90-F-19

# Asymptotic Robustness of Tests of Overidentification and Exogeneity

by

T.W. Anderson
Department of Statistics
Stanford University

and

Naoto Kunitomo
Faculty of Economics
University of Tokyo

July 1990

# Asymptotic Robustness of Tests of Overidentification and Exogeneity

T. W. Anderson and Naoto Kunitomo

## 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 based on a priori grounds 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 by Anderson and Rubin (1949), Koopmans and Hood (1953), Basman (1960), Wu (1973), Byron (1972), Revankar and Hartley (1973), Revankar (1978), Hausman (1978), Kariya and Hodoshima (1980), Hwang (1980a), and Revankar and Yoshino (1989) among many others have drawn some attention and have been applied in empirical work.

Anderson and Kunitomo (1989a) derived systematically several test procedures and obtained relationships among different test statistics. For these purposes they considered a subsystem of structural equations and regarded the single equation method as a special case of this formulation. They obtained three types of test procedures, namely, the likelihood ratio (LR) test, the Lagrange Multiplier (LM) test, and the Wald test for the block identifiability restrictions and the predeterminedness restrictions in the subsystem of structural equations; the test statistics derived include all test statistics mentioned above as special cases and give new interpretations to some of them.

In this paper we shall derive the asymptotic distributions of test statistics discussed in Anderson and Kunitomo (1989a) under a set of fairly general conditions on the disturbance terms. For this purpose we shall use a new martingale central limit theorem and a martingale convergence theorem based on a Lindeberg-type condition for martingale difference sequences developed by Anderson and Kunitomo (1989b) and apply them to the present situation. We allow a finite number of lagged endogenous variables, and the disturbance terms are not necessarily independent. We shall show that the limiting distributions of test statistics considered in this paper are noncentral $\chi^2$-distributions under local alternative hypotheses and are central $\chi^2$-distributions under the null hypotheses when the disturbances are the martingale difference sequences. Because test statistics have often been proposed under a set of relatively restrictive assumptions, it is important to show that the assumptions usually made are not essential for the testing procedures in practice.

In Section 2 we formulate the two hypotheses and in Section 3 we summarize the test statistics in a subsystem of structural equations. In Section 4 we give some general results on the asymptotic distributions of test statistics introduced in Section 3. Detailed proofs of the theorems are given in Section 5.

## 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 $K_1$ predetermined variables, $B$ and $\Gamma$ 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. The columns of matrix $B$ are linearly independent; that is, the rank of $B$ is $G_0$. When $G_0 = 1$, (2.1) is the usual single structural equation.

The reduced form equation for the endogenous variables $Y$ appearing in the first $G_0$ structural equations (2.1) with $K$ $(= K_1 + K_2)$ 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 the 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 $\Omega$ is a $G \times G$ non negative 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

$\Gamma$ 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 $\Pi$ has been partitioned into submatrices of $K_1$ and $K_2$ rows:

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

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

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

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

say. $\Sigma$ is a $G_0 \times G_0$ nonnegative 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 of the block identifiability in (2.1),

$$\text{rank}(\Pi_{2\cdot}) = G - G_0 = G_*. \tag{2.12}$$

The order condition is

$$L = K_2 - G_* \geq 0. \tag{2.13}$$

In the above notation $L$ is often 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

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

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

Then from (2.10), it is clear that the block identifiability condition is 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 $H_\nu$. This testing problem is mathematically equivalent to the hypothesis for the rank test in multivariate analysis. (See Anderson (1984), Chapter 8.)

## 2.3. Predeterminedness

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 ($G = G_1 + G_2$), $V = (V_1, V_2)$, $v_t = (v_{1t}', v_{2t}')'$, 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 $H_\xi$ and $H_\eta$ imply the hypothesis $H_{\xi,\eta} : \xi = 0$, $\eta = 0$. When the disturbance terms follow the multivariate normal distribution, the uncorrelatedness implies independence between any subset of regressor $Y_2$ and disturbance terms in (2.1), This testing problem has been sometimes called the test of independence. The hypothesis of predeterminedness in this paper may be called weak exogeneity. (See Engle, Hendry, and Richard (1983) and Holly (1987).)

## 3. Test Statistics for Two Hypotheses

In this section we summarize test statistics for the block identifiability condition and the predeterminedness condition in a subsystem of structural equations. The derivations of the statistics have been given in Anderson and Kunitomo (1989a). Three types of test procedures were discussed: the likelihood ratio (LR) test, the Lagrange Multiplier (LM) test, and Wald test for two hypotheses. In order to derive test procedures the multivariate normal distributions for the disturbances $v_t$ was assumed. However, this assumption shall be relaxed considerably in this paper.

### 3.1. Tests for block identifiability

Under the assumption of multivariate normality of the disturbance terms $\{v_t\}$, the log likelihood ratio (LR) times $-2$ for $H_\xi : \xi = 0$ vs $H_A : \xi \neq 0$ is

$$LR_1 = T \sum_{i=1}^{G_0} \log(1 + \lambda_i), \tag{3.1}$$

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

$$\left| Y'(P_Z - P_{Z_1})Y - \lambda Y' \overline{P}_Z Y \right| = 0, \tag{3.2}$$

$P_Z = Z(Z'Z)^{-1}Z'$ denotes the projection operator onto the space spanned by the column vectors of $Z$, and $\overline{P}_Z = I_T - P_Z$ for any (full column) matrix $Z$.

The equation (3.2) is a sample analogue of (2.14). For $G_0 = 1$ the likelihood ratio statistic (3.1) was derived by Anderson and Rubin (1949). In this case $LR_1$ is a function of the smallest characteristic root in the limited information maximum likelihood (LIML) estimation method. When $G_0 = 2$, $LR_1$ is the statistic proposed by Koopmans and Hood (1953) for the nonidentification test.

The Lagrange Multiplier (LM) statistic, which is identical to the Rao Score statistic, has been developed as a statistic to test a hypothesis $H$ about a vector parameter $\theta$ in a likelihood $L$. In these general terms the criterion is

$$LM = \left( \frac{\partial \log L}{\partial \theta} \bigg|_H \right) \left( - \frac{\partial^2 \log L}{\partial^2 \theta} \bigg|_H \right)^{-1} \left( \frac{\partial \log L}{\partial \theta} \bigg|_H \right), \tag{3.3}$$

where $H$ denotes the null hypothesis and the value of the parameter in (3.3) maximizes the likelihood under the null hypothesis. Applying this general principle to our present testing problem for $H_\xi := 0$ vs $H_A : \xi \neq 0$ yields

$$LM_1 = \text{tr } \widehat{B}'Y'(P_Z - P_{Z_1})Y\widehat{B}\widehat{\Sigma}^{-1}, \tag{3.4}$$

where $\widehat{B}$ and $\widehat{\Sigma} = \widehat{B}'Y'\overline{P}_{Z_1}Y\widehat{B}$ are the maximum likelihood estimators of $B$ and $\Sigma$ under the null hypothesis. When we use the roots of (3.2), this statistic is

$$LM_1 = T\sum_{i=1}^{G_0} \frac{\lambda_i}{1+\lambda_i}. \tag{3.5}$$

When $G_0 = 1$, this statistic $LM_1$ is identical to the $LM$ statistic proposed by Byron (1972). When $G_0 = G$, (3.5) is the Bartlett-Nanda-Pillai trace criterion in multivariate statistical analysis.

In general terms the Wald test is based on the statistic

$$h(\hat{\theta})'\left[C(\hat{\theta})\right]^{-1}h(\hat{\theta}), \tag{3.6}$$

where $\hat{\theta}$ is the maximum likelihood estimator of the parameter vector $\theta$ under the alternative hypothesis and $C(\hat{\theta})$ is an estimator of the asymptotic covariance matrix of $h(\hat{\theta})$. In our problem the null hypothesis is that rank of $\Pi_{2\cdot}$ is $G - G_0 = G_*$, say. We partition the matrix $Y = (Y_0, Y_*)$ as $T \times (G_0 + G_*)$. By expressing (2.10) in the form of $h(\theta) = \mathrm{vec}(\Pi_{2\cdot}B) = 0$, we obtain a Wald test as

$$W_1 = \mathrm{tr}\left[\widehat{\Sigma}^{-1}\widetilde{B}_{TS}'Y'(P_Z - P_{Z_1})Y\widetilde{B}_{TS}\right], \tag{3.7}$$

where $\widetilde{B}_{TS}' = (I_{G_0}, -\widetilde{B}_*')$ is the two-stage least-squares estimator and

$$\widetilde{B}_* = \left[Y_*'(P_Z - P_{Z_1})Y_*)\right]^{-1}Y_*'(P_Z - P_{Z_1})Y_0. \tag{3.8}$$

The numbering of the columns of $Y$ may be arbitrary. When we use the unrestricted estimator $\widehat{\Omega} = (1/T)Y'\overline{P}_Z Y$ for $\Omega$, the resulting Wald statistic $W_1$ is the statistic derived by Wegge (1978) for $G_0 = 1$, which is also identical to the Wald statistic derived by Byron (1974). When we use the maximum likelihood estimator of $\Omega$ under the null hypothesis $\widehat{\Omega} = (1/T)Y'\overline{P}_{Z_1}Y$, the resulting $W_1$ is the statistic proposed by Basman (1960) for the case of $G_0 = 1$.

The limited information maximum likelihood estimator $\widehat{B}_{LI}$ under $H_\xi : \xi = 0$ is asymptotically equivalent to $\widetilde{B}_{TS}$ in the sense that $\sqrt{T}(\widehat{B}_{LI} - \widetilde{B}_{TS}) \xrightarrow{\text{P}} 0$. Thus we may substitute $\widehat{B}_{LI}$ for $\widetilde{B}_{TS}$ for an estimator of $\Sigma$. (See Lemma 1 in Section 4; it is assumed that the rank of $\Pi_{2*}$ consisting of the last $G_*$ columns of $\Pi_{2\cdot}$ is $G_*$.) To construct $\widehat{B}_{LI}$ define the vector $c_i$ by

$$\left[Y'(P_Z - P_{Z_1})Y - \lambda_i Y'\overline{P}_Z Y\right]c = 0 \tag{3.9}$$

and $c'Y'\overline{P}_Z Y c = T$, $i = 1, \ldots, G_0$, and define the matrices $C$, $C_0$, and $C_*$ by

$$C = (c_1, \ldots, c_{G_0}) = \begin{pmatrix} C_0 \\ C_* \end{pmatrix}, \tag{3.10}$$

where $C_0$ is $G_0 \times G_0$. Then

$$\widehat{B}_* = C_* C_0^{-1} \tag{3.11}$$

and $\widehat{B}_{LI} = (I_{G_0}, -\widehat{B}_*')'$. The statistic $W_1$ can be modified by replacing $\widehat{B}_{TS}$ by $\widehat{B}_{LI}$ to obtain

$$W_1' = T \sum_{i=1}^{G_0} \lambda_i, \tag{3.12}$$

where $\lambda_i$, $i = 1, \ldots, G_0$, are the $G_0$ smallest roots of (3.2). When $G = G_0$, $W_1'$ is the Lawley-Hotelling Trace Criterion.

### 3.2. Test statistics for predeterminedness against unrestricted alternatives

Under the assumption of the multivariate normal distribution for the disturbance terms $\{v_t\}$, the log likelihood ratio criterion (LR) times $-2$ for $H_{\xi,\eta} : \xi = 0,\ \eta = 0$ vs $H_A : \xi \neq 0,\ \eta \neq 0$ is

$$LR_2 = T \sum_{i=1}^{G_0} \log(1 + \lambda_i^*), \tag{3.13}$$

where $\lambda_{G_0}^* \geq \cdots \geq \lambda_1^* \geq 0$ are the roots of

$$\left| Y_1'(P_{Y_2,Z} - P_{Y_2,Z_1})Y_1 - \lambda^* Y_1' \overline{P}_{Y_2,Z} Y_1 \right| = 0. \tag{3.14}$$

A Lagrange multiplier statistic for testing $H_{\xi,\eta} : \xi = 0,\ \eta = 0$ vs $H_A$ is similar to $LM_1$; it is

$$LM_2 = T \sum_{i=1}^{G_0} \frac{\lambda_i^*}{1 + \lambda_i^*}. \tag{3.15}$$

A (modified) Wald statistic for $H_{\xi,\eta} : \xi = 0,\ \eta = 0$ vs $H_A$ is similar to $W_1'$; it is

$$W_2 = T \sum_{i=1}^{G_0} \lambda_i^*. \tag{3.16}$$

When $G_0 = G_1 = 1$, $W_2$ reduces to the statistic proposed by Revankar and Hartley (1973).

### 3.3. Test statistics for predeterminedness against the alternative of overidentification

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

$$LR_3 = T\sum_{i=1}^{G_0} \log\left(\frac{1+\lambda_i^*}{1+\lambda_i}\right), \tag{3.17}$$

where $\lambda_i^*$ and $\lambda_i$ are the roots of equations (2.14) and (3.14), respectively. For $G_0 = 1$ and $G_1 \geq 1$, $LR_3$ is the statistic obtained by Hwang (1980a). For $G_0 = G_1 = 1$, $LR_3$ reduces to the statistic obtained by Kariya and Hodoshima (1980).

The development of the Lagrange multiplier statistic for testing for $H_{\xi,\eta}$ vs $H_\xi$ is more complicated. The statistic is

$$\begin{aligned} LM_3 &= \mathrm{tr}\,[\widehat{B}_1' Y_1' \overline{P}_{Y_2,Z_1} \overline{P}_Z Y_2' (Y_2' \overline{P}_Z \overline{P}_{RF} \overline{P}_Z Y_2)^{-1} Y_2' \overline{P}_Z \overline{P}_{Y_2,Z_1} Y_1 \widehat{B}_1 \widehat{\Sigma}^{-1}] \\ &= \mathrm{tr}\,[\widehat{B}_1 Y_1' (P_X - P_{R\widehat{F}}) Y_1 \widehat{B}_1 \widehat{\Sigma}^{-1}], \end{aligned} \tag{3.18}$$

where $R = (Y_2, Z)$, $X = (Y_2, Z_1, \overline{P}_Z Y_2)$, $J_1 = (0, I_{G_1 - G_0})'$, $\rho = \Omega_{22}^{-1}\Omega_{21}$, and

$$F = \left[\begin{pmatrix} \rho \\ \Pi_{1\cdot} - \Pi_{2\cdot}\rho \end{pmatrix} J_1,\ \begin{pmatrix} I_{G_2+K_1} \\ 0 \end{pmatrix}\right] \tag{3.19}$$

is evaluated at its maximum likelihood estimator. We have used Lemma A.6 in Anderson and Kunitomo (1989a). When $G_1 = G_0$, we have $P_{RF} = P_{Y_2,Z_1}$ and

$$\begin{aligned} LM_3 &= \mathrm{tr}\,[Y_1'(\overline{P}_{Y_2,Z_1} - \overline{P}_X) Y_1 \widehat{\Sigma}^{-1}] \\ &= T\sum_{i=1}^{G_0} \frac{\lambda_i^{**}}{1+\lambda_i^{**}}, \end{aligned} \tag{3.20}$$

where $\lambda_i^{**}$ are the roots of

$$\left| Y_1'(P_X - P_{Y_2,Z_1})Y_1 - \lambda^{**} Y_1' \overline{P}_X Y_1 \right| = 0, \tag{3.21}$$

and $X = (Y_2, Z_1, \overline{P}_Z Y_2)$ is a $T \times (G_2 + K_1 + G_2)$ matrix. In the present formulation of the LM test, $\widehat{\Sigma}$ should be the maximum likelihood estimator $\Sigma$ under the null hypothesis

$$\widehat{\Sigma} = \widehat{B}_1' \widehat{\Omega}_{11\cdot 2} \widehat{B}_1 = \frac{1}{T} \widehat{B}_1' Y_1' \overline{P}_{Y_2,Z_1} Y_1 \widehat{B}_1. \tag{3.22}$$

Several alternative estimators of $\Sigma$ could be used. If $\widehat{\Sigma} = (1/(T-2G_2-K_1))\widehat{B}_1'Y_1'\overline{P}_XY_1\widehat{B}_1$, $LM_3$ is the statistic proposed by Wu (1973) and Wu (1974) when $G_0 = G_1 = 1$. Hausman (1978) proposed a specification test statistic that is proportional to $LM_3$ when $G_0 = G_1 = 1$. (See Nakamura and Nakamura (1980) and Hwang (1985).) Another possible estimator of $\Sigma$ is $\widehat{\Sigma} = [1/(T-K-G_2)]\widehat{B}_1'Y_1'\overline{P}_{Y_2,Z}Y_1\widehat{B}_1$ because it is an unrestricted sum of squares of the regression residuals. Then the statistic $LM_3$ is the statistic proposed by Revankar (1978) when $G_0 = G_1 = 1$.

A Wald statistic for the null hypothesis $H_{\xi,\eta} : \xi = 0,\ \eta = 0$ vs $H_\xi$ is

$$(3.23)\quad W_3 = \mathrm{tr}\ \widehat{B}'Y'\overline{P}_{Z_1}Y_2\left(T\widehat{\Omega}_{22}(\hat{\rho}J_1, I_{G_2}, 0)(\widehat{D}'Z'Z\widehat{D})^{-1}(\hat{\rho}J_1, I_{G_2}, 0)'\widehat{\Omega}_{22} + \widehat{\Omega}_{22}\right)^{-1} Y_2'\overline{P}_{Z_1}Y\widehat{B}\widehat{\Sigma}^{-1},$$

where $\hat{\rho} = \widehat{\Omega}_{22}^{-1}\widehat{\Omega}_{21}$ and $\widehat{B}$ are the maximum likelihood estimators of $B$ and $\rho$ under $H_\xi$, that is, $\widehat{B} = C$ given by (3.10). This Wald-type statistic is similar to that of Smith (1985).

## 4. Asymptotic Distributions of Statistics

We shall show that the test statistics given in the previous sections have $\chi^2$- distributions under conditions much more general than the conditions under which the tests were derived. Let the $\sigma$-field $F_{t-1}$ be generated by $z_1, v_1, \ldots, z_{t-1}, v_{t-1}, z_t$. We assume that

$$(4.1)\qquad E(v_t | F_{t-1}) = 0 \quad \text{a.s.},$$

$$(4.2)\qquad E(v_t'v_t | F_{t-1}) = \Omega_t \quad \text{a.s.}\ .$$

Note that $\Omega_t$ can be a function of $z_1, v_1, \ldots, z_{t-1}, v_{t-1}, z_t$. Since $u_t = Bv_t$, we obtain

$$(4.3)\qquad E(u_t | F_{t-1}) = 0 \quad \text{a.s.},$$

$$(4.4)\qquad E(u_t'u_t | F_{t-1}) = \Sigma_t \quad \text{a.s.}\ .$$

In the conditional expectation operator in (4.1) to (4.4) $F_{t-1}$ is the information set available at $t-1$. The predetermined variables $z_t$ may include a finite number of past endogenous variables $y_{t-1}, y_{t-2}, \ldots, y_{t-p}$. In order to investigate the asymptotic distribution of the test statistics, we use two theorems for martingale difference sequences given by Anderson and Kunitomo (1989b).

**Theorem 1.** Let $\{z_t, v_t\}$, $t = 1, 2, \ldots$, be a sequence of pairs of random vectors, and let $\{F_t\}$ be an increasing sequence of $\sigma$-fields such that $z_t$ is $F_{t-1}$-measurable and $v_t$ is $F_t$-measurable. Suppose

$$\frac{1}{T}\sum_{t=1}^{T} z_t z_t' \xrightarrow{\text{P}} M, \tag{4.5}$$

where $M$ is a constant matrix, and

$$\max_{1 \le t \le T} \frac{z_t' z_t}{T} \xrightarrow{\text{P}} 0 \tag{4.6}$$

as $T \to \infty$. Suppose further that $E(v_t | F_{t-1}) = 0$ a.s., $E(v_t v_t' | F_{t-1}) = \Omega_t$ a.s.,

$$\frac{1}{T}\sum_{t=1}^{T} (\Omega_t \otimes z_t z_t') \xrightarrow{\text{P}} (\Omega \otimes M), \tag{4.7}$$

where $\Omega$ is a nonnegative constant matrix, and

$$\sup_{t=1,2,\ldots} E\big(v_t' v_t I(v_t' v_t > a) \big| F_{t-1}\big) \xrightarrow{\text{P}} 0 \tag{4.8}$$

as $a \to \infty$. Then

$$\text{vec}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} z_t v_t'\right) \xrightarrow{\text{L}} N(0, \Omega \otimes M). \tag{4.9}$$

**Theorem 2.** Let $\{v_t\}$, $t = 1, 2, \ldots$, be a sequence of random vectors, and let $\{F_t\}$ be an increasing sequence of $\sigma$-fields such that $v_t$ is $F_t$-measurable. Let $E(v_t | F_{t-1}) = 0$ a.s., $E(v_t v_t' | F_{t-1}) = \Omega_t$ a.s., and

$$\frac{1}{T}\sum_{t=1}^{T} \Omega_t \xrightarrow{\text{P}} \Omega, \tag{4.10}$$

where $\Omega$ is a constant matrix, and for any $\varepsilon > 0$,

$$\frac{1}{T}\sum_{t=1}^{T} E\big(v_t' v_t I(v_t' v_t > T\varepsilon) \big| F_{t-1}\big) \xrightarrow{\text{P}} 0 \tag{4.11}$$

as $T \to \infty$. Then

$$\frac{1}{T}\sum_{t=1}^{T} v_t' v_t \xrightarrow{\text{P}} \Omega. \tag{4.12}$$

The proofs of the above theorems are given in Anderson and Kunitomo (1989b). Theorem 1 generalizes Theorem 5(i) of Lai and Robbins (1981), where the scalar $v_t$'s are independently identically distributed. Theorem 2 is a martingale convergence result. Because it is relatively easy to check the conditions in the theorems, they may be useful for many applications. The most important point is that we do not require any condition other than the conditional second-order moments. Both Theorems 1 and 2 allow conditional (as well as unconditional) heteroscedasticities for the martingale difference disturbance terms. We note that condition (4.8) implies condition (4.11). Hence conditions (4.5)-(4.8) and (4.10) are sufficient for the following results.

Consider a sequence of local alternatives for the identifiability restrictions,

$$\Pi B = \begin{pmatrix} \Gamma \\ 0 \end{pmatrix} + \frac{1}{\sqrt{T}}\xi_1 = \begin{pmatrix} \Gamma \\ 0 \end{pmatrix} + \frac{1}{\sqrt{T}}\begin{pmatrix} \xi_{11} \\ \xi_{21} \end{pmatrix}, \tag{4.13}$$

where $\xi_1$ is a $K \times G_0$ matrix. We consider $\Pi = \Pi(T)$ as depending on $T$ in such a way that $\Pi(T) \rightarrow \Pi(0)$, where $\Pi_{2*}(0)$ has rank $G_*$. The matrices $B$ and $\Gamma$ do not depend on $T$. In the rest of the paper we suppress the index $T$. The condition $\Pi_{2*}(0)$ having rank $G_*$ is written as $\Pi_{2*}$ having rank $G_*$. When $\xi_1 = 0$, (4.13) reduces to the block identifiability restrictions. Kunitomo (1988) discussed the formulation of these local alternatives in some detail. We obtain the following theorems.

**Theorem 3.** Suppose (4.5) to (4.8) and (4.10) hold, $M$ is nonsingular, and $\Pi_{2*}$ has rank $G_*$. Let $\widetilde{B}$ be defined by (3.8) and $\widetilde{\Gamma}$ by

$$\widetilde{\Gamma} = (Z_1'Z_1)^{-1}Z_1'(Y_0 - Y_*\widetilde{B}_*). \tag{4.14}$$

Then under the local alternatives (4.13) $\widetilde{B}_* \xrightarrow{p} B_*$, $\widetilde{\Gamma} \xrightarrow{p} \Gamma$, and

$$\operatorname{vec} \sqrt{T}\begin{pmatrix} \widetilde{B}_* - B_* \\ \widetilde{\Gamma} - \Gamma \end{pmatrix} \xrightarrow{d} N\{\operatorname{vec}[(D'MD)^{-1}D'M\xi_1], \Sigma \otimes (D'MD)^{-1}\}, \tag{4.15}$$

where

$$D = \left[\Pi_{\cdot *}, \begin{pmatrix} I_{K_1} \\ 0 \end{pmatrix}\right], \tag{4.16}$$

$$\Pi = (\Pi_{\cdot 0}, \Pi_{\cdot *}). \tag{4.17}$$

**Corollary 1.** Suppose (4.5) to (4.8) and (4.10) hold and $M$ is nonsingular. Let $\widehat{B}_*$ be defined by (3.11) and $\widehat{\Gamma}$ by

$$\widehat{\Gamma} = (Z_1'Z_1)^{-1}Z_1'(Y_0 - Y_*\widehat{B}_*). \tag{4.18}$$

Then under local alternatives (4.13) $\widehat{B}_* \xrightarrow{p} B_*$, $\widehat{\Gamma} \xrightarrow{p} \Gamma$, and

$$\text{(4.19)} \qquad \operatorname{vec} \sqrt{T} \begin{pmatrix} \widehat{B}_* - B_* \\ \widehat{\Gamma} - \Gamma \end{pmatrix} \xrightarrow{d} N\{ \operatorname{vec} \left[(D'MD)^{-1} D'M\xi_1\right],\ \Sigma \otimes (D'MD)^{-1} \}$$

The matrices in the limiting distributions are

$$\text{(4.20)} \qquad D'M = \begin{pmatrix} \Pi'_{\cdot *} M_{\cdot 1} & \Pi'_{\cdot *} M_{\cdot 2} \\ M_{11} & M_{12} \end{pmatrix},$$

$$\text{(4.21)} \qquad D'MD = \begin{pmatrix} \Pi'_{\cdot *} M \Pi_{\cdot *} & \Pi'_{\cdot *} M_{\cdot 1} \\ M_{1\cdot} \Pi_{\cdot *} & M_{11} \end{pmatrix},$$

$$\text{(4.22)} \qquad (D'MD)^{-1} =$$

$$\begin{pmatrix} (\Pi'_{2*} M_{22\cdot 1} \Pi_{2*})^{-1} & -(\Pi'_{2*} M_{22\cdot 1} \Pi_{2*})^{-1} \Pi'_{\cdot *} M_{\cdot 1} M_{11}^{-1} \\ -M_{11}^{-1} M_{1\cdot} \Pi_{\cdot *} (\Pi'_{2*} M_{22\cdot 1} \Pi_{2*})^{-1} & M_{11}^{-1} M_{1\cdot} \Pi_{\cdot *} (\Pi'_{2*} M_{22\cdot 1} \Pi_{2*})^{-1} \Pi'_{\cdot *} M_{\cdot 1} M_{11}^{-1} + M_{11}^{-1} \end{pmatrix},$$

where

$$\text{(4.23)} \qquad M = \begin{pmatrix} M_{1\cdot} \\ M_{2\cdot} \end{pmatrix} = (M_{\cdot 1}, M_{\cdot 2}),$$

$$\text{(4.24)} \qquad M_{22\cdot 1} = M_{22} - M_{21} M_{11}^{-1} M_{12},$$

Then

$$\text{(4.25)} \qquad (D'MD)^{-1} D'M\xi_1$$
$$= \begin{pmatrix} 0 & (\Pi'_{2*} M_{22\cdot 1} \Pi_{2*})^{-1} \Pi'_{2*} M_{22\cdot 1} \\ I_{K_1} & -M_{11}^{-1} M_{1\cdot} \Pi_{\cdot *} (\Pi'_{2*} M_{22\cdot 1} \Pi_{2*})^{-1} \Pi'_{2*} M_{22\cdot 1} + M_{11}^{-1} M_{12} \end{pmatrix} \begin{pmatrix} \xi_{11} \\ \xi_{21} \end{pmatrix}.$$

For the case of $G_0 = 1$ Anderson and Rubin (1950) gave the result of Theorem 3 under different conditions and using a different proof. Textbooks in econometrics often refer to Mann and Wald (1943) for the asymptotic results in the case of $G_0 = 1$ under the identifiability conditions. (See Chapter 10 of Theil (1971), for instance.) Anderson (1951) proved the theorem for arbitrary $G_0$ under the null hypothesis and normality. However, earlier results were obtained under stronger conditions than ours here. Our proof seems much shorter and simpler than the ones already known.

**Theorem 4.** Suppose (4.5) to (4.8) and (4.10) hold, $M$ is nonsingular, and $\Pi_{2*}$ has rank $G_*$. Then under the local alternatives (4.13) each of the statistics $LR_1$, $LM_1$, $W_1$, and $W_1'$ has the limiting distribution of the noncentral $\chi^2$ with $G_0 \times [K_2 - (G - G_0)] = G_0 \times [K_2 - G_*]$ degrees of freedom and noncentrality parameter

(4.26) $$\delta_1^2 = \mathrm{tr}\,(\Theta_1 \Sigma^{-1}),$$

where

(4.27) $$\begin{aligned}\Theta_1 &= \xi_1' [M - MD(D'MD)^{-1} D'M] \xi_1 \\ &= \xi_{21}' [M_{22\cdot1} - M_{22\cdot1}\Pi_{2*}(\Pi_{2*}' M_{22\cdot1}\Pi_{2*})^{-1}\Pi_{2*}' M_{22\cdot1}] \xi_{21}.\end{aligned}$$

When $\xi_{21} = 0$, each of the above statistics has the limiting distribution of $\chi^2$ with $G_0 \times [K_2 - (G - G_0)] = G_0 \times [K_2 - G_*]$ degrees of freedom. This result for the case of $G_0 = 1$ has been obtained previously under the assumptions that disturbances are independently, identically, and normally distributed and there are no lagged endogenous variables in the explanatory variables.

Next we consider a local alternative hypothesis for the predeterminedness condition,

(4.28) $$(\rho, I_{G_2})B = \frac{1}{\sqrt{T}}\eta_1$$

where $\rho = \Omega_{22}^{-1}\Omega_{21}$ and $\eta_1$ is a nonzero $G_2 \times G_0$ matrix. We consider $\Omega = \Omega(T)$ as depending on $T$ such that $\Omega(T) \to \Omega(0)$, which is nonsingular. In order to avoid severe complications in Theorems 5 and 6 we shall assume that

(4.29) $$\mathcal{E}(v_t v_t' | \mathcal{F}_{t-1}) = \Omega(T) \quad \text{a.s.}$$

Then (4.7) and (4.10) are satisfied automatically (for $\Omega_t = \Omega(T)$, $t = 1, \ldots, T$). More general theorems could be stated, but to prove them would require a central limit theorem more general than our Theorem 1.

**Theorem 5.** Suppose $\mathcal{E}(v_t|\mathcal{F}_{t-1}) = 0$ a.s., (4.5), (4.6), (4.8), and (4.29) hold, $M$ and $\Omega$ are nonsingular, and $\Pi_{2*}$ has rank $G_*$. Then under the local alternatives (4.13) and (4.28) each of the statistics $LR_2$, $LM_2$, and $W_2$ has the limiting distribution of the noncentral $\chi^2$ with $G_0 \times [K_2 - (G_1 - G_0)]$ degrees of freedom and noncentrality parameter

(4.30) $$\delta_2^2 = \mathrm{tr}(\theta_2 \Sigma^{-1})$$

where

$$(4.31)\quad \begin{aligned}\theta_2 &= \zeta_1'\left[Q^{-1} - F(F'QF)^{-1}\right]\zeta_1 \\ &= (Q^{-1}\zeta_1)'\left[Q - QF(F'QF)^{-1}Q\right](Q^{-1}\zeta_1) \\ &= (\xi_{21} - \Pi_{22}\eta_1)'\left[M_{22\cdot 1} - M_{22\cdot 1}\Pi_{2*}(\Pi_{2*}'M_{22\cdot 1}\Pi_{2*} + \Omega_{*2}\Omega_{22}^{-1}\Omega_{2*})^{-1}\Pi_{2*}'M_{22\cdot 1}\right] \\ &\quad\cdot(\xi_{21} - \Pi_{22}\eta_1),\end{aligned}$$

$$(4.32)\qquad Q = \begin{pmatrix} \Pi_{\cdot 2}'M\Pi_{\cdot 2} & \Pi_{\cdot 2}'M \\ M\Pi_{\cdot 2} & M \end{pmatrix},$$

$$(4.33)\qquad F = \left[\begin{pmatrix} \rho \\ \Pi_{11} - \Pi_{12}\rho \\ \Pi_{21} - \Pi_{22}\rho \end{pmatrix} J_1, \begin{pmatrix} I_{G_2} \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ I_{K_1} \\ 0 \end{pmatrix}\right],$$

$$(4.34)\qquad \zeta_1 = \begin{pmatrix} \Pi_{\cdot 2}' \\ I_K \end{pmatrix} M\xi_1 + \begin{pmatrix} \Omega_{22}\eta_1 \\ 0 \end{pmatrix},$$

$$(4.35)\qquad Q^{-1}\zeta_1 = \begin{pmatrix} \eta_1 \\ \xi_1 - \Pi_{\cdot 2}\eta_1 \end{pmatrix},$$

$$(4.36)\qquad J_1 = \begin{pmatrix} 0 \\ I_{G_1 - G_0} \end{pmatrix}, \qquad G_1 \times G_0,$$

$$(4.37)\qquad (\Omega_{21}, \Omega_{22}) = (\Omega_{20}, \Omega_{2*}).$$

**Theorem 6.** Assume $\mathcal{E}(v_t|\mathcal{F}_{t-1}) = 0$ a.s., (4.5), (4.6), (4.8), and (4.29) hold, $M$ and $\Omega$ are nonsingular, and $\Pi_{2*}$ has rank $G_*$. Then under the local alternatives (4.13) and (4.28) each of the statistics $LR_3$, $LM_3$, and $W_3$ has the limiting distribution of the noncentral $\chi^2$ with $G_0 \times G_2$ degrees of freedom and noncentrality parameter $\delta_3^2 = \delta_2^2 - \delta_1^2 = \operatorname{tr}\left[(\theta_2 - \theta_1)\Sigma^{-1}\right]$.

The noncentrality parameter can also be written

$$(4.38)\qquad \theta_3 = \zeta_2^{*\prime}\left[\Omega_{22} - \Omega_{2*}(\Pi_{2*}'M_{22\cdot 1}\Pi_{2*} + \Omega_{*2}\Omega_{22}^{-1}\Omega_{2*})^{-1}\Omega_{*2}\right]\zeta_2^*,$$

where

$$(4.39) \qquad \zeta_2^* = \eta_1 - \Omega_{22}^{-1}\Omega_{2*}(\Pi_{2*}'M_{22\cdot1}\Pi_{2*})^{-1}\Pi_{2*}'M_{22\cdot1}\xi_{21}.$$

Note that when $\xi_{21} = 0$ the noncentrality parameter for $LR_2$, $LM_2$, and $W_2$ is the same as for $LR_3$, $LM_3$, and $W_3$. However, since the number of degrees of freedom for the first set of statistics is greater than for the second set, the second set yield greater asymptotic power. If, in addition, $G_0 = G_1$, the noncentrality parameter is

$$\begin{aligned}(4.40)\ \theta_2 &= \theta_3 \\ &= \eta_1'\left[\Pi_{2*}'M_{22\cdot1}\Pi_{2*} - \Pi_{2*}'M_{22\cdot1}\Pi_{2*}(\Pi_{2*}'M_{22\cdot1}\Pi_{2*} + \Omega_{22})^{-1}\Pi_{2*}'M_{22\cdot1}\Pi_{2*}\right]\eta_1 \\ &= \eta_1'\left[\Pi_{2*}'M_{22\cdot1}\Pi_{2*}(\Pi_{2*}'M_{22\cdot1}\Pi_{2*} + \Omega_{22})^{-1}\Omega_{22}\right]\eta_1.\end{aligned}$$

When $\eta_1 = 0$, the noncentrality parameter for $LR_3$, $LM_3$, and $W_3$ is

$$\begin{aligned}(4.41) \qquad \theta_2 - \theta_1 &= \theta_3 \\ &= \xi_{21}'M_{22\cdot1}\Pi_{2*}\left[(\Pi_{2*}'M_{22\cdot1}\Pi_{2*})^{-1}\right. \\ &\qquad \left. - (\Pi_{2*}'M_{22\cdot1}\Pi_{2*} + \Omega_{*2}\Omega_{22}^{-1}\Omega_{2*})^{-1}\right]\Pi_{2*}'M_{22\cdot1}\xi_{21}.\end{aligned}$$

Thus $LR_3$, $LM_3$, and $W_3$ could be used to test $H_\xi$. The noncentrality parameter for $LR_1$, $LM_1$, and $W_1$ is smaller than the parameter for $LR_2 LM_2$, and $W_2$, but the number of degrees of freedom is also smaller. A comparison of asymptotic powers may depend on the significance level.

When $\eta_1 = 0$ and $\xi_1 = 0$, each of the three statistics $LR_2$, $LM_2$, and $W_2$ has a limiting distribution of $\chi^2$ with $G_0 \times \left[K_2 - (G_1 - G_0)\right]$ degrees of freedom. When $\eta_1 = 0$ and $\xi_1 = 0$, $LR_3$, $LM_3$, and $W_3$ are asymptotically distributed as $\chi^2$ with $G_0 \times G_2$ degrees of freedom. Some of these results for the case of $G_0 = G_1 = 1$ have been obtained previously under the assumptions that the disturbances are independently and identically distributed and there are no lagged endogenous variables. Furthermore, in this case it is known that Wu's statistic, Revankar's statistic, and the Revankar-Hartley statistic adjusted by their numbers of degrees of freedom are distributed as $F$ when the disturbances are normally distributed.

## 5. Proofs of Theorems

### 5.1. Asymptotic normality

**Proof of Theorem 3.** To prove the consistency of $\widetilde{B}_*$ we use the fact that

$$\frac{1}{T}Y'(P_Z - P_{Z_1})Y \xrightarrow{\text{P}} \Pi'(M - M_{.1}M_{11}^{-1}M_{1.})\Pi \qquad (5.1)$$
$$= \Pi_{2.}'M_{22\cdot 1}\Pi_{2.}\ .$$

From (5.1) we obtain

$$\widetilde{B}_* \xrightarrow{\text{P}} \left(\Pi_{2*}'M_{22\cdot 1}\Pi_{2*}\right)^{-1}\Pi_{2*}M_{22\cdot 1}\Pi_{20}\ . \qquad (5.2)$$

Since $\Pi_{2*}'M_{22\cdot 1}\Pi_{2*}$ is nonsingular and $\Pi_{20} = \Pi_{2*}B_*$, it follows that $\widetilde{B}_* \xrightarrow{\text{P}} B_*$. Then with $\widetilde{B} = (I, -\widetilde{B}_*')'$

$$\widetilde{\Gamma} = \left(\frac{1}{T}Z_1'Z_1\right)^{-1}\frac{1}{T}Z_1'(Z\Pi + V)\widetilde{B} \qquad (5.3)$$
$$\xrightarrow{\text{P}} M_{11}^{-1}(M_{11}\ M_{12})\begin{pmatrix}\Pi_{1.}\\ \Pi_{2.}\end{pmatrix}B$$
$$= \Gamma.$$

From (3.8) and (4.14) we derive

$$Y_*'P_ZY_*\widetilde{B}_* + Y_*'Z_1\widetilde{\Gamma} = Y_*'P_ZY_0, \qquad (5.4)$$

$$Z_1'Y_*\widetilde{B}_* + Z_1'Z_1\widetilde{\Gamma} = Z_1'Y_0. \qquad (5.5)$$

From these two equations we obtain

$$\frac{1}{T}\begin{pmatrix}Y_*'P_ZY_* & Y_*'Z_1\\ Z_1'Y_* & Z_1'Z_1\end{pmatrix}\sqrt{T}\begin{pmatrix}\widetilde{B}_* - B_*\\ \widetilde{\Gamma} - \Gamma\end{pmatrix} = \frac{1}{\sqrt{T}}\begin{pmatrix}Y_*'P_ZU + \frac{1}{\sqrt{T}}Y_*'P_ZZ\xi_1\\ Z_1'U + \frac{1}{\sqrt{T}}Z_1'Z\xi_1\end{pmatrix}. \qquad (5.6)$$

The matrix on the left-hand side of (5.6) converges in probability to $D'MD$. The right-hand side is

$$\frac{1}{\sqrt{T}}\begin{bmatrix}\frac{1}{T}(\Pi_{.*}'Z' + V')Z(\frac{1}{T}Z'Z)^{-1}Z'\\ Z_1'\end{bmatrix}U + \begin{bmatrix}\frac{1}{T}(\Pi_{.*}'Z' + V')Z\\ \frac{1}{T}Z_1'Z\end{bmatrix}\xi_1. \qquad (5.7)$$

The first vector in (5.7) converges in distribution to $N[0, \Sigma \otimes (D'MD)]$ and the second vector converges in probability to $D'M\xi_1$. Theorem 3 follows. ∎

To prove Corollary 1 we want to show that $\widehat{B}_*$ is asymptotically equivalent to $\widetilde{B}_*$. We use the following lemma.

**Lemma 1.** Under the local alternatives given by (4.13), for any $0 \leq \delta < 1$

$$T^{\delta}\lambda_i \xrightarrow{\text{P}} 0, \quad i = 1, 2, \ldots, G_0, \tag{5.8}$$

where $\lambda_i$ are the $G_0$ smallest roots of (3.2).

**Proof of Lemma 1.** From (3.2) we have

$$T\sum_{i=1}^{G_0} \lambda_i = \min_B \text{ tr } B'Y'(P_Z - P_{Z_1})YB \tag{5.9}$$

under the condition

$$\frac{1}{T}B'Y'\overline{P}_Z YB = I_{G_0}. \tag{5.10}$$

(This is a modification of Lemma A.3 in Anderson and Kunitomo (1989a) for a sum replacing a product of roots.) However,

$$\begin{aligned}(5.11)\quad \min_B \text{tr } B'Y'(P_Z - P_{Z_1})YB &\leq \text{tr } B'Y'(P_Z - P_{Z_1})YB \\ &= \text{tr } (B'\Pi'Z' + B'V')(P_Z - P_{Z_1})(Z\Pi B + VB) \\ &= \text{tr } \left(U + \frac{1}{\sqrt{T}}Z\xi_1\right)'(P_Z - P_{Z_1})\left(U + \frac{1}{\sqrt{T}}Z\xi_1\right) \\ &\leq \text{tr } \left(U + \frac{1}{\sqrt{T}}Z\xi_1\right)' P_Z \left(U + \frac{1}{\sqrt{T}}Z\xi_1\right)\end{aligned}$$

since $P_Z - P_{Z_1}$ is positive semidefinite. The minimum on the left-hand side of (5.11) is over matrices B satisfying (5.10) and the parameter matrix in the second term also satisfies (5.10). In turn, the right-hand side of (5.11) is not greater than

$$2 \text{ tr } U'P_Z U + \frac{2}{T} \text{ tr } \xi_1' Z' P_Z Z \xi_1. \tag{5.12}$$

The second term converges to $2\xi_1' M \xi_1$ and the first term converges in distribution to a Wishart matrix with covariance matrix $\Sigma$ and $K$ degrees of freedom. Then for $0 \leq \delta < 1$

$$T^{\delta} \sum_{i=1}^{G_0} \lambda_i \xrightarrow{\text{P}} 0. \tag{5.13}$$

Because $\lambda_i \geq 0$, $i = 1, \ldots, G_0$, we obtain (5.1). ∎

**Lemma 1′.** Under the conditions of Theorem 3

$$\sqrt{T}(\widehat{B}_* - \widetilde{B}_*) \xrightarrow{\text{P}} 0. \tag{5.14}$$

**Proof of Lemma 1′.** From (3.8) and $c'Y'\overline{P}_Z Y c = T$ we obtain

$$\frac{1}{T} Y'(P_Z - P_{Z_1}) Y C = \frac{1}{T} Y' \overline{P}_Z Y C \Lambda, \tag{5.15}$$

$$C' \frac{1}{T} Y' \overline{P}_Z Y C = I_{G_0}, \tag{5.16}$$

where $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_{G_0})$. As $T \to \infty$, the limits of these equations are

$$\Pi_{2\cdot}' M_{22\cdot 1} \Pi_{2\cdot} C = 0, \tag{5.17}$$

$$C' \Omega C = I_{G_0}. \tag{5.18}$$

The solutions $C$ to (5.17) and (5.18) are not unique; they are generated by multiplying a given solution on the right by orthogonal matrices. Define $\overline{C} = (\overline{C}_0', \overline{C}_*')'$ and the orthogonal $Q$ by

$$C = \begin{pmatrix} C_0 \\ C_* \end{pmatrix} = \begin{pmatrix} \overline{C}_0 \\ \overline{C}_* \end{pmatrix} Q \tag{5.19}$$

and the requirement that $\overline{C}_0$ be lower triangular. Similarly define $\widetilde{C} = (\widetilde{C}_0', \widetilde{C}_*')'$ by

$$Y_*'(P_Z - P_{Z_1}) Y \widetilde{C} = 0, \tag{5.20}$$

$\widetilde{C}' \overline{P}_Z \widetilde{C} = T I_{G_0}$, and $\widetilde{C}_0$ lower triangular. We can write the last $G_*$ components of (5.15) as

$$\frac{1}{T} Y_*'(P_Z - P_{Z_1}) Y C Q' = \frac{1}{T} Y' \overline{P}_Z Y C \Lambda Q'. \tag{5.21}$$

Subtraction of (5.20) from (5.21) yields

$$\frac{1}{T}Y_*'(P_Z - P_{Z_1})Y\sqrt{T}(\overline{C} - \widetilde{C}) = \frac{1}{T}Y'\overline{P}_Z Y\overline{C}\sqrt{T}\,Q\Lambda Q'. \tag{5.22}$$

By Lemma 1 and the fact that $Q\Lambda Q'$ is positive definite, $\sqrt{T}Q\Lambda Q' \xrightarrow{\text{P}} 0$. Hence $\sqrt{T}(\overline{C} - \widetilde{C}) \xrightarrow{\text{P}} 0$. The lemma follows because $\widehat{B}_* = -\overline{C}_*\overline{C}_0^{-1}$ and $\widetilde{B}_* = -\widetilde{C}_*\widetilde{C}_0^{-1}$. ∎

**Proof of Corollary 1.** We have

$$\sqrt{T}(\widehat{\Gamma} - \widetilde{\Gamma}) = \left(\frac{1}{T}Z_1'Z_1\right)^{-1}\frac{1}{T}Z_1'(Z\Pi + V)\sqrt{T}(\widehat{B} - \widetilde{B}) \xrightarrow{\text{P}} 0. \tag{5.23}$$

∎

### 5.2. Tests of overidentification

In order to prove Theorem 4, we give the following two lemmas on the convergence of two random matrices.

**Lemma 2.** Under the local alternatives given by (4.13),

$$\frac{1}{T}\widehat{B}'Y\overline{P}_Z Y\widehat{B} \xrightarrow{\text{P}} \Sigma, \tag{5.24}$$

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

**Proof of Lemma 2.** From Theorem 3 $\widehat{B} \xrightarrow{\text{P}} B$ as $T \to \infty$. By Theorem 2 we have

$$\frac{1}{T}V'\overline{P}_Z V = \frac{1}{T}V'V - \left(\frac{1}{T}V'Z\right)\left(\frac{1}{T}Z'Z\right)^{-1}\left(\frac{1}{T}Z'V\right) \xrightarrow{\text{P}} \Omega. \tag{5.25}$$

∎

**Lemma 3.** Under the local alternatives given by (4.13),

$$\widehat{B}'Y'(P_Z - P_{Z_1})Y\widehat{B} \xrightarrow{\text{L}} W_{G_0}(\Sigma, L, \theta_1), \tag{5.26}$$

where $W_{G_0}(\Sigma, L, \theta_1)$ is the noncentral Wishart distribution with $L$ degrees of freedom, covariance matrix $\Sigma$, and noncentrality matrix $\theta_1$ given by (4.27).

**Proof of Lemma 3.** Since $\widehat{B} = B + (\widehat{B} - B)$, the left-hand side of (5.26) is decomposed as

$$
\begin{aligned}
(5.27) \qquad & B'Y'(P_Z - P_{Z_1})YB + B'Y'(P_Z - P_{Z_1})Y(\widehat{B} - B) \\
& + (\widehat{B} - B)'Y'(P_Z - P_{Z_1})YB + (\widehat{B} - B)'Y'(P_Z - P_{Z_1})Y(\widehat{B} - B).
\end{aligned}
$$

Since $YB = (Z\Pi + V)B = Z\xi_1 + U = Z_1\Gamma + U^*$, where $U^* = U + \frac{1}{\sqrt{T}} Z\xi_1$, the first term of (5.24) is

$$
(5.28) \qquad U^{*\prime}(P_Z - P_{Z_1})U^*.
$$

By the standardization of $\widehat{B}_0 = B_0 = I_{G_0}$, we have $Y(\widehat{B} - B) = -Y_*(\widehat{B}_* - B_*)$. Then $\left[\text{since}\,(P_Z - P_{Z_1})Z_1 = (\overline{P}_{Z_1} - \overline{P}_Z)Z_1 = 0\right]$

$$
\begin{aligned}
(5.29) \quad (P_Z - P_{Z_1})Y(\widehat{B} - B) &= (P_Z - P_{Z_1})(Y_*, Z_1)\begin{bmatrix} -(\widehat{B}_* - B_*) \\ -(\widehat{\Gamma} - \Gamma) \end{bmatrix} \\
&= -(P_Z - P_{Z_1})\left[\frac{1}{\sqrt{T}} ZD + \frac{1}{\sqrt{T}}(V_*, 0)\right]\begin{bmatrix} \sqrt{T}(\widehat{B}_* - B_*) \\ \sqrt{T}(\widehat{\Gamma} - \Gamma) \end{bmatrix}.
\end{aligned}
$$

By Theorem 3 $(P_Z - P_{Z_1})Y(\widehat{B} - B)$ is asymptotically equivalent to

$$
(5.30) \qquad -(P_Z - P_{Z_1})\left[\frac{1}{\sqrt{T}} ZD + \frac{1}{\sqrt{T}}(V_*, 0)\right]\left(D'\frac{1}{T}Z'ZD\right)^{-1}\left(\frac{1}{\sqrt{T}} ZD\right)' U^*.
$$

Note $P_Z P_{ZD} = P_{ZD}$ and $P_{Z_1} P_{ZD} = P_{Z_1}$. Then the second term of (5.27) is asymptotically equivalent to

$$
(5.31) \qquad -U^{*\prime}(P_Z - P_{Z_1})P_{ZD}U^* = -U^{*\prime}(P_{ZD} - P_{Z_1})U^*.
$$

By similar consideration of the third and fourth terms of (5.24), we find that (5.27) is asymptotically equivalent to

$$
\begin{aligned}
(5.32) \quad & U^{*\prime}(P_Z - P_{ZD})U^* = \\
& \frac{1}{\sqrt{T}} U^{*\prime} Z \left(\frac{1}{T}Z'Z\right)^{-1/2}\left[I_K - \left(\frac{1}{T}Z'Z\right)^{1/2} D \left(\frac{1}{T}D'Z'ZD\right)^{-1} D' \left(\frac{1}{T}Z'Z\right)^{1/2}\right] \\
& \qquad \left(\frac{1}{T}Z'Z\right)^{-1/2} \frac{1}{\sqrt{T}} Z'U^*.
\end{aligned}
$$

Since the matrix in brackets in (5.32) is idempotent and of rank $K - (G_* + K_1) = K_2 - G_*$ (the rank of $P_Z - P_{ZD}$), we obtain (5.26) by applying Theorem 1 to $Z'U^*/\sqrt{T}$. ∎

**Proof of Theorem 4.** From $\widehat{B} = CC_0^{-1}$ we obtain

$$
\begin{aligned}
(5.33) \qquad W_1' &= \operatorname{tr} \Lambda \\
&= \operatorname{tr}\left(C' \frac{1}{T} Y' \overline{P}_Z Y C\right)^{-1} C' Y' (P_Z - P_{Z_1}) Y C \\
&= \operatorname{tr} (\widehat{B}' Y' \overline{P}_Z Y \widehat{B})^{-1} \widehat{B} Y' (P_Z - P_{Z_1}) Y \widehat{B} \\
&= \operatorname{tr}\left(\Sigma^{-\frac{1}{2}} \widehat{B} Y' \overline{P}_Z Y \widehat{B} \Sigma^{-\frac{1}{2}}\right)^{-1} \Sigma^{-\frac{1}{2}} \widehat{B} Y' (P_Z - P_{Z_1}) Y \widehat{B} \Sigma^{-\frac{1}{2}}
\end{aligned}
$$

converges in distribution to the trace of a noncentral Wishart matrix with covariance matrix $I$, $L$ degrees of freedom, and noncentrality parameter $\Sigma^{-\frac{1}{2}}\theta_1\Sigma^{-\frac{1}{2}}$. Then Theorem 4 follows for $W_1'$. Since $LR_1$, $LM_1$, and $W_1$ are asymptotically equivalent to $W_1'$ we have Theorem 4. ∎

### 5.3. Tests of exogeneity and overidentification

If $Y = (Y_1, Y_2)$ is normally distributed, the model for the conditional distribution of $Y_1$ given $Y_2$ is

$$(5.34) \qquad Y_1 = Y_2\rho + Z_1\Pi_{11}^{**} + Z_2\Pi_{21}^{**} + V_1^*,$$

where

$$(5.35) \qquad \rho = \Omega_{22}^{-1}\Omega_{21},$$

$$(5.36) \qquad \Pi_{11}^{**} = \Pi_{11} - \Pi_{12}\rho,$$

$$(5.37) \qquad \Pi_{21}^{**} = \Pi_{21} - \Pi_{22}\rho.$$

The random matrix

$$(5.38) \qquad V_1^* = V_1 - V_2\rho$$

is normally distributed with mean 0; each row has covariance matrix

$$(5.39) \qquad \Omega_{11\cdot 2} = \Omega_{11} - \Omega_{12}\Omega_{22}^{-1}\Omega_2 \, .$$

In (5.34) $Y_2$ is treated as predetermined.

The set of structural equations under the null hypothesis is

$$(5.40)\quad \begin{aligned} Y_1 B_1 &= Y_2 \rho B_1 + Z_1 \Pi_1^{**} B_1 + Z_2 \Pi_2^{**} B_1 + V_1^* B_1 \\ &= Y_2 \rho B_1 + Z_1(\Pi_{11} - \Pi_{12}\rho) B_1 + Z_2(\Pi_{21} - \Pi_{22}\rho) B_1 + (V_1 - V_2\rho) B_1 \\ &= Y_2 B_2 + Z_1 \Gamma + U. \end{aligned}$$

This has the same form as (2.1) and (5.34) has the same form as (2.2) with the following correspondences:

| (2.1) and (2.2) | (5.34) and (5.40) |
|---|---|
| $Y$ | $Y_1$ |
| $G$ | $G_1$ |
| $Z_1$ | $(Y_2, Z_1)$ |
| $K_1$ | $K_1 + G_2$ |
| $V$ | $V_1^*$ |
| $\Omega$ | $\Omega_{11\cdot 2}$ |
| $B$ | $B_1$ |
| $\Gamma$ | $(B_2, \Gamma)$ |
| $Z_2$ | $Z_2$ |
| $Z$ | $(Y_2, Z) = R$ |
| $\Pi_{1\cdot}$ | $\begin{bmatrix} \rho \\ \Pi_{11}^{**} \end{bmatrix}$ |
| $\Pi_{2\cdot}$ | $\Pi_{21}^{**}$ |
| $\Pi_{2*}$ | $\Pi_{21}^{**} \begin{pmatrix} 0 \\ I_{G_1} - G_0 \end{pmatrix}$ |
| $\begin{bmatrix} \Pi_{1\cdot} \\ \Pi_{2\cdot} \end{bmatrix} B = \begin{bmatrix} \Gamma \\ 0 \end{bmatrix} + \frac{1}{\sqrt{T}} \begin{bmatrix} \xi_{11} \\ \xi_{21} \end{bmatrix}$ | $\begin{bmatrix} \rho \\ \Pi_{11}^{**} \\ \Pi_{21}^{**} \end{bmatrix} B_1 = \begin{bmatrix} B_2 \\ \Gamma \\ 0 \end{bmatrix} + \frac{1}{\sqrt{T}} \begin{bmatrix} \eta_1 \\ \xi_{11} + \Pi_{12}\eta_1 \\ \xi_{21} + \Pi_{22}\eta_1 \end{bmatrix}$ |
| $M = \operatorname*{plim}_{T\to\infty} \frac{1}{T} Z'Z$ | $\begin{aligned} Q &= \operatorname*{plim}_{T\to\infty} \frac{1}{T} \begin{bmatrix} Y_2'Y_2 & Y_2'Z \\ Z'Y_2' & Z'Z \end{bmatrix} \\ &= \begin{bmatrix} \Pi_{\cdot 2}' M \Pi_{\cdot 2} + \Omega_{22} & \Pi_{\cdot 2}' M \\ M\Pi_{\cdot 2} & M \end{bmatrix} \\ &= \begin{bmatrix} \Pi_{\cdot 2}' \\ I_K \end{bmatrix} M(\Pi_{\cdot 2}, I_K) + \begin{bmatrix} \Omega_{22} & 0 \\ 0 & 0 \end{bmatrix} \end{aligned}$ |

$$D = \left[\Pi_{\cdot *}, \begin{pmatrix} I_{K_1} \\ 0 \end{pmatrix}\right] \qquad F = \left[\begin{pmatrix} \rho \\ \Pi_{11}^{**} \\ \Pi_{21}^{**} \end{pmatrix} \begin{pmatrix} 0 \\ I_{G_1-G_0} \end{pmatrix}, \begin{pmatrix} I_{G_2+K_1} \\ 0 \end{pmatrix}\right]$$

$$M_{22\cdot 1} \qquad M_{22\cdot 1}^{*} = M_{22\cdot 1} - M_{22\cdot 1}\Pi_{22}\left(\Pi_{22}' M_{22\cdot 1}\Pi_{22} + \Omega_{22}\right)^{-1}\Pi_{22}' M_{22\cdot 1}.$$

Finally

$$M_{22\cdot 1} - M_{22\cdot 1}\Pi_{2*}\left(\Pi_{2*}' M_{22\cdot 1}\Pi_{2*}\right)^{-1}\Pi_{2*}' M_{2*}' M_{22\cdot 1} \tag{5.41}$$

corresponds to

$$M_{22\cdot 1}^{*} - M_{22\cdot 1}^{*}\Pi_{21}^{**}\begin{pmatrix} 0 \\ I_{G_1-G_0} \end{pmatrix}\left[(0, I_{G_1-G_0})\Pi_{21}^{**\prime} M_{22\cdot 1}^{*}\Pi_{21}^{**}\begin{pmatrix} 0 \\ I_{G_1-G_0} \end{pmatrix}\right]^{-1} (0, I_{G_1-G_0})\Pi_{21}^{**\prime} M_{22\cdot 1}^{*}. \tag{5.42}$$

When (5.34) is multiplied by $B_1$ under the local alternatives

$$\begin{aligned} Y_1 B_1 &= Y_2\rho B_1 + Z_1(\Pi_{11} - \Pi_{12}\rho)B_1 + Z_2(\Pi_{21} - \Pi_{12}\rho)B_1 \\ &\quad + (V_1 - V_2\rho)B, \\ &= Y_2 B_2 + Z_1\Gamma + \frac{1}{\sqrt{T}} Z\xi_1 + U - \frac{1}{\sqrt{T}} V_2\eta_1 \\ &= Y_2 B_2 + Z_1\Gamma + U_1^{*}, \end{aligned} \tag{5.43}$$

where

$$U_1^{*} = U + \frac{1}{\sqrt{T}}\,\Xi, \tag{5.44}$$

and

$$\Xi = Z\xi_1 + V_2\eta_1 \tag{5.45}$$

under the local alternatives (4.28). Then the limiting distribution of $\widehat{B}_1' Y_1'(P_{Y_2,Z} - P_{Y_2,Z_1})Y_1\widehat{B}_1$ is the limiting distribution of

$$\begin{aligned} &U_1^{*\prime}(P_R - P_{RF})U_1^{*} \\ &= \left(\frac{1}{\sqrt{T}} U_1^{*\prime} R\right)\left(\frac{1}{T}R'R\right)^{-1/2}\left[I_{G_2+K} - \left(\frac{1}{T}R'R\right)^{1/2}\right. \\ &\quad \left.\times F\left(F'\frac{1}{T}R'RF\right)^{-1} F'\left(\frac{1}{T}R'R\right)^{1/2}\right]\left(\frac{1}{T}R'R\right)^{-1/2}\left(\frac{1}{\sqrt{T}}R'U_1^{*}\right). \end{aligned} \tag{5.46}$$

This limiting distribution is the distribution of

$$A_1 = C'Q^{-\frac{1}{2}}(I_{G_2+K} - Q^{\frac{1}{2}}F(F'QF)^{-1}F'Q^{\frac{1}{2}})Q^{-\frac{1}{2}}C, \tag{5.47}$$

where vec $C$ has the distribution $N(\text{vec}\ \zeta_1, \Sigma \otimes Q)$ and

$$\zeta_1 = \begin{pmatrix} \Pi'_{\cdot 2} \\ I_K \end{pmatrix} M\xi_1 + \begin{pmatrix} \Omega_{22}\eta_1 \\ 0 \end{pmatrix}. \tag{5.48}$$

Since the rank of $A_1$ is $G_2 + K - (G_2 + G_1 - G_0 + K_1) = K_2 - (G_1 - G_0)$ and $\widehat{\Sigma}$ is a consistent estimator of $\Sigma$, we obtain the asymptotic distribution of $LR_2$, $LM_2$, and $W_2$ in Theorem 5 by applying the same argument as in the proof of Theorem 4.

To justify the application of Theorem 3 we have to show that

$$\frac{1}{\sqrt{T}}\begin{pmatrix} Y_2' \\ Z' \end{pmatrix} V_1^* \xrightarrow{\text{L}} N(0, \Omega_{11\cdot 2} \otimes Q); \tag{5.49}$$

that is, we have to show that the conditions of Theorem 3 are met when $v_t$ is replaced by $v_{1t}^* = v_{1t} - \rho' v_{2t}$ and $z_t$ is replaced by $(y_{2t}', z_t')'$. The condition corresponding to (4.5) is met by the facts $\frac{1}{T}Z'Z \xrightarrow{\text{P}} M$ and $\frac{1}{T}V'V \xrightarrow{\text{P}} \Omega$. The condition corresponding to (4.6) is met by (4.6) and $\max_{t=1,\ldots,T} \|v_t\|^2 \xrightarrow{\text{P}} 0$, which is a consequence of (4.8). Clearly $\mathcal{E}(v_{1t}^* | \mathcal{F}_{t-1}) = 0$,

$$\begin{aligned} \mathcal{E}(v_{1t}^* v_{1t}^{*\prime} | \mathcal{F}_{t-1}) &= \Omega_{11}(t) - \rho'\Omega_{21}(t) - \Omega_{12}(t)\rho + \rho'\Omega_{22}(t)\rho \\ &= \Omega_{11\cdot 2}(t), \end{aligned} \tag{5.50}$$

$$\frac{1}{T}\sum_{t=1}^{T} \Omega_{11\cdot 2}(t) \xrightarrow{\text{P}} \Omega_{11\cdot 2}, \tag{5.51}$$

which corresponds to (4.10). From the assumptions of Theorem 5

$$\frac{1}{T}\sum_{t=1}^{T} \left[\Omega_{11\cdot 2}(t) \otimes \begin{pmatrix} y_{2t} \\ z_t \end{pmatrix} (y_{2t}', z_t')\right] \xrightarrow{\text{P}} \Omega_{11\cdot 2} \otimes Q. \tag{5.52}$$

Finally (4.8) for $v_{1t}^*$ follows from (4.8) for $v_t$. This justifies the application of Theorem 3. ∎

### 5.4. Tests of exogeneity against alternative of overidentification

**Proof of Theorem 6 (a).** First, we consider the asymptotic distribution of $LR_3$. Let $J_3' = (0, I_K)$ be a $K \times (G - G_0 + K)$ choice matrix and $Z = RJ_3$. Then from the derivation the derivation of (5.36), the limiting distribution of $\widehat{B}'Y'(P_Z - P_{Z_1})Y\widehat{B}$ is the limiting distribution of $U^{*\prime}(P_Z - P_{ZD})U^*$, which can be written as

(5.53)

$$\Big(\frac{1}{\sqrt{T}}U_1^{*\prime}R\Big)J_3\Big\{\Big[J_3'\Big(\frac{1}{T}R'R\Big)J_3\Big]^{-1} - D\Big[D'J_3'\Big(\frac{1}{T}R'R\Big)J_3D\Big]^{-1}D'\Big\}J_3'\Big(\frac{1}{\sqrt{T}}R'U_1^*\Big).$$

The limiting distribution of (5.53) is the distribution of

$$\begin{aligned}(5.54)\quad A_2 &= C'J_3M^{-\frac{1}{2}}(I_K - M^{\frac{1}{2}}D(D'MD)^{-1}D'M^{\frac{1}{2}})M^{-\frac{1}{2}}J_3'C \\ &= C'Q^{-\frac{1}{2}}\Big\{Q^{\frac{1}{2}}\begin{bmatrix}0 & 0\\ 0 & M^{-1}\end{bmatrix}Q^{\frac{1}{2}} - Q^{\frac{1}{2}}\begin{bmatrix}0 & 0\\ 0 & D(D'MD)^{-1}D'\end{bmatrix}Q^{\frac{1}{2}}\Big\}Q^{-\frac{1}{2}}C.\end{aligned}$$

Let

$$(5.55)\qquad P_1 = I_{G_2+K} - Q^{\frac{1}{2}}F(F'QF)^{-1}F'Q^{\frac{1}{2}},$$

$$(5.56)\quad P_2 = Q^{\frac{1}{2}}\begin{pmatrix}I_{G_2} & 0\\ 0 & M^{-\frac{1}{2}}\end{pmatrix}\begin{bmatrix}0 & 0\\ 0 & I_K - M^{\frac{1}{2}}D(D'MD)^{-1}D'M^{\frac{1}{2}}\end{bmatrix}\begin{pmatrix}I_{G_2} & 0\\ 0 & M^{-\frac{1}{2}}\end{pmatrix}Q^{\frac{1}{2}}.$$

Then

$$(5.57)\qquad A_i = C'Q^{-\frac{1}{2}}P_iQ^{-\frac{1}{2}}C, \qquad i = 1, 2.$$

We see that $P_1^2 = P_1$, $P_2^2 = P_2$, $P_1P_2 = P_2P_1 = P_2$ (since $P_2QF = 0$), and $(P_1 - P_2)^2 = P_1 - P_2$. Since tr $P_1 = G_2 + K - (G_1 - G_0 + G_2 + K_1) = G_1 - G_0 + K_2$ and tr $P_2 = K - (G_* + K_1) = G - G_0 + K_2$, it follows that $\mathrm{tr}(P_1 - P_2) = G_2$. Hence, $A_1 - A_2$ has the distribution $W_{G_1}(\Sigma, G_2, \theta_3)$, where

$$(5.58)\qquad \theta_3 = \zeta_1'Q^{-\frac{1}{2}}(P_1 - P_2)Q^{-\frac{1}{2}}\zeta_1 = \theta_2 - \theta_1.$$

The limiting distribution of $LR_3$ follows because it has the same limiting distribution as $T\sum_{i=1}^{G_0}\lambda_i^* - T\sum_{i=1}^{G_0}\lambda_i$. ∎

**Proof of Theorem 6(b).** Next we obtain the asymptotic distribution of $LM_3$. In Anderson and Kunitomo (1989a) the Lagrange multiplier matrix is

$$(5.59)\qquad \Lambda_0 = -Y_2'P_Z\overline{P}_{Y_2,Z_1}Y_1\widehat{B}_1\widehat{\Sigma}^{-1},$$

where $\widehat{B}_1$ and $\widehat{\Sigma}$ are the maximum likelihood estimators of $B_1$ and $\Sigma$ under the null hypothesis. The matrix $\Lambda_0$ is asymptotically equivalent to

$$(5.60)\qquad \Lambda_0^* = -Y_2'P_Z\overline{P}_{Y_2,Z_1}Y_1B_1\Sigma^{-1} - Y_2'P_Z\overline{P}_{Y_2,Z_1}Y_1(\widehat{B}_1 - B_1)\Sigma^{-1}.$$

Under the local alternatives (4.13) and (4.28), the first term of $\Lambda_0^*\Sigma/\sqrt{T}$ is $Y_2'P_Z\overline{P}_{Y_2,Z_1}U_1^*/\sqrt{T}$ and the second term is

$$(5.61)\qquad \frac{1}{\sqrt{T}}Y_2'P_Z\overline{P}_{Y_2,Z_1}(Y_1,Y_2,Z_1)\begin{bmatrix}0\\-(\widehat{B}_*-B)\\-(\widehat{\Gamma}-\Gamma)\end{bmatrix}$$
$$=\frac{1}{\sqrt{T}}Y_2'P_Z\overline{P}_{Y_2,Z_1}\left[\frac{1}{\sqrt{T}}RF+\frac{1}{\sqrt{T}}(V_1^*J_1,0,0)\right]\sqrt{T}\begin{bmatrix}-(\widehat{B}_*-B_*)\\-(\widehat{\Gamma}-\Gamma)\end{bmatrix}.$$

The limiting distribution of $\sqrt{T}\begin{bmatrix}\widehat{B}_*-B_*\\\widehat{\Gamma}-\Gamma\end{bmatrix}$ is the limiting distribution of

$$(5.62)\qquad \left(\frac{1}{T}F'R'RF\right)^{-1}\frac{1}{\sqrt{T}}(RF)'U_1^*.$$

Then by the similar argument as in the proof of Lemma 3 the limiting distribution of vec $\Lambda_0^*/\sqrt{T}$ is the limiting distribution of

$$(5.63)\qquad \frac{1}{\sqrt{T}}Y_2'P_Z\overline{P}_{Y_2,Z_1}U_1^*-\frac{1}{\sqrt{T}}Y_2'P_Z\overline{P}_{Y_2,Z_1}P_{RF}U_1^*$$
$$=\frac{1}{\sqrt{T}}Y_2'(P_Z-P_ZP_{Y_2,Z_1}-P_ZP_{RF}+P_ZP_{Y_2,Z_1}P_{RF})U_1^*$$
$$=\frac{1}{\sqrt{T}}Y_2'\overline{P}_Z\overline{P}_{RF}U_1^*$$
$$=\frac{1}{\sqrt{T}}V_2'\overline{P}_Z\overline{P}_{RF}\left(U+\frac{1}{\sqrt{T}}\Xi\right),$$

where $U_1^*$ is given by (5.35) and $u_t^*$ is the $t$-th row of $U_1^*$. (We have used the relations $P_{Y_2,Z_1}P_{RF}=P_{Y_2,Z_1}$ and $Y_2'P_Z\overline{P}_{RF}=-Y_2'\overline{P}_Z\overline{P}_{RF}$.) Since the first term of $Y_1^*$ is asymptotically uncorrelated to $P_{Y_2,Z_1}$ and $P_{X^*}$, the noncentrality parameter is given by $\delta_4^2=\mathrm{tr}(\theta_4\Sigma^{-1})$, where

$$(5.64)\qquad \theta_4=\operatorname*{plim}_{T\to\infty}\frac{1}{T}\Xi'\,\overline{P}_{RF}\overline{P}_ZY_2\left(\frac{1}{T}Y_2'\overline{P}_Z\overline{P}_{RF}\overline{P}_ZY_2\right)^{-1}Y_2'\overline{P}_Z\overline{P}_{RF}\,\Xi$$
$$=\operatorname*{plim}_{T\to\infty}\frac{1}{T}\Xi'\left(\overline{P}_{RF}-\overline{P}_{X^*}\right)\Xi$$
$$=\operatorname*{plim}_{T\to\infty}\frac{1}{T}\Xi'\left(P_{X^*}-P_{RF}\right)\Xi,$$

where $X^*=(RF,\overline{P}_ZY_2)$ is a $T\times(G_1-G_0+G_2+K_1+G_2)$ matrix. The second equality is based on Lemma 6 of Anderson and Kunitomo (1989a). Since $V'V/T\to\Omega$ in probability and

$$(5.65)\qquad \frac{1}{T}R'\overline{P}_ZY_2=\frac{1}{T}\begin{pmatrix}V_2'\\0\end{pmatrix}\overline{P}_ZV_2\xrightarrow{\ p\ }\begin{pmatrix}\Omega_{22}\\0\end{pmatrix},$$

we have

$$P = \operatorname*{plim}_{T\to\infty} \frac{1}{T} X^{*\prime} X^* = \begin{bmatrix} F'QF & F' \begin{pmatrix} \Omega_{22} \\ 0 \end{pmatrix} \\ (\Omega_{22}, 0)F & \Omega_{22} \end{bmatrix}. \tag{5.66}$$

Similarly,

$$\begin{aligned} \zeta_2 = \operatorname*{plim}_{T\to\infty} \frac{1}{T} X^{*\prime} \Xi &= \operatorname*{plim}_{T\to\infty} \frac{1}{T} \begin{bmatrix} F' \begin{pmatrix} Y_2' \\ Z' \end{pmatrix} \\ Y_2'\overline{P_Z} \end{bmatrix} (Z\xi_1 + V_2\eta_1) \\ &= \begin{bmatrix} F' \begin{pmatrix} \Pi_{\cdot 2}' \\ I_K \end{pmatrix} \\ 0 \end{bmatrix} M\xi_1 + \begin{bmatrix} F' \begin{pmatrix} I_{G_2} \\ 0 \end{pmatrix} \\ I_{G_2} \end{bmatrix} \Omega_{22}\eta_1 \\ &= \begin{pmatrix} D' \\ 0 \end{pmatrix} M\xi_1 + \begin{pmatrix} J_1'\rho' \\ I_{G_2} \\ 0 \\ I_{G_2} \end{pmatrix} \Omega_{22}\eta_1, \end{aligned} \tag{5.67}$$

where $J_4' = \left(I_{G_2}, 0, I_{G_2}\right)'$ is a $G_2 \times (G_2 + K_1 + G_2)$ choice matrix. We note that $|P| = |D'MD||\Omega_{22}| \neq 0$ if both $D'MD$ and $\Omega$ are nonsingular. Express $RF = X^* J_2$, where $J_2' = (I_{G-G_0+K_1}, 0)$ is a $(G - G_0 + K_1) \times (G - G_0 + K_1 + G_2)$ choice matrix. Then

$$\theta_4 = \zeta_2'(P^{-1} - J_2(J_2'PJ_2)^{-1}J_2')\zeta_2. \tag{5.68}$$

By applying Theorem 2 to (5.68), $LM_3$ is asymptotically distributed with $W_{G_0}(\Sigma, G_2, \theta_4)$. Under the local alternatives (4.13) and (4.20), the estimator of $\widehat{\Sigma}$ is written as

$$\widehat{\Sigma} = \frac{1}{T}\widehat{B}_1'\left(V_1^* + \frac{1}{\sqrt{T}}\Xi\right)'\overline{P}_X\left(V_1^* + \frac{1}{\sqrt{T}}\Xi\right)\widehat{B}_1. \tag{5.69}$$

As $T \to \infty$, $\widehat{\Sigma} \to \Sigma$ in probability, which is the limit of the covariance matrix of the rows of $U_1^*$. Some matrix algebra shows that $\theta_4 = \theta_3$ and, hence, $\delta_4^2 = \delta_3^2$.

**Proof of Theorem 6(c).** We now turn to the asymptotic distribution of $W_3$. We write

$$\begin{aligned} \frac{1}{\sqrt{T}} Y_2'\overline{P}_Z Y\widehat{B} &= \sqrt{T}J_5'\Omega B + J_5'\Omega\sqrt{T}(\widehat{B} - B) \\ &+ \sqrt{T}J_5\left(\frac{1}{T}Y'\overline{P}_{Z_1}YB - \Omega B\right) + \sqrt{T}J_5\left(\frac{1}{T}Y'\overline{P}_{Z_1}YB - \Omega B\right)(\widehat{B} - B), \end{aligned} \tag{5.70}$$

where $J_5' = (0, I_{G_2})$ is the $G_2 \times (G_1 + G_2)$ choice matrix. The limiting distribution of $J_5'\Omega\sqrt{T}(\widehat{B} - B)$ is the limiting distribution of

$$-(\Omega_{21}J_1, \Omega_{22}, 0)(D'MD)^{-1}D'\frac{1}{\sqrt{T}}Z'U_1^*. \tag{5.71}$$

Similarly,

$$\sqrt{T}J_5\left(\frac{1}{T}Y'\overline{P}_{Z_1}YB - \Omega B\right) = \frac{1}{\sqrt{T}}V_2'\overline{P}_{Z_1}U^* - \eta_1, \tag{5.72}$$

and the limiting covariance matrix of vec $\left(\frac{1}{\sqrt{T}}V_2'\overline{P}_2U^* - \eta_1\right)$ is $\Sigma \otimes \Omega_{22}$. Because the last term in (5.70) is asymptotically negligible and $U^* = V_1^*B_1 + \frac{1}{\sqrt{T}}\Xi$, the noncentrality parameter is

$$\begin{aligned}\zeta_3 &= \Omega_{22}\eta_1 - (\Omega_{21}J_1, \Omega_{22}, 0)(D'MD)^{-1}D' \operatorname*{plim}_{T\to\infty}\frac{1}{T}Z'\,\Xi \\ &\quad + \operatorname*{plim}_{T\to\infty}\left(\frac{1}{T}V_2'\overline{P}_Z\,\Xi - \eta_1\right) \\ &= \Omega_{22}\eta_1 - (\Omega_{21}J_1, \Omega_{22}, 0)(D'MD)^{-1}D'M\xi_1.\end{aligned} \tag{5.73}$$

Furthermore

$$\begin{aligned}&\widehat{\Omega}_{22}(\hat{\rho}J_1, I_{G_2}, 0)\left(\frac{1}{T}\widehat{D}'Z'ZD\right)^{-1}\begin{bmatrix} J_1'\hat{\rho}' \\ I_{G_2} \\ 0 \end{bmatrix}\widehat{\Omega}_{22} + \widehat{\Omega}_{22} \\ &\qquad \xrightarrow{\text{p}} \Omega_{22}(\rho J_1, I_{G_2}, 0)(D'MD)^{-1}\begin{bmatrix} J_1'\rho' \\ I_{G_2} \\ 0 \end{bmatrix}\Omega_{22} + \Omega_{22}.\end{aligned} \tag{5.74}$$

Further matrix algebra verifies that the noncentrality parameter is $\delta_3^2$. ∎

## References


Anderson, T. W. (1951). Estimating linear restrictions on regression coefficients for multivariate normal distributions, *Annals of Mathematical Statistics,* **22**, 327–351.

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

Anderson, T. W., and Kunitomo, Naoto (1989a). Tests of overidentification and exogeneity in simultaneous equation models, Technical Report No. 36, Econometric Workshop, Stanford University.

Anderson, T. W. , and Kunitomo, Naoto (1989b). Asymptotic robustness in regression and autoregression based on Lindeberg conditions, Technical Report No. 23, ARO Contract No. DAAL03-89-K-0033, Department of Statistics, Stanford University.

Anderson, T. W., and Rubin, Herman (1949). Estimation of the parameters of a single equation in a complete system of stochastic equations, *Annals of Mathematical Statistics,* **20**, 46–63.

Anderson, T. W., and Rubin, Herman (1950). The asymptotic properties of estimates of the parameters of a single equation in a complete system of stochastic equations, *Annals of Mathematical Statistics,* **21**, 570–582.

Basmann, R. L. (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.

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

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

Holly, Alberto. (1987). Testing for exogeneity: a survey, *Economics 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 Hodoshima, J. (1980), Finite sample properties of the tests for independnece 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, Chapter 6, *Studies in Econometric Method* (W. C. Hood and T. C. Koopmans, eds.), Yale University Press.

Kunitomo, N. (1988). Approximate distributions and power of the test statistics for overidentifying restrictions in a system of simultaneous equations, *Econometric Theory,* **4**, 248–274.

Lai, Tze Leung, and Robbins, Herbert (1981). Consistency and asymptotic efficiency of slope estimates in stochastic approximation schemes, *Zeitschrift fur Wahrscheinlichkeitstheorie und verwandte Gebiete,* **56**, 329–360.

Mann, H. B., and Wald, A. (1943). On the statistical treatment of linear stochastic difference equations, *Econometrica,* **11**, 173–220.

Nakamura, A., and Nakamura, M. (1982). On the relationships among several specification error tests presented by Durbin, Wu and Hausman, *Econometrica,* **49**, 1583–1588.

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

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

Smith, R. (1989). Wald tests for the independence of stochastic variables and disturbances of a single linear stochastic simultaneous equation, *Economics Letters,* **17**, 87–90.

Theil, H. (1971). *Principles of Econometrics*, John Wiley & Sons.

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. (1974). Alternative tests of independence between stochastic regressors and disturbances: finite sample results, *Econometrica,* **42**, 529–546.