wald test for $H_{\xi,\eta}$ against H_A. W_2 may be called the generalized Revankar-Hartley test.

We now derive a Wald-type statistic for the null hypothesis $H_{\xi,\eta} : \xi = 0,\ \eta = 0$ against the alternative $H_\xi : \xi = 0$. We note that from (3.18) under H_ξ

$$\begin{aligned} T\widehat{\Omega} &= (Y - Z\widehat{\Pi})'(Y - Z\widehat{\Pi}) \\ &= Y'\overline{P}_Z Y + \widehat{\Omega}\widehat{B}\begin{pmatrix} 0 \\ \Lambda \end{pmatrix}'(Z'Z)^{-1}\begin{pmatrix} 0 \\ \Lambda \end{pmatrix}\widehat{B}'\widehat{\Omega}. \end{aligned} \tag{4.45}$$

Using (3.21), we have

$$T\widehat{\Omega}\widehat{B} = Y'\overline{P}_Z Y\widehat{B} + \widehat{\Omega}\widehat{B}\widehat{\Sigma}^{-1}\widehat{B}'Y'(P_Z - P_{Z_1})Y\widehat{B}. \tag{4.46}$$

Because $T\widehat{\Sigma}_H = \widehat{B}_H Y'\bar{P}_{Z_1} Y\widehat{B}_H$, where $\widehat{B}_H = CC_0^{-1}$, we obtain an unrestricted estimator of η as

$$\begin{aligned} \hat{\eta} &= J_2'\widehat{\Omega}_H\widehat{B}_H \\ &= J_2'\frac{1}{T}Y'\bar{P}_Z Y\widehat{B}_H C_0(I + \widehat{\Lambda})C_0^{-1} \\ &= J_2'\frac{1}{T}Y'\bar{P}_Z Y\widehat{B} + J_2'\frac{1}{T}Y'\bar{P}_Z Y C_0\widehat{\Lambda}C_0^{-1}, \end{aligned} \tag{4.47}$$

where $J_2' = (0, I_{G_2})$ and $\widehat{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_{G_0})$. Since $\lambda_i = o_p(1/T^{1-\varepsilon})$ for any $\varepsilon > 0$, $\sqrt{T} C_0 \widehat{\Lambda} C_0^{-1} \xrightarrow{\mathrm{p}} 0$. (See Anderson and Kunitomo (1989b).) Then the limiting distribution of $\sqrt{T}(\hat{\eta} - \eta)$ is the same as the limiting distribution of

$$\text{(4.49)} \qquad \mathrm{vec}(J_2' \Omega \sqrt{T}(\widehat{B} - B)) = \mathrm{vec}\big[(\Omega_{21}, \Omega_{22}) J_*, 0\big] \sqrt{T} \begin{bmatrix} -(\widehat{B}_* - B_*) \\ -(\widehat{\Gamma} - \Gamma) \end{bmatrix}$$

is the limiting distribution of

$$\text{(4.50)} \qquad -I_{G_0} \otimes \big[(\Omega_{21}, \Omega_{22}) J_*, 0\big] (D' M D)^{-1} D' \frac{1}{\sqrt{T}} \mathrm{vec}\,(Z' U),$$

where $J_*' = (0, I_{G_*})$ and

$$\text{(4.51)} \qquad D = \left[\Pi_{\cdot *}, \ \begin{pmatrix} I_{K_1} \\ 0 \end{pmatrix} \right]$$

is a $K \times (G_* + K_1)$ matrix.

The limiting distribution of $\sqrt{T} \mathrm{vec}\left[(1/T) Y' \bar{P}_Z Y - \Omega\right]$ is $N(0, \Omega \otimes \Omega)$. Hence the limiting distribution of

$$\text{(4.52)} \qquad \sqrt{T} \mathrm{vec}\left[J_2' \left(\frac{1}{T} Y' \bar{P}_Z Y - \Omega \right) B \right]$$

is $N(0, \Sigma \otimes \Omega_{22})$. From (4.50) and (4.52), the asymptotic covariance matrix of $\mathrm{vec}\,\eta^*$ is given by

$$\text{(4.53)} \qquad \Sigma \otimes \big(\Omega_{22} (\rho J_1, I_{G_2}, 0) (D' M D)^{-1} (\rho J_1, I_{G_2}, 0)' \Omega_{22} + \Omega_{22} \big),$$

where $J_1' = (0, I_{G_1 - G_0})$. Hence we define a Wald statistic by

(4.54)

$$W_3 = T (\mathrm{vec}\,\widehat{\eta})' \big\{ \widehat{\Sigma} \otimes \big[T \widehat{\Omega}_{22} (\hat{\rho} J_1, I_{G_2}, 0) (\widehat{D}' Z' Z \widehat{D})^{-1} (\hat{\rho} J_1, I_{G_2}, 0)' \widehat{\Omega}_{22} + \widehat{\Omega}_{22} \big] \big\}^{-1} (\mathrm{vec}\,\widehat{\eta})$$

where $\widehat{\Sigma}, \hat{\rho}$, and $\widehat{\Omega}_{22}$ are the maximum likelihood estimators of Σ, ρ, and Ω_{22}, and $\widehat{D}$ is the maximum likelihood estimator of D, respectively. Again, using Lemma A.5, we have

$$\text{(4.55)} \qquad \begin{aligned} W_3 = \mathrm{tr} \big\{ \widehat{B}' Y' \overline{P}_{Z_1} Y_2 \big[& T \widehat{\Omega}_{22} (\hat{\rho} J_1, I_{G_2}, 0) (\widehat{D}' Z' Z \widehat{D})^{-1} (\hat{\rho} J_1, I_{G_2}, 0)' T \widehat{\Omega}_{22} \\ & + T \widehat{\Omega}_{22} \big]^{-1} Y_2' \overline{P}_{Z_1} Y \widehat{B} \widehat{\Sigma}^{-1}, \end{aligned}$$

where $\widehat{B}$ is the maximum likelihood estimator of B under H_ξ.

This Wald-type statistic is similar to the Wald statistic proposed by Smith (1985) when $G_0 = 1$. Although it is complicated in general, it can be further simplified in the case when the subsystem of structural equations is just-identified as the alternative hypothesis. In this case, since $\lambda_i = 0$, $i = 1, \ldots, G_0$, in (3.13) we have $T\widehat{\Omega}_{22} = Y_2'\overline{P}_Z Y_2$, $\widehat{\Pi}_{\cdot 2} = (Z'Z)^{-1}Z'Y_2$, $T\widehat{\Sigma} = \widehat{B}'Y'\overline{P}_{Z_1}Y\widehat{B}$. Then, in particular, when $G_0 = G_1 = 1$, it can be shown that W_3 in (4.55) is equivalent to the statistic proposed by Wu (1973) and Wu (1974) except $\widehat{\Sigma}$. This may give the Wu test procedure another new interpretation.

4.3. An inequality among statistics.

We have derived three types of statistics for the predeterminedness restriction in a subsystem of structural equations. There is a simple inequality among the statistics we have derived for $H_{\xi,\eta} : \xi = 0,\ \eta = 0$ vs $H_A : \xi \neq 0$. Using Lemma A.7, we have

$$0 \leq LM_2 \leq LR_2 \leq W_2. \tag{4.56}$$

This inequality is an analogue to (3.50) for the problem of testing of the block identifying restriction in Section 3. However, a similar inequality can not be obtainable for the testing problem of $H_{\xi,\eta} : \xi = 0,\ \eta = 0$ vs $H_\xi : \xi = 0$.

5. Conclusion

In this paper we have derived systematically a number of procedures for testing the block identifiability condition and the predeterminedness condition in a subsystem of structural equations. We generalized the test statistics proposed previously and derived the LR test, LM test, and Wald test for these two problems. This formulation enables us to give new interpretations to a number of testing procedures. We explored the relationship between test statistics in econometrics and those in multivariate statistical analysis and obtained some new interpretations for some test statistics commonly known in multivariate statistical analysis.

Among three types of test statistics discussed in this paper, the LR test procedures have often turned out to be considerably simpler than the other two procedures. It is especially evident for the econometric exogeneity hypothesis when $G_0 < G_1$. This finding may be important for practical implementation of the testing procedure in empirical studies.

Appendix A

In this appendix we present some useful lemmas. Most of these lemmas are known in multivariate statistical analysis and their proofs can be found in the works of Anderson (1984) or Rao (1973). We shall present only the proof of Lemma A.2, which may be new in econometrics.

Lemma A.1: Let D and G be $p \times p$ positive definite matrices. Then the function

$$f(G) = -N \log |G| - \operatorname{tr}(G^{-1}D) \tag{A.1}$$

is maximized at $G = (1/N)D$.

Lemma A.2: Let a $p \times p$ positive definite matrix A be decomposed into $(p_1 + p_2) \times (p_1 + p_2)$ submatrices $A = (A_{ij})$. For any $q \times p_1$ matrix B and $q \times p_2$ matrix C,

$$\begin{aligned} \min_{C} \left| A + \begin{pmatrix} B' \\ C' \end{pmatrix} (B, C) \right| &= \frac{|A|}{|A_{11}|} |A_{11} + B'B| \\ &= \left|A_{11} + B'B\right| \left|A_{22} - A_{21}A_{11}^{-1}A_{12}\right| \end{aligned} \tag{A.2}$$

and the minimum occurs at $C = BA_{11}^{-1}A_{12}$.

Proof: Let $D = (B, C)$. Then

$$|A + D'D| = \begin{vmatrix} A & -D' \\ D & I_q \end{vmatrix} = |A| \left|I_q + DA^{-1}D'\right|. \tag{A.3}$$

Let also the inverse matrix A be decomposed into $(p_1 + p_2) \times (p_1 + p_2)$ submatrices $A^{-1} = (A^{ij})$. Then

$$\begin{aligned} DA^{-1}D' &= (C + BA^{12}(A^{22})^{-1})A^{22}(C + BA^{12}(A^{22})^{-1})' + B(A^{11} - A^{12}(A^{22})^{-1}A^{21})B' \\ &\geq B(A^{11} - A^{12}(A^{22})^{-1}A^{21})B' = BA_{11}^{-1}B'. \end{aligned}$$

Hence,

$$|A + D'D| \geq |A| \left|I_q + BA_{11}^{-1}B'\right|. \tag{A.4}$$

Finally, we obtain (A.2) by using (A.3). ∎

Lemma A.3: Let A be a $p \times p$ positive semidefinite matrix and $0 \leq \lambda_1 \leq \cdots \leq \lambda_p$ be its characteristic roots. Let B be a $p \times q$ ($p > q$) matrix. Then

$$\min_{B'B=I} |B'AB| = \prod_{i=1}^{q} \lambda_i. \tag{A.5}$$

Proof. If A is singular, the left-hand and right-hand sides of (A.5) are 0. For A positive definite, we use

$$\mathrm{ch}_i(UV) \geq \mathrm{ch}_j(U)\mathrm{ch}_k(V) \tag{A.6}$$

for $j + k \leq i + 1$, U positive definite, and V positive semidefinite. Here $\mathrm{ch}_i(W)$ is the i-th smallest characteristic root of W. [See, for example, Theorem 2.2 of Anderson and Das Gupta (1963).] For any B such that $B'B = I_q$

$$\mathrm{ch}_{p-q+i}(BB') = \mathrm{ch}_i(B'B) = \mathrm{ch}_i(I_q) = 1, \quad i = 1, \ldots, q. \tag{A.7}$$

In (A.6) let $U = A$, $V = BB'$, $i = p - q + j$ and $k = p - q + 1$ to obtain

$$\begin{aligned} \mathrm{ch}_j(B'AB) &= \mathrm{ch}_{p-q+j}(ABB') \geq \mathrm{ch}_j(A)\mathrm{ch}_{p-q+1}(BB') \\ &= \mathrm{ch}_j(A), \qquad j = 1, \ldots, q. \end{aligned} \tag{A.8}$$

Then

$$|B'AB| = \prod_{j=1}^{q} \mathrm{ch}_j(B'AB) \geq \prod_{j=1}^{q} \mathrm{ch}_j(A) = \prod_{j=1}^{j} \lambda_j. \tag{A.9}$$

Equality is obtained when the columns of B are the characteristic vectors of A corresponding to the roots $\lambda_1, \ldots, \lambda_q$. ∎

Lemma A.4:

$$\frac{\partial \operatorname{tr}(AB)}{\partial B} = A', \quad \frac{\partial \operatorname{tr}(B'ABC)}{\partial B} = ABC + A'BC'. \tag{A.10}$$

Lemma A.5: For any $m \times n$ matrix $A = (a_1, \ldots, a_n)$, we define an $mn \times 1$ vector $\operatorname{vec} A = (a_1', \ldots, a_n')'$. Then for any conformable matrices,

$$\operatorname{vec}(BXC) = (C' \otimes B)(\operatorname{vec} X), \tag{A.11}$$

(A.12) $$\mathrm{tr}\,(BCD) = (\mathrm{vec}\,B')'(I \otimes C)(\mathrm{vec}\,D),$$

(A.13) $$\mathrm{tr}\,(BX'CXD) = (\mathrm{vec}\,X)'(DB \otimes C')(\mathrm{vec}\,X),$$

where each K is a commutation matrix defined by $\mathrm{vec}\,(C) = K\mathrm{vec}\,(C')$ for an arbitrary matrix C of suitable order.

Lemma A.6:

(A.14) $$\overline{P}_{B,C} = \overline{P}_B - \overline{P}_B C(C'\overline{P}_B C)^{-1} C'\overline{P}_B,$$

where D^{-1} stands for the generalized inverse matrix of any matrix D.

Lemma A.7: For non-negative λ_i, $i = 1, \ldots, p$,

(A.15) $$\sum_{i=1}^{p} \frac{\lambda_i}{1+\lambda_i} \leq \log \prod_{i=1}^{p} (1+\lambda_i) \leq \sum_{i=1}^{p} \lambda_i.$$

Appendix B

Maximum Likelihood Estimators

Maximum likelihood estimators of Π, B, and Ω under H_ξ as well as the likelihood ratio test of H_ξ were developed by Anderson (1951). This appendix summarizes the results needed in the present paper. The exposition will refer to the table of correspondence between the notation of the present paper and that of Anderson (1951) at the end of this appendix. The reduced form model is specified in B.1 of the table and the basic statistics in B.2.

A matrix B satisfying $\Pi_{2\cdot}B = 0$ can be multiplied on the right by an arbitrary nonsingular matrix F to obtain BF, which also satisfies the condition, $\Pi_{2\cdot}(BF) = 0$. The maximum likelihood estimators of B similarly can be transformed by multiplication on the right by an arbitrary nonsingular matrix. Any such estimator is composed of linear combinations of the G_0 characteristic vectors of

$$(Y'\bar{P}_Z Y)^{-1} Y'(P_Z - P_{Z_1})Y \tag{B.1}$$

corresponding to the G_0 smallest characteristic roots. With the normalization defined in B.3 of the table the matrix is C. When B is normalized so

$$B = \begin{pmatrix} I \\ -B_* \end{pmatrix}, \tag{B.2}$$

the maximum likelihood estimator is

$$\widehat{B}_H = CC_0^{-1} = \begin{pmatrix} I \\ -C_* C_0^{-1} \end{pmatrix}. \tag{B.3}$$

In Anderson (1951) the likelihood was maximized under the condition that $\Gamma'\Sigma\Gamma = I$, but it was shown that the maximum of the likelihood function was independent of the normalization. The estimator of Γ was normalized by

$$\frac{1}{N}\widehat{\Gamma}' A \widehat{\Gamma} = \widehat{\Gamma}' H \widehat{\Gamma} = (I + \Phi^*)^{-1}.$$

Then

$$\widehat{\Gamma} = C(I + \Phi^*)^{-\frac{1}{2}},$$

where Φ^* is the diagonal matrix composed of the smallest characteristic roots of (B.1). The maximum likelihood estimators of B, Ω, Σ, and $\Pi_{2\cdot}$ are given in B.4.

Table. Correspondence of Notations

This paper	Anderson (1951)
B.1. Model	
$Y = Z\Pi + V$	$\mathcal{E}X = \bar{B}Z$
$T \times G \quad T \times K \quad K \times G \quad T \times G$	$p \times N \quad p \times q \quad q \times N$
$\mathcal{E}v_t v_t' = \Omega$	Σ
$Z = (Z_1, Z_2), \Pi = \begin{pmatrix} \Pi_{1\cdot} \\ \Pi_{2\cdot} \end{pmatrix} \begin{matrix} K_1 \\ K_2 \end{matrix}$	$\bar{B} = (\bar{B}_1, \bar{B}_2), Z = \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} \begin{matrix} q_1 \\ q_2 \end{matrix}$
$\Pi_{2\cdot}B = 0$	$\Gamma'\bar{B}_2 = 0$
$K \times G \quad G \times G_0 \quad K \times G_0$	$m \times p \quad p \times q \quad m \times q$
K, G, G_0, T	q, p, m, N
B.2. Descriptive Statistics	
$(Z'Z)^{-1}Z'Y$	$B = XZ'(ZZ')^{-1}$
$S = \widehat{V}'\widehat{V}$	$A = NH = (N - q)S$
$= Y'\bar{P}_Z Y$	$= (X - BZ)(X - BZ)'$
	$= X\bar{P}_Z X'$
$A_{22\cdot 1} = Z_2'Z_2 - Z_2'Z_1(Z_1'Z_1)^{-1}Z_1'Z_2$	$Q = Z_2Z_2' - Z_2Z_1'(Z_1Z_1')^{-1}Z_1Z_2'$
$= Z_2'\bar{P}_{Z_1}Z_2$	$= Z_2\bar{P}_{Z_1}Z_2'$
$Y'(P_Z - P_{Z_1})Y$	$B_2'QB_2$
B.3. Determinantal Roots and Associated Matrices	
$\lvert Y'(P_Z - P_{Z_1})Y - \lambda Y'\bar{P}_Z Y \rvert = 0$	$\lvert B_2QB_2' - \phi A \rvert = 0$
$\lambda_G \geq \lambda_{G-1} \geq \cdots \geq \lambda_1 > 0$	$\phi_1 \geq \phi_2 \geq \cdots \geq \phi_p > 0$
$\left.\begin{array}{r} [Y'(P_Z - P_{Z_1})Y - \lambda_i Y'\bar{P}_Z Y]c = 0 \\ \frac{1}{T}c'Y'\bar{P}_Z Yc = 1 \end{array}\right\} c_i$	$\left.\begin{array}{r} (B_2QB_2' - \phi_i A)c = 0 \\ \frac{1}{N}c'Ac = c'H\ c = 1 \end{array}\right\} c_i$

This paper	Anderson (1951)
$C = (c_1, \ldots, c_{G_0}) = \begin{pmatrix} C_0 \\ -C_* \end{pmatrix} \begin{matrix} G_0 \\ G_* \end{matrix}$	$C = (c_{p-m+1}, \ldots, c_p)$
$\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_{G_0})$	$\Phi^* = \mathrm{diag}(\phi_{p-m+1}, \ldots, \phi_p)$
$C' \frac{1}{T} Y' \bar{P}_Z Y C = I$	$C' \frac{1}{N} A C = C' H C = I$
$Y'(P_Z - P_{Z_1}) Y C = Y' \bar{P}_Z Y C \Lambda$	$B_2 Q B_2' C = A C \Phi^*$
$\frac{1}{T} C' Y' (P_Z - P_{Z_1}) Y C = \Lambda$	$\frac{1}{N} C' B_2 Q B_2' C = \Phi^*$

B.4. Estimators

This paper	Anderson (1951)
$\widehat{B}_H = C C_0^{-1} = \begin{pmatrix} I \\ -C_* C_0^{-1} \end{pmatrix}$	$\widehat{\Gamma} = C(I + \Phi^*)^{-\frac{1}{2}}$
$\widehat{\Omega} = \frac{1}{T} Y' \bar{P}_Z Y + \frac{1}{T} Y' \bar{P}_Z Y C \Lambda C' \frac{1}{T} Y' \bar{P}_Z Y$ $= \frac{1}{T} Y' \bar{P}_Z Y + \frac{1}{T} Y'(P_Z - P_{Z_1}) Y C C' \frac{1}{T} Y' \bar{P}_Z Y$	$\widehat{\Sigma} = H + H \widehat{\Gamma}(I + \Phi^*) \Phi^* \widehat{\Gamma}' H$ $= H + H C \Phi^* C' H$
$\widehat{\Sigma}_H = \widehat{B}_H' \widehat{\Omega} \widehat{B}_H$ $= (C_0')^{-1}(I + \Lambda) C_0^{-1}$ $= \frac{1}{T} \widehat{B}_H' Y' \bar{P}_{Z_1} Y \widehat{B}_H$	$\widehat{\Gamma}' \widehat{\Sigma} \widehat{\Gamma} = I$ $C' \widehat{\Sigma} C = I + \Phi^*$
$\widehat{\Omega} C = \frac{1}{T} Y' \bar{P}_Z Y C + \frac{1}{T} Y'(P_Z - P_{Z_1}) Y C$ $= \frac{1}{T} Y' \bar{P}_{Z_1} Y C$ $= \frac{1}{T} Y' \bar{P}_Z Y C (I + \Lambda)$	
$\widehat{\Pi}_{2\cdot} = (0, I_{K_2})(Z'Z)^{-1} Z' Y \left(I - C C' \frac{1}{T} Y' \bar{P}_Z Y\right)$	$\widehat{B}_2 = (I - \widehat{\Sigma} \widehat{\Gamma} \widehat{\Gamma}') B_2$

References

Anderson, T. W. (1951). Estimating linear restrictions on regression coefficients for multivariate normal distributions, *Annals of Mathematical Statistics* **34**, 327-351.

Anderson, T. W. (1984). *An Introduction to Multivariate Statistical Analysis*, Second Edition, John Wiley and Sons.

Anderson, T. W., and S. Das Gupta (1963). Some inequalities on characteristic roots of matrices, *Biometrika* **50**, 671-672.

Anderson, T. W., and N. Kunitomo (1989a). Asymptotic robustness in regression and autoregression based on Lindeberg conditions. Technical Report No. 23, ARO Contract No. DAAG-29-85-K-0239, Department of Statistics, Stanford University, and Discussion Paper 89-F-10, Faculty of Economics, University of Tokyo.

Anderson, T. W., and N. Kunitomo (1989b). Asymptotic robustness of tests in simultaneous equations. Unpublished manuscript.

Anderson, T. W., and H. Rubin (1949). Estimation of the parameters of a single equation in a complete system of stochastic equations, *Annals of Mathematical Statistics* **20**, 570-582.

Basmann, R. (1960). On finite sample distributions of generalized classical linear identifiability test statistics, *Journal of the American Statistical Association* **55**, 650-659.

Byron, R. P. (1972). Testing for misspecification in econometric systems using full information, *International Economic Review* **13**, 745-756.

Byron, R. P. (1974). Testing structural specification using the unrestricted reduced form, *Econometrica* **42**, 869-883.

Durbin, J. (1954). Errors in variables, *Review of the International Statistical Institute* **22**, 23-32.

Engle, R. F. (1984). Wald, likelihood ratio, and lagrange multiplier tests in econometrics, in *Handbook of Econometrics* Vol. 2, (Z. Griliches and M. Intrigator, eds.), North-Holland.

Engle, R. D., Hendry, and J-F. Richard (1983). Exogeneity, *Econometrica* **51**, 277-304.

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

Hillier, G. (1987). Classes of similar regions and their power properties for some econometric testing problems, *Econometric Theory* **3**, 1-44.

Holly, A. (1987). Testing for exogeneity: A survey, *Economic Notes,* Monte Dei Paschi di Siena, 108-130.

Hwang, H. (1980a). Test of independence between a subset of stochastic regressors and disturbances, *International Economic Review* **21**, 749-760.

Hwang, H. (1980b). A comparison of tests of overidentifying restrictions, *Econometrica* **48**, 1821-1825.

Hwang, H. (1985). The equivalence of Hausman and lagrange multiplier tests of independence between disturbance and a subset of stochastic regressors, *Economics Letters* **17**, 83-86.

Kariya, T., and J. Hodoshima (1980). Finite sample properties of the tests for independence in structural systems and the LRT, *Economic Studies Quarterly* **31**, 45-56.

Koopmans, T. C., and Hood, W. C. (1953). The estimation of simultaneous linear economic relationships, *Studies in Econometric Method* (W. C. Hood and T. C. Koopmans, eds.), Yale University Press, 112-199.

Nakamura, A., and M. Nakamura (1981). On the relationships among several specification error tests presented by Durbin, Wu and Hausman, *Econometrica* **49**, 1583-1588.

Rao, C. R. (1973). *Linear Statistical Inference and Its Applications*, Second Edition, John Wiley and Sons.

Revankar, N. (1978). Asymptotic relative efficiency analysis of certain tests of independence in structural systems, *International Economic Review* **1**, 165-179.

Revankar, N. S., and M. J. Hartley (1973). An independence test and conditional unbiased predictions in the context of simultaneous equation systems, *International Economic Review* **14**, 625-631.

Revankar, N., and N. Yoshino (1989). An 'expanded equation' approach to weak-exogeneity tests in structural systems and an application, *Review of Economics and Statistics*, in press.

Smith, R. (1985). Wald tests for the independence of stochastic variables and disturbances of a single linear stochastic simultaneous equation, *Economics Letters* **17**, 87-90.

Wegge, L. L. (1978). Constrained indirect least squares estimators, *Econometrica* **46**, 435-449.

Wu, D-M. (1973). Alternative tests of independence between stochastic regressors and disturbances, *Econometrica* **41**, 733-750.

Wu, D-M. (1973). Alternative tests of independence between stochastic regressors and disturbances: Finite sample results, *Econometrica* **42**, 529-546.