Testing a Hypothesis about a Structural Coefficient in a Simultaneous Equation Model for Known Covariance Matrix *

T. W. Anderson †

and

Naoto Kunitomo ‡

In honor of Raul P. Mentz

June 30, 2009

*This is TWA09-6-30-5.

†Department of Statistics and Department of Economics, Stanford University

‡Graduate School of Economics, University of Tokyo

1. Introduction

There is a considerable econometrics literature about *estimating* coefficients of one equation in a simultaneous equation model (Anderson and Rubin (1949), for example). More recently, econometicians have been studying the problem of *testing* a hypothesis concerning the coefficients of a single equation. See Andrews, Moreira and Stock (2006), for example. In a previous study Anderson and Kunitomo (2007) have derived the likelihood ratio test of a vector of coefficients when it is assumed that the single equation is identified and the covariance is estimated. In the present paper we derive the likelihood ratio test in the case of the error covariance matrix known.

2. The model

Let $\mathbf{Y}$ be a $T \times G$ matrix of endogenous or dependent variables; let $\mathbf{Z}$ be a $T \times K$ matrix of exogenous or independent variables; and let $\mathbf{V}$ be a $T \times G$ matrix of unobservable errors or shocks, which are statistically independent of $\mathbf{Z}$. The rows of $\mathbf{V}$ are assumed independent, the t-th row of which has mean $\mathcal{E}(\mathbf{v}_t) = \mathbf{0}$ and covariance $\mathcal{E}(\mathbf{v}_t\mathbf{v}_t') = \mathbf{\Omega}$. (The matrix $\mathbf{\Omega}$ is assumed known.) The ”reduced form” of the model is

$$\mathbf{Y} = \mathbf{Z\Pi} + \mathbf{V} \ , \tag{2.1}$$

where $\mathbf{\Pi}$ is a $K \times G$ matrix of parameters. Statisticians will recognize (2.1) as a multivariate regression.

A single *structural* equation of interest may be written as

$$\mathbf{Y}\boldsymbol{\beta} = \mathbf{Z}\boldsymbol{\gamma} + \mathbf{u} \ , \tag{2.2}$$

where $\mathbf{u} = \mathbf{V}\boldsymbol{\beta}$. To distinguish this equation from other linear combinations of $\mathbf{Y}$ and $\mathbf{Z}$, some "identifying" assumptions concerning $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are made; in particular some components of $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ may be assumed $\mathbf{0}$. The relation between (2.1) and (2.2) implies

$$\boldsymbol{\Pi}\boldsymbol{\beta} = \boldsymbol{\gamma} \ . \tag{2.3}$$

The equation (2.1) is said to be *identified* by $0's$ if given $\boldsymbol{\Pi}$ (2.3) has a unique solution for $(\boldsymbol{\beta}', \boldsymbol{\gamma}')$ except a constant of proportionality.

In this paper first we pay special attention to the case that $\boldsymbol{\gamma} = 0$. Then (the relation (2.3) is

$$\boldsymbol{\Pi}\boldsymbol{\beta} = \mathbf{0} \ . \tag{2.4}$$

Then $\boldsymbol{\beta}$ is identified if $\boldsymbol{\Pi}$ has rank $G - 1$.

We consider the null hypothesis $H_0 \ : \boldsymbol{\beta} = \boldsymbol{\beta}_0$, where $\boldsymbol{\beta}_0$ is a specified vector. This null hypothesis is equivalent to the hypothesis that $\boldsymbol{\Pi}\boldsymbol{\beta}_0 = \mathbf{0}$.

The unrestricted regression coefficient $\boldsymbol{\Pi}$ can be estimated by the sample regression

$$\mathbf{P} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{Y} \ . \tag{2.5}$$

When $\boldsymbol{\Omega}$ is known and $\mathbf{V}$ is normal, the matrix $\mathbf{P}$ constitutes a sufficient set of statistics. The density of $\mathbf{P}$ for $\boldsymbol{\Omega} = \mathbf{I}_G$ is

$$\mathrm{L}(\boldsymbol{\Pi}) = (2\pi)^{-\frac{1}{2}TG} \exp\{-\frac{1}{2}\mathrm{tr}(\mathbf{P} - \boldsymbol{\Pi})'\mathbf{A}(\mathbf{P} - \boldsymbol{\Pi})\} \ , \tag{2.6}$$

where

$$\mathbf{A} = \mathbf{Z}'\mathbf{Z} \ . \tag{2.7}$$

The likelihood maximized with respect to $\boldsymbol{\Pi}$ is

$$\mathrm{L}_{H_2} = (2\pi)^{-\frac{1}{2}TG} \ , \tag{2.8}$$

where H_2 denotes the model with no restriction on $\boldsymbol{\Pi}$.

The model and the inference problem have some invariance properties. If $\mathbf{Z}$ is replaced by $\tilde{\mathbf{Z}} = \mathbf{ZC}$ and $\mathbf{\Pi}$ is replaced by $\tilde{\mathbf{\Pi}} = \mathbf{C}^{-1}\mathbf{\Pi}$, then the model (2.1) is unchanged since $\mathbf{\Pi}\boldsymbol{\beta} = \mathbf{0}$ is transformed to $\tilde{\mathbf{\Pi}}\boldsymbol{\beta} = \mathbf{0}$ The null hypothesis is unaffected. Note that $\mathbf{C}$ is an arbitrary nonsingular matrix.

If $\mathbf{Y}$ is replaced by $\mathbf{Y\Phi} = \mathbf{Y}^*$, $\mathbf{V}$ is replaced by $\mathbf{V\Phi} = \mathbf{V}^*$ ($\mathbf{v}_t^* = \mathbf{\Phi}'\mathbf{v}_t$), $\mathbf{\Pi}$ is replaced by $\mathbf{\Pi\Phi} = \mathbf{\Pi}^*$, $\boldsymbol{\beta}$ is repalced by $\mathbf{\Phi}^{-1}\boldsymbol{\beta} = \boldsymbol{\beta}^*$, and $\boldsymbol{\beta}_0$ is replaced by $\mathbf{\Phi}^{-1}\boldsymbol{\beta}_0 = \boldsymbol{\beta}_0^*$, the model is unchanged with $\mathbf{\Omega}$ replaced by $\mathbf{\Phi}'\mathbf{\Omega\Phi} = \mathbf{\Omega}^*$. When $\mathbf{\Omega} = \mathbf{\Omega}^* = \mathbf{I}_G$ is required, $\mathbf{\Phi}'\mathbf{\Phi} = \mathbf{I}_G$, that is, $\mathbf{\Phi}$ is an orthogonal matrix.

3. Likelihood Ratio Tests

We denote the model with $\mathbf{\Pi}$ having rank $G-1$ as H_1. We shall derive the likelihood ratio criterion to test H_0 when H_1 is assumed, namely $\mathrm{L}_{H_0}/\mathrm{L}_{H_1}$, as the ratio of the criterion to test H_0 vs H_2 and the criterion to test H_1 vs H_2.

$$\frac{\mathrm{L}_{H_0}}{\mathrm{L}_{H_1}} = \frac{\mathrm{L}_{H_0}/\mathrm{L}_{H_2}}{\mathrm{L}_{H_1}/\mathrm{L}_{H_2}} \ . \tag{3.1}$$

The hypothesis that $\mathbf{\Pi}\boldsymbol{\beta}_0 = \mathbf{0}$ can be considered as the hypothesis that $\mathbf{\Pi}$ has rank $G-1$ and $\mathbf{\Pi}\boldsymbol{\beta}_0 = \mathbf{0}$. When $G=2$, $\mathbf{\Pi}$ has rank 1, that is, can be written as $\boldsymbol{\mu}\boldsymbol{\gamma}'$ for the K-vector $\boldsymbol{\mu}$ and the 2 vector $\boldsymbol{\gamma}$, and that $\boldsymbol{\gamma}'\boldsymbol{\beta}_0 = 0$.

Lemma : If $G=2$ and $\mathbf{\Pi} = \boldsymbol{\mu}\boldsymbol{\gamma}'$, then the likelihood function of $\mathbf{P}$ maximized with respect to $\boldsymbol{\mu}$ is

$$\max_{\boldsymbol{\mu}} \mathrm{L}(\boldsymbol{\mu}\boldsymbol{\gamma}') = (2\pi)^{-T}\exp\{-\frac{1}{2}\mathrm{tr}\mathbf{P}'\mathbf{AP} + \frac{1}{2}\boldsymbol{\gamma}'\mathbf{P}'\mathbf{AP}\boldsymbol{\gamma}\} \ . \tag{3.2}$$

Proof :

$$\begin{aligned}
\mathrm{L}(\boldsymbol{\mu}\boldsymbol{\gamma}') &= (2\pi)^{-T}\exp\{-\frac{1}{2}\mathrm{tr}(\mathbf{P}-\boldsymbol{\mu}\boldsymbol{\gamma}')'\mathbf{A}(\mathbf{P}-\boldsymbol{\mu}\boldsymbol{\gamma}')\} \\
&= (2\pi)^{-T}\exp\{-\frac{1}{2}\mathrm{tr}\left[\mathbf{PAP}' - \boldsymbol{\gamma}\boldsymbol{\mu}'\mathbf{AP} - \mathbf{P}'\mathbf{A}\boldsymbol{\mu}\boldsymbol{\gamma}' + \boldsymbol{\gamma}\boldsymbol{\mu}'\mathbf{A}\boldsymbol{\mu}\boldsymbol{\gamma}'\right]\}
\end{aligned}$$

is maximized at $\boldsymbol{\mu} = \boldsymbol{P\gamma}/\boldsymbol{\gamma}'\boldsymbol{\gamma}$. **Q.E.D.**

To eliminate the indeterminacy of a scale factor in $\boldsymbol{\beta}$ we require

$$\boldsymbol{\beta}'\boldsymbol{\beta} = 1 \ . \tag{3.3}$$

An alternative normalization is

$$\boldsymbol{\beta}'\boldsymbol{\Psi}\boldsymbol{\beta} = 1 \ , \tag{3.4}$$

where $\boldsymbol{\Psi}$ is a known positive semi-definite matrix. To specify $\boldsymbol{\beta}' = (1, \boldsymbol{\beta}_2)$, for example, we take

$$\boldsymbol{\Psi} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \ . \tag{3.5}$$

When $\boldsymbol{\Pi}$ is $K \times 2$, then $\boldsymbol{\Pi\beta} = \mathbf{0}$ implies

$$\boldsymbol{\beta}'\boldsymbol{\gamma} = 0 \ \ . \tag{3.6}$$

Let $\boldsymbol{\gamma}_0$ be a vector such that $\boldsymbol{\beta}_0'\boldsymbol{\gamma}_0 = 0$. Then

$$\begin{aligned} \mathrm{L}_{H_0} &= (2\pi)^{-T} \exp\{-\frac{1}{2}\mathrm{tr}\mathbf{PAP}' + \frac{1}{2}\boldsymbol{\gamma}_0'\mathbf{P}'\mathbf{AP}\boldsymbol{\gamma}_0\} \\ &= (2\pi)^{-T} \exp\{-\frac{1}{2}\boldsymbol{\beta}_0'\mathbf{P}'\mathbf{AP}\boldsymbol{\beta}_0\} \end{aligned} \tag{3.7}$$

since

$$\mathrm{tr}\mathbf{PAP}' = \boldsymbol{\beta}'\mathbf{P}'\mathbf{AP}\boldsymbol{\beta} + \boldsymbol{\gamma}'\mathbf{P}'\mathbf{AP}\boldsymbol{\gamma} \ \ . \tag{3.8}$$

The likelihood ratio criterion for testing H_0 against H_2 is

$$\frac{\mathrm{L}_{H_0}}{\mathrm{L}_{H_2}} = e^{-\frac{1}{2}\boldsymbol{\beta}_0'\mathbf{P}'\mathbf{AP}\boldsymbol{\beta}_0} \ . \tag{3.9}$$

When the null hypothesis H_0 is true,

$$-2\log\frac{\mathrm{L}_{H_0}}{\mathrm{L}_{H_2}} = \boldsymbol{\beta}_0'\mathbf{P}'\mathbf{AP}\boldsymbol{\beta}_0 \tag{3.10}$$

has a χ^2-distribution with K degrees of freedom since $\mathbf{P}\boldsymbol{\beta}_0$ has the distribution $N(\mathbf{0}, \sigma^2\mathbf{A}^{-1})$.

Now consider

$$H_1 \ : \ \text{rank}\boldsymbol{\Pi} = G-1 \ , \tag{3.11}$$

that is, $\boldsymbol{\Pi}$ is written as $\boldsymbol{\Pi} = \boldsymbol{\mu}\boldsymbol{\gamma}^{'}$. The likelihood of $\mathbf{P}$ maximized subject to $\boldsymbol{\Pi} = \boldsymbol{\mu}\boldsymbol{\gamma}^{'}$ is

$$\begin{aligned} \mathrm{L}_{H_1} &= \max_{\boldsymbol{\gamma}} \max_{\boldsymbol{\mu}} \mathrm{L}(\boldsymbol{\mu}\boldsymbol{\gamma}^{'}) \\ &= (2\pi)^{-T} \exp\{-\frac{1}{2}\text{tr}\mathbf{P}\mathbf{A}\mathbf{P}^{'} + \frac{1}{2}\max_{\boldsymbol{\gamma}} \frac{\boldsymbol{\gamma}^{'}\mathbf{P}^{'}\mathbf{A}\mathbf{P}\boldsymbol{\gamma}}{\boldsymbol{\gamma}^{'}\boldsymbol{\gamma}}\} \\ &= (2\pi)^{-T} \exp\{-\frac{1}{2}\min_{\boldsymbol{\beta}} \frac{\boldsymbol{\beta}^{'}\mathbf{P}^{'}\mathbf{A}\mathbf{P}\boldsymbol{\beta}}{\boldsymbol{\beta}^{'}\boldsymbol{\beta}}\} \ . \end{aligned}$$

The likelihood ratio criterion for testing H_1 against H_2 is

$$\begin{aligned} \frac{\mathrm{L}_{H_1}}{\mathrm{L}_{H_2}} &= e^{-\frac{1}{2}\min_{\boldsymbol{\beta}} \frac{\boldsymbol{\beta}^{'}\mathbf{P}^{'}\mathbf{A}\mathbf{P}\boldsymbol{\beta}}{\boldsymbol{\beta}^{'}\boldsymbol{\beta}}} \\ &= e^{-\frac{1}{2}\frac{\hat{\boldsymbol{\beta}}^{'}\mathbf{P}^{'}\mathbf{A}\mathbf{P}\hat{\boldsymbol{\beta}}}{\hat{\boldsymbol{\beta}}^{'}\hat{\boldsymbol{\beta}}}} \ , \end{aligned} \tag{3.12}$$

where $\hat{\boldsymbol{\beta}}$ is the maximum likelihood estimator of $\boldsymbol{\beta}$ for H_1.

When the null hypothesis H_1 is true,

$$\begin{aligned} -2\log\frac{\mathrm{L}_{H_1}}{\mathrm{L}_{H_2}} &= \min_{\boldsymbol{\beta}} \frac{\boldsymbol{\beta}^{'}\mathbf{P}^{'}\mathbf{A}\mathbf{P}\boldsymbol{\beta}}{\boldsymbol{\beta}^{'}\boldsymbol{\beta}} \\ &= \frac{\hat{\boldsymbol{\beta}}^{'}\mathbf{P}^{'}\mathbf{A}\mathbf{P}\hat{\boldsymbol{\beta}}}{\hat{\boldsymbol{\beta}}^{'}\hat{\boldsymbol{\beta}}} \end{aligned} \tag{3.13}$$

has a limiting χ^2-distribution with 1 degree of freedom (Anderson and Rubin (1950)). Note that $\min_{\mathbf{b}} \frac{\mathbf{b}^{'}\mathbf{P}^{'}\mathbf{A}\mathbf{P}\mathbf{b}}{\mathbf{b}^{'}\mathbf{b}}$ is the smaller root of

$$\left|\mathbf{P}^{'}\mathbf{A}\mathbf{P} - d\mathbf{I}_2\right| = 0 \ . \tag{3.14}$$

The likelihood ratio criterion for testing H_0 against H_1 is

$$(3.15)\qquad \begin{aligned} \frac{\mathrm{L}_{H_0}}{\mathrm{L}_{H_1}} &= \frac{\mathrm{L}_{H_0}/\mathrm{L}_{H_2}}{\mathrm{L}_{H_1}/\mathrm{L}_{H_2}} \\ &= e^{-\frac{1}{2}\boldsymbol{\beta}_0^{'}\mathbf{P}^{'}\mathbf{A}\mathbf{P}\boldsymbol{\beta}_0 + \frac{1}{2}\min_{\boldsymbol{\beta}} \frac{\boldsymbol{\beta}^{'}\mathbf{P}^{'}\mathbf{A}\mathbf{P}\boldsymbol{\beta}}{\boldsymbol{\beta}^{'}\boldsymbol{\beta}}} \\ &= e^{\frac{1}{2}\left[\frac{\hat{\boldsymbol{\beta}}^{'}\mathbf{P}^{'}\mathbf{A}\mathbf{P}\hat{\boldsymbol{\beta}}}{\hat{\boldsymbol{\beta}}^{'}\hat{\boldsymbol{\beta}}} - \frac{\boldsymbol{\beta}_0^{'}\mathbf{P}^{'}\mathbf{A}\mathbf{P}\boldsymbol{\beta}_0}{\boldsymbol{\beta}_0^{'}\boldsymbol{\beta}_0}\right]} . \end{aligned}$$

Equivalently, the likelihood ratio criterion for testing H_0 against H_1 is

$$(3.16)\qquad -2\log\frac{\mathrm{L}_{H_0}}{\mathrm{L}_{H_1}} = \frac{\boldsymbol{\beta}_0^{'}\mathbf{P}^{'}\mathbf{A}\mathbf{P}\boldsymbol{\beta}_0}{\boldsymbol{\beta}_0^{'}\boldsymbol{\beta}_0} - \frac{\hat{\boldsymbol{\beta}}^{'}\mathbf{P}^{'}\mathbf{A}\mathbf{P}\hat{\boldsymbol{\beta}}}{\hat{\boldsymbol{\beta}}^{'}\hat{\boldsymbol{\beta}}} .$$

Note that the criterion does not depend on the normalization of $\boldsymbol{\beta}_0$ and of $\hat{\boldsymbol{\beta}}$. The criterion (3.18) has been proposed by Moreira (2003). He called (3.18) the conditional likelihood ratio.

Corollary : The likelihood ratio criterion (3.18) is invariant with respect to the groups of transformations

$$(3.17)\qquad \tilde{\mathbf{Z}} = \mathbf{Z}\mathbf{C},\ \ \tilde{\mathbf{\Pi}} = \mathbf{C}^{-1}\mathbf{\Pi}$$

and

$$(3.18)\qquad \mathbf{Y}^{*} = \mathbf{Y}\mathbf{\Phi},\ \ \mathbf{P}^{*} = \mathbf{P}\mathbf{\Phi},\ \ \boldsymbol{\beta}^{*} = \mathbf{\Phi}^{'}\boldsymbol{\beta},\ \ \boldsymbol{\beta}_0^{*} = \mathbf{\Phi}^{'}\boldsymbol{\beta}_0,$$

where $\mathbf{\Phi}\mathbf{\Phi}^{'} = \mathbf{I}$.

When $\mathbf{\Omega} = \sigma^2\mathbf{I}$, the transformation (3.19) is appropriate. A normalization of $\boldsymbol{\beta}^{'}\mathbf{\Psi}\boldsymbol{\beta} = 1$ is transformed to

$$(3.19)\qquad \boldsymbol{\beta}^{*'}\mathbf{\Phi}^{'}\mathbf{\Psi}\mathbf{\Phi}\boldsymbol{\beta}^{*} = 1 .$$

When $G = 2$ and $\mathbf{\Omega} = \mathbf{I}$, the condition $\boldsymbol{\beta}^{'}\boldsymbol{\beta} = 1$ implies that $\boldsymbol{\beta}$ can be written as

$$(3.20)\qquad \boldsymbol{\beta} = \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix}$$

and $\boldsymbol{\gamma}$ can be written as

$$\boldsymbol{\gamma} = \text{constant} \times \begin{bmatrix} -\sin\theta \\ \cos\theta \end{bmatrix} . \tag{3.21}$$

The term $\max_{\boldsymbol{\mu}} \mathrm{L}(\boldsymbol{\mu}\boldsymbol{\gamma}')$ is independent of the constant which can be taken as 1. The null hypothesis $\boldsymbol{\beta} = \boldsymbol{\beta}_0$ where

$$\boldsymbol{\beta}_0 = \begin{bmatrix} \cos\theta_0 \\ \sin\theta_0 \end{bmatrix} \tag{3.22}$$

implies that the null hypothesis $\boldsymbol{\beta} = \boldsymbol{\beta}_0$ is equivalent to $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ and

$$\boldsymbol{\gamma} = \boldsymbol{\gamma}_0 = \begin{bmatrix} -\sin\theta_0 \\ \cos\theta_0 \end{bmatrix} . \tag{3.23}$$

Anderson and Kunitomo (2007) derived the likelihood ratio criterion for testing H_0 against H_1 when $\boldsymbol{\Omega}$ is unknown and must be estimated. In that case let

$$\mathbf{H} = \mathbf{Y}'\mathbf{Y} - \mathbf{P}'\mathbf{A}\mathbf{P} . \tag{3.24}$$

The maximum likelihood estimators of $\boldsymbol{\Pi}$ and $\boldsymbol{\Omega}$ are $\mathbf{P}$ and $(1/T)\mathbf{H}$ (under H_2). Then

$$-2\log\frac{\mathrm{L}^*_{H_0}}{\mathrm{L}^*_{H_1}} = \log\left[1 + \frac{\boldsymbol{\beta}_0'\mathbf{P}'\mathbf{A}\mathbf{P}\boldsymbol{\beta}_0}{\boldsymbol{\beta}_0'\mathbf{H}\boldsymbol{\beta}_0}\right] - \log\left[1 + \min_{\mathbf{b}} \frac{\mathbf{b}'\mathbf{P}'\mathbf{A}\mathbf{P}\mathbf{b}}{\mathbf{b}'\mathbf{H}\mathbf{b}}\right] \tag{3.25}$$

is the likelihood ratio criterion for testing H_0 against H_1. It might be observed that Anderson and Rubin (1949) derived the likelihood ratio criterion for testing H_1 against H_2 explicitly and the LR criterion for testing H_0 against H_2 implicitly (the converse of a confidence region), but did not derive the LR criterion for testing H_0 against H_1.

When the null hypothesis is true, $\mathbf{P}\boldsymbol{\beta}_0 \xrightarrow{p} 0$, $(1/T)\mathbf{H} \xrightarrow{p} \boldsymbol{\Omega}$ as $T \to \infty$. Assume $(1/T)\mathbf{A} \to \mathbf{A}_0$. Then expansion of $\log(1+x) = x + o_p(1)$ for $x \to 0$ yields

$$-2\log\frac{\mathrm{L}^*_{H_0}}{\mathrm{L}^*_{H_1}} = \frac{\boldsymbol{\beta}_0'\mathbf{P}'\mathbf{A}\mathbf{P}\boldsymbol{\beta}_0}{\boldsymbol{\beta}_0'\mathbf{H}\boldsymbol{\beta}_0} - \min_{\mathbf{b}} \frac{\mathbf{b}'\mathbf{P}'\mathbf{A}\mathbf{P}\mathbf{b}}{\mathbf{b}'\mathbf{H}\mathbf{b}} + o_p(1) . \tag{3.26}$$

The limiting distribution of $-2\log[\mathrm{L}_{H_0}/\mathrm{L}_{H_1}]$ is the same as the limiting distribution of $-2\log[\mathrm{L}^*_{H_0}/\mathrm{L}^*_{H_1}]$. Under the null hypothesis the limiting distribution is χ^2 with 1 degree of freedom.

4. More General Models

Suppose that $\mathbf{Y}, \mathbf{Z}, \boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ in (2.2) are

$$\mathbf{Y} = (\mathbf{Y}_1, \mathbf{Y}_2),\ \mathbf{Z} = (\mathbf{Z}_1, \mathbf{Z}_2),\ \boldsymbol{\beta} = \begin{bmatrix} \boldsymbol{\beta}_1 \\ \mathbf{0} \end{bmatrix}\ \gamma = \begin{bmatrix} \gamma_1 \\ \mathbf{0} \end{bmatrix}, \tag{4.1}$$

where $\mathbf{Y}_1$ and $\mathbf{Y}_2$ have G_1 and G_2 columns, respectively, and $\mathbf{Z}_1$ and $\mathbf{Z}_2$ have K_1 and K_2 columns, respectively. Then (2.2) is

$$\mathbf{Y}_1\boldsymbol{\beta}_1 = \mathbf{Z}_1\gamma_1 + \mathbf{u}\ . \tag{4.2}$$

Let

$$\boldsymbol{\Pi} = \begin{bmatrix} \boldsymbol{\Pi}_{11} & \boldsymbol{\Pi}_{12} \\ \boldsymbol{\Pi}_{21} & \boldsymbol{\Pi}_{22} \end{bmatrix},\ \boldsymbol{\Omega} = \begin{bmatrix} \boldsymbol{\Omega}_{11} & \boldsymbol{\Omega}_{12} \\ \boldsymbol{\Omega}_{21} & \boldsymbol{\Omega}_{22} \end{bmatrix}. \tag{4.3}$$

Then (2.2) and (4.2) imply

$$\begin{bmatrix} \boldsymbol{\Pi}_{11}\boldsymbol{\beta}_1 \\ \boldsymbol{\Pi}_{21}\boldsymbol{\beta}_1 \end{bmatrix} = \begin{bmatrix} \gamma_1 \\ \mathbf{0} \end{bmatrix} \tag{4.4}$$

and $\mathbf{u} = \mathbf{V}_1\boldsymbol{\beta}_1$, where $\mathbf{V} = (\mathbf{V}_1, \mathbf{V}_2)$. The second subvector of (4.4) corresponds to (2.4).

It will be convenient to transform the model so that the two sets of exogenous variables are orthogonal. Let

$$\mathbf{A} = \mathbf{Z}'\mathbf{Z} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix} = \begin{bmatrix} \mathbf{Z}_1'\mathbf{Z}_1 & \mathbf{Z}_1'\mathbf{Z}_2 \\ \mathbf{Z}_2'\mathbf{Z}_1 & \mathbf{Z}_2'\mathbf{Z}_2 \end{bmatrix} \tag{4.5}$$

and

$$\mathbf{Z}_{2.1} = \mathbf{Z}_2 - \mathbf{Z}_1\mathbf{A}_{11}^{-1}\mathbf{A}_{12},\ \mathbf{A}_{22.1} = \mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12}\ . \tag{4.6}$$

Define $(\mathbf{\Pi}_{11}^*, \mathbf{\Pi}_{12}^*) = (\mathbf{I}_{K_1}, \mathbf{A}_{11}^{-1}\mathbf{A}_{12})\mathbf{\Pi}$. Then

$$\mathbf{Z}\mathbf{\Pi} = (\mathbf{Z}_1, \mathbf{Z}_{2.1}) \begin{bmatrix} \mathbf{\Pi}_{11}\boldsymbol{\beta}_1 \\ \mathbf{\Pi}_{21}\boldsymbol{\beta}_1 \end{bmatrix} = (\mathbf{Z}_1, \mathbf{Z}_{2.1})\mathbf{\Pi}^* \ , \tag{4.7}$$

by defining $\mathbf{\Pi}^*$. Define also

$$\mathbf{A}^* = \begin{bmatrix} \mathbf{Z}_1' \\ \mathbf{Z}_{2.1} \end{bmatrix} [\mathbf{Z}_1, \mathbf{Z}_{2.1}] = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{O} \\ \mathbf{O} & \mathbf{A}_{22.1} \end{bmatrix} , \tag{4.8}$$

$$\mathbf{P}^* = (\mathbf{A}^*)^{-1} \begin{bmatrix} \mathbf{Z}_1' \\ \mathbf{Z}_{2.1} \end{bmatrix} \mathbf{Y} = \begin{bmatrix} \mathbf{A}_{11}^{-1}\mathbf{Z}_1'\mathbf{Y} \\ \mathbf{A}_{22.1}^{-1}\mathbf{Z}_{2.1}\mathbf{Y} \end{bmatrix} = \begin{bmatrix} \mathbf{P}_1^* \\ \mathbf{P}_2^* \end{bmatrix} . \tag{4.9}$$

The density of $\mathbf{P}^*$ is defined from the density of $\mathbf{P}$ given (2.5) as

$$\begin{aligned} \mathrm{L}(\mathbf{\Pi}^*) &= (2\pi)^{-\frac{1}{2}TG_1} \exp\{-\frac{1}{2}\mathrm{tr}(\mathbf{P}_1^* - \mathbf{\Pi}_1^*)'\mathbf{A}_{11}(\mathbf{P}_1^* - \mathbf{\Pi}_1^*)\} \\ &\quad \times (2\pi)^{-\frac{1}{2}TG_2} \exp\{-\frac{1}{2}\mathrm{tr}(\mathbf{P}_2^* - \mathbf{\Pi}_2^*)'\mathbf{A}_{22.1}(\mathbf{P}_2^* - \mathbf{\Pi}_2^*)\} \ , \end{aligned} \tag{4.10}$$

where $\mathbf{\Pi}_1^* = (\mathbf{\Pi}_{11}^*, \mathbf{\Pi}_{12}^*)$ and $\mathbf{\Pi}_2^* = (\mathbf{\Pi}_{21}, \mathbf{\Pi}_{22})$. The maximum of the first part of $\mathrm{L}(\mathbf{\Pi}^*)$ occurs at $\mathbf{\Pi}_1^* = \mathbf{P}_1^*$ and is $(2\pi)^{-\frac{1}{2}TG_1}$. The maximum of the second part with respect to $\mathbf{\Pi}_{22}$ is

$$\mathrm{L}(\mathbf{\Pi}_{21}) = (2\pi)^{-\frac{1}{2}TG_2} \exp\{-\frac{1}{2}\mathrm{tr}(\mathbf{P}_{21} - \mathbf{\Pi}_{21})'\mathbf{A}_{22.1}(\mathbf{P}_{21} - \mathbf{\Pi}_{21})\} \ . \tag{4.11}$$

This corresponds to (2.5).

5. Distribution of Test Criterion

The matrix

$$\mathbf{G} = \mathbf{P}'\mathbf{A}\mathbf{P} \tag{5.1}$$

has the noncentral Wishart distribution with $T-K$ degrees of freedom, covariance matrix $\mathbf{\Omega} = \mathbf{I}_G$, and the noncentrality parameter

$$\mathbf{\Gamma} = \mathbf{\Pi}'\mathbf{A}\mathbf{\Pi} \tag{5.2}$$

when $\mathbf{P}$ has the density (2.6). The noncentral Wishart distribution was derived by Anderson and Girshick (1944) with correction (1964) for $\mathbf{\Pi}$ of rank 1. The criterion (3.18) can be written

$$-2\log\frac{\mathrm{L}_{H_0}}{\mathrm{L}_{H_1}} = \frac{\boldsymbol{\beta}_0'\mathbf{G}\boldsymbol{\beta}_0}{\boldsymbol{\beta}_0'\boldsymbol{\beta}_0} - \min_{\mathbf{b}} \frac{\mathbf{b}'\mathbf{G}\mathbf{b}}{\mathbf{b}'\mathbf{b}} \ . \tag{5.3}$$

Since the criterion is invariant with respect to orthogonal transformations (3.20), we can write (5.3) for $\boldsymbol{\beta}_0 = (1,0)'$ and

$$\mathbf{\Pi} = \boldsymbol{\mu}\boldsymbol{\gamma}' = \boldsymbol{\mu}(0,1) = (0,\boldsymbol{\mu}) \tag{5.4}$$

and

$$\min_{\mathbf{b}} \frac{\mathbf{b}'\mathbf{G}\mathbf{b}}{\mathbf{b}'\mathbf{b}} = d_1 \ , \tag{5.5}$$

the smaller root of

$$0 = |\mathbf{G} - d\mathbf{I}_2| = d^2 - (g_{11}+g_{22})d + g_{11}g_{22} - g_{12}^2 \ . \tag{5.6}$$

Thus (5.3) is

$$g_{11} - d_1 = \frac{1}{2}\left[g_{11} - g_{22} + \sqrt{(g_{11}-g_{22})^2 + 4g_{12}^2}\right] \ . \tag{5.7}$$

The distribution of (5.7) can be found from the noncentral Wishart distribution. See Anderson, Kunitomo, and Matsushita (2009).

References

[1] Anderson, T.W. and N. Kunitomo (2007), "On Likelihood Ratio Tests of Structural Coefficients : Anderson-Rubin (1949) revisited," Discussion Paper CIRJE-F-499, Graduate School of Economics, University of Tokyo (http://www.e.u-tokyo.ac.jp/cirje/research/dp/2007).

[2] Anderson, T.W., N. Kunitomo, and Y. Matsushita (2009), “The Limited Information Maximum Likelihood Estimator as an Angle,” Discussion Paper CIRJE-F-619, Graduate School of Economics, University of Tokyo (http://www.e.u-tokyo.ac.jp/cirje/research/dp/2009).

[3] Anderson, T.W. and M.A. Girshick (1944), “Some extensions of the Wishart distribution,” *Annals of Mathematical Statistics*, Vol. 15, 345-357. (Corrections, Vol.35 (1964),923-925).

[4] 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*, Vol. 20, 46-63.

[5] Anderson, T.W. and H. Rubin (1950), “The Asymptotic Properties of Estimates of the Parameters of a Single Equation in a Complete System of Stochastic Equation,” *Annals of Mathematical Statistics*, Vol. 21, 570-582.

[6] Andrews, Donald W.H., Marcelo J. Moreira, and James H. Stock (2006), “Optimal Two-Sided Invariant Similar Tests for Instrumental Variables Regression,” *Econometrica*, Vol.74, 715-752.

[7] Moreira, M. (2003), “A Conditional Likelihood Ratio Test for Structural Models,” *Econometrica*, Vol. 71, 1027-1048.