Optimal Significance Tests in Simultaneous Equation Models*

T. W. Anderson†
Stanford University

April 27, 2011

Abstract

Consider testing the null hypothesis that a single structural equation has specified coefficients. The alternative hypothesis is that the relevant part of the reduced form matrix has proper rank, that is, that the equation is identified. The usual linear model with normal disturbances is invariant with respect to linear transformations of the endogenous and of the exogenous variables. When the disturbance covariance matrix is known, it can be set to the identity, and the invariance of the endogenous variables is with respect to orthogonal transformations. The likelihood ratio test is invariant with respect to these transformations and is the best invariant test. Furthermore it is admissible in the class of all tests. Any other test has lower power and/or higher significance level.

Keywords: Admissible invariant tests, likelihood ratio tests, Bayes tests

*The author thanks Naoto Kunitomo Graduate School of Economics, University of Tokyo) for his generous assistance. An early version of this paper was presented to the James Durbin Seminar October 29, 2009, sponsored by the London School of Economics and University College, London.

†Department of Statistics and Economics, Stanford University, Stanford, CA 94305-4065, USA; twa@stanford.edu

1 Introduction

There is a considerable literature on statistical inference concerning a single structural equation in a simultaneous equation model. A predominance of the literature concerns estimation of the coefficients of the single equation. Anderson and Rubin (1949) developed the Limited Information Maximum Likelihood (LIML) estimator on the basis of normality of the disturbances. When the disturbance covariance matrix was known, the corresponding estimator was known as LIMLK. They also suggested a test of the null hypothesis, say, H_0, the vector of coefficients of the endogenous variables, say, $\boldsymbol{\beta}$, is a specified vector, say, $\boldsymbol{\beta}_0$; the alternative hypothesis, say H_2, $\boldsymbol{\beta}$ was unrestricted. When the single equation was over-identified (a term defined later), the test was inefficient in the sense that the power against the alternative was not optimum. Moreira (2003) derived an alternative test which he called the *conditional likelihood ratio* test. Anderson and Kunitomo (2007) derived an equivalent test by testing H_0 against H_1 : the equation is identified. This likelihood ratio criterion is the ratio of the likelihood ratio criterion for testing H_0 vs H_2 to the likelihood ratio criterion for testing H_1 vs H_2. (These two likelihood ratio criteria were given in Anderson and Rubin, 1949.)

The current paper treats the testing problem when the disturbances matrix is known and is assumed to be proportional to $\mathbf{I}$. Further, the number of endogenous variables in the single equation is restricted to two. In this case it is convenient to use polar coordinates for the vector $\boldsymbol{\beta}$.

The likelihood ratio criterion for testing H_0 against H_1 is developed in polar coordinates. The criterion has an intuitively appealing interpretation and some invariance properties; that is, the criterion is invariant to rotations of the coordinate system.

We show that the likelihood ratio test is the best *invariant* test by showing that it is a Bayes solution. It follows that it is *admissible* among the class of all tests. This means that there is no test with better significance level and better power. (The precise definition of admissibility will be given later.) The result is one of few properties of tests in the field that is not approximate or asymptotic. Chamberlain (2007) has also considered these problems in polar coordinates.

Anderson (1976) pointed out that a structural equation in a simultaneous equation model is the same as a *linear functional relationship* in the statistical literature. Creasy (1956) derived the likelihood ratio test of the slope parameter in this model.

Anderson, Stein and Zaman (1985) showed that the LIMLK estimator is admissible for a loss function to be defined later. They first showed that the LIMLK estimator was the best invariant estimator and then deduced that it was admissible in the class of all estimators.

Inference for the LIML estimator and the likelihood ratio test when the disturbance covariance matrix is estimated will be treated in a subsequent paper.

2 A simultaneous equation model

The observed data consists of a $T \times G$ matrix of endogenous or dependent variables $\mathbf{Y}$ and a $T \times K$ matrix of exogenous or independent variables $\mathbf{Z}$ ($G < K$). A linear model (the reduced form) is

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

where $\boldsymbol{\Pi}$ is a $K \times G$ matrix of parameters and $\boldsymbol{V}$ is a $T \times G$ matrix of unobservable disturbances. The rows of $\boldsymbol{V}$ are assumed independent; each row has a normal distribution

$N(\mathbf{0}, \boldsymbol{\Omega})$.

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

$$\boldsymbol{P} = (\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{Y}. \tag{2.2}$$

The covariance matrix $\boldsymbol{\Omega}$ can be estimated by $(1/T)\boldsymbol{H}$, where

$$\boldsymbol{H} = (\boldsymbol{Y} - \boldsymbol{Z}\boldsymbol{P})'(\boldsymbol{Y} - \boldsymbol{Z}\boldsymbol{P}) = \boldsymbol{Y}'\boldsymbol{Y} - \boldsymbol{P}'\boldsymbol{A}\boldsymbol{P} \tag{2.3}$$

and $\boldsymbol{A} = \boldsymbol{Z}'\boldsymbol{Z}$. The matrices $\boldsymbol{P}$ and $\boldsymbol{H}$ consititute sufficient statistics for the model.

A structural or behavioral equation may involve a $T \times G_1$ subset of the endogenous variables $\boldsymbol{Y}_1$, a $T \times K_1$ subset of the exogenous variables $\boldsymbol{Z}_1$, and a $T \times G_1$ subset of disturbances $\boldsymbol{V}_1$. The structural equation of interest is

$$\boldsymbol{Y}_1\boldsymbol{\beta}_1 = \boldsymbol{Z}_1\gamma_1 + \boldsymbol{u}, \tag{2.4}$$

where $\boldsymbol{u} = \boldsymbol{V}_1\boldsymbol{\beta}_1$ and $\boldsymbol{V} = (\boldsymbol{V}_1, \boldsymbol{V}_2)$. A component of $\boldsymbol{u}$ has the normal distribution $N(0, \sigma^2)$, where $\sigma^2 = \boldsymbol{\beta}_1'\boldsymbol{\Omega}_{11}\boldsymbol{\beta}_1$ and $\boldsymbol{\Omega}_{11}$ is the $G_1 \times G_1$ upper-left submatrix of

$$\boldsymbol{\Omega} = \begin{bmatrix} \boldsymbol{\Omega}_{11} & \boldsymbol{\Omega}_{12} \\ \boldsymbol{\Omega}_{21} & \boldsymbol{\Omega}_{22} \end{bmatrix} \tag{2.5}$$

When $\boldsymbol{Y}$, $\boldsymbol{Z}$, $\boldsymbol{V}$, and $\boldsymbol{\Pi}$ are partitioned similarly, the reduced form (2.1) can be written

$$(\boldsymbol{Y}_1, \boldsymbol{Y}_2) = (\boldsymbol{Z}_1, \boldsymbol{Z}_2) \begin{bmatrix} \boldsymbol{\Pi}_{11} & \boldsymbol{\Pi}_{12} \\ \boldsymbol{\Pi}_{21} & \boldsymbol{\Pi}_{22} \end{bmatrix} + (\boldsymbol{V}_1, \boldsymbol{V}_2), \tag{2.6}$$

where $(\boldsymbol{Y}_1, \boldsymbol{Y}_2)$ is a $T \times (G_1 + G_2)$ matrix. The relation between the reduced form and the structural equation is

$$\begin{bmatrix} \gamma_1 \\ \mathbf{0} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Pi}_{11} & \boldsymbol{\Pi}_{12} \\ \boldsymbol{\Pi}_{21} & \boldsymbol{\Pi}_{22} \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}_1 \\ \mathbf{0} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Pi}_{11}\boldsymbol{\beta}_1 \\ \boldsymbol{\Pi}_{21}\boldsymbol{\beta}_1 \end{bmatrix}. \tag{2.7}$$

The second submatrix of (2.7),

$$\boldsymbol{\Pi}_{21}\boldsymbol{\beta}_1 = \mathbf{0}, \tag{2.8}$$

defines $\boldsymbol{\beta}_1$ except for a multiplicative constant if and only if the rank of $\boldsymbol{\Pi}_{21}$ is $G_1 - 1$ $(G_1 < K_1)$. In that case the structural equation is said to be *identified.*

In this paper we derive the likelihood ratio test of the null hypothesis

$$H_0 : \boldsymbol{\beta}_1 = \boldsymbol{\beta}_0$$

against the alternative

$$H_1 : \boldsymbol{\beta}_1 \text{ is identified.}$$

The goal of this paper is to show that this test is admissible. Roughly speaking, it means that there is no other test that can have better power everywhere. In developing this thesis it will be convenient to carry out the detail when $\boldsymbol{\gamma}_1$ is vacuous, that is $K_1 = 0$. Furthermore, we set $G_2 = 0$ so that $G = G_1$. Then the structural equation is

$$\boldsymbol{Y}\boldsymbol{\beta} = (\boldsymbol{Z}\boldsymbol{\Pi} + \boldsymbol{V})\boldsymbol{\beta} = \boldsymbol{u}. \tag{2.9}$$

3 Invariance and normalization

The model (2.1), $\boldsymbol{\Omega}$, (2.9), and $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0$ is invariant with respect to linear transformations of the exogenous variables

$$\boldsymbol{Z}^+ = \boldsymbol{Z}\boldsymbol{C}, \qquad \boldsymbol{\Pi}^+ = \boldsymbol{C}^{-1}\boldsymbol{\Pi} \tag{3.1}$$

for $\boldsymbol{C}$ being nonsingular. Then

$$\boldsymbol{Z}^+\boldsymbol{\Pi}^+ = \boldsymbol{Z}\boldsymbol{\Pi}, \quad \boldsymbol{A}^+ = \boldsymbol{C}'\boldsymbol{A}\boldsymbol{C}, \quad \boldsymbol{P}^+ = \boldsymbol{C}^{-1}\boldsymbol{P}, \tag{3.2}$$

and

$$\begin{aligned} \boldsymbol{G}^+ &= \boldsymbol{P}^{+\prime}\boldsymbol{A}^+\boldsymbol{P}^+ = \boldsymbol{P}'\boldsymbol{A}\boldsymbol{P} = \boldsymbol{G}, \\ \boldsymbol{H}^+ &= \boldsymbol{Y}'\boldsymbol{Y} - \boldsymbol{P}^{+\prime}\boldsymbol{A}^+\boldsymbol{P}^+ = \boldsymbol{H}. \end{aligned} \tag{3.3}$$

If the rank of $\boldsymbol{\Pi}$ is $G-1$ ($\leq K$), the equation $\boldsymbol{\Pi}\boldsymbol{\beta} = \boldsymbol{0}$ determines $\boldsymbol{\beta}$ except for a multiplicative constant. The "natural normalization" is

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

which determines the constant except for sign. The model (2.1), (2.9), and (3.4) is invariant with respect to transformations

$$\boldsymbol{Y}^* = \boldsymbol{Y}\boldsymbol{\Phi}, \quad \boldsymbol{\Pi}^* = \boldsymbol{\Pi}\boldsymbol{\Phi}, \quad \boldsymbol{\beta}^* = \boldsymbol{\Phi}^{-1}\boldsymbol{\beta}, \quad \boldsymbol{V}^* = \boldsymbol{V}\boldsymbol{\Phi}, \tag{3.5}$$

and

$$\boldsymbol{\Omega}^* = \boldsymbol{\Phi}'\boldsymbol{\Omega}\boldsymbol{\Phi}, \qquad \boldsymbol{\beta}_0^* = \boldsymbol{\Phi}^{-1}\boldsymbol{\beta}_0, \tag{3.6}$$

where $\boldsymbol{\Phi}$ is nonsingular. Then

$$\boldsymbol{P}^* = \boldsymbol{P}\boldsymbol{\Phi}, \quad \boldsymbol{G}^* = \boldsymbol{P}^{*\prime}\boldsymbol{A}\boldsymbol{P}^* = \boldsymbol{\Phi}'\boldsymbol{P}'\boldsymbol{A}\boldsymbol{P}\boldsymbol{\Phi} = \boldsymbol{\Phi}'\boldsymbol{G}\boldsymbol{\Phi}, \tag{3.7}$$

and

$$\boldsymbol{H}^* = \boldsymbol{\Phi}'\boldsymbol{H}\boldsymbol{\Phi}, \quad \boldsymbol{\Pi}^*\boldsymbol{\beta}^* = \boldsymbol{\Pi}\boldsymbol{\beta} = \boldsymbol{0}, \quad \boldsymbol{\beta}^{*\prime}\boldsymbol{\Omega}^*\boldsymbol{\beta}^* = 1. \tag{3.8}$$

We consider the model (2.1) and (2.9) when $\boldsymbol{\Omega}$ (the covariance matrix of a row of $\boldsymbol{V}$) is known. In this case we can make a transformation (3.5) and (3.6) so $\boldsymbol{\Omega} = \boldsymbol{I}$. Then the first equation in (3.6) is

$$\boldsymbol{I} = \boldsymbol{O}'\boldsymbol{O}, \tag{3.9}$$

that is, the invariance with respect to transformations is with respect to *orthogonal* transformations. We shall use $\boldsymbol{O}$ to indicate an orthogonal matrix. We can write (3.5) and (3.6) as

$$\begin{aligned} \boldsymbol{Y}^* = \boldsymbol{Y}\boldsymbol{O}, \quad \boldsymbol{\Pi}^* &= \boldsymbol{\Pi}\boldsymbol{O}, \quad \boldsymbol{\beta}^* = \boldsymbol{O}'\boldsymbol{\beta}, \\ \boldsymbol{\beta}_0^* = \boldsymbol{O}'\boldsymbol{\beta}_0, &\qquad \boldsymbol{\beta}^{*\prime}\boldsymbol{\beta}^* = \boldsymbol{\beta}'\boldsymbol{\beta} = 1. \end{aligned} \tag{3.10}$$

The null hypothesis is $\boldsymbol{\beta} = \boldsymbol{\beta}_0$.

4 A canonical form for $G = 2$ and polar coordinates

The main part of this paper concerns the model for $\boldsymbol{\Omega} = \boldsymbol{I}_2$ and

$$G_1 = G = 2, \quad G_2 = 0, \quad K_1 = 0, \quad K_2 = K \geq 2. \tag{4.1}$$

Then the vector $\boldsymbol{\beta}$ with natural parameterization satisfies

$$\boldsymbol{\Pi\beta} = \mathbf{0}, \qquad \boldsymbol{\beta}'\boldsymbol{\beta} = 1. \tag{4.2}$$

We can parameterize $\boldsymbol{\beta}$ as

$$\boldsymbol{\beta} = \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix} = \boldsymbol{\beta}_\theta, -\pi \leq \theta \leq \pi. \tag{4.3}$$

This is the *polar* or *angular* representation of the coefficient vector.

When the $K \times 2$ matrix $\boldsymbol{\Pi}$ has rank 1, it can be parameterized as

$$\boldsymbol{\Pi} = \boldsymbol{\gamma}\boldsymbol{\alpha}_\theta', \tag{4.4}$$

where $\boldsymbol{\gamma}$ is a $K \times 1$ vector and

$$\boldsymbol{\alpha}_\theta = \begin{bmatrix} -\sin\theta \\ \cos\theta \end{bmatrix}. \tag{4.5}$$

Note that

$$(\boldsymbol{\beta}_\theta, \boldsymbol{\alpha}_\theta) = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} = \boldsymbol{O}_\theta \tag{4.6}$$

is an orthogonal matrix. The model is identified.

Since $\boldsymbol{\Omega}$ is known, the sufficient statistic in the model is $\boldsymbol{P}$.

Now make a transformation (3.1) so $\boldsymbol{A}^+ = \boldsymbol{C}'\boldsymbol{AC} = \boldsymbol{I}_K$; define $\boldsymbol{Q} = \boldsymbol{P}^+ = \boldsymbol{C}^{-1}\boldsymbol{P}$ and $\boldsymbol{W} = \boldsymbol{A}^{-1}\boldsymbol{Z}'\boldsymbol{V}$,

$$\boldsymbol{\Pi}^+ = \boldsymbol{\nu}\boldsymbol{\alpha}_\theta', \quad \boldsymbol{P}'\boldsymbol{AP} = \boldsymbol{Q}'\boldsymbol{Q}, \quad \boldsymbol{\nu} = \boldsymbol{C}^{-1}\gamma. \tag{4.7}$$

The model is

$$\boldsymbol{Q} = \boldsymbol{\nu}\boldsymbol{\alpha}_\theta' + \boldsymbol{W}. \tag{4.8}$$

Here $\boldsymbol{W} = (\boldsymbol{w}_1, \boldsymbol{w}_2)$, $\mathcal{E}(\boldsymbol{W}) = \mathbf{0}$,

$$\mathcal{E}(\boldsymbol{w}_1\boldsymbol{w}_1') = \mathcal{E}(\boldsymbol{w}_2\boldsymbol{w}_2') = \boldsymbol{I}_K, \qquad \mathcal{E}(\boldsymbol{w}_1\boldsymbol{w}_2') = \mathbf{0}. \tag{4.9}$$

The hypothesis $\boldsymbol{\beta} = \boldsymbol{\beta}_0$ is equivalent to the hypothesis $\theta = \theta_0$ when $\boldsymbol{\beta} = (\cos\theta, \sin\theta)'$ and is equivalent to the hypothesis $\boldsymbol{\alpha} = \boldsymbol{\alpha}_0$ when $\boldsymbol{\alpha}' = (-\sin\theta, \cos\theta)$ and $\theta = \theta_0$.

5 The density of $\boldsymbol{Q}$

The density of $\boldsymbol{Q}$ is

$$(5.1)\qquad \frac{1}{(2\pi)^K} e^{-\frac{1}{2}\mathrm{tr}\boldsymbol{W}'\boldsymbol{W}} = \frac{1}{(2\pi)^K} e^{-\frac{1}{2}\mathrm{tr}(\boldsymbol{Q}'\boldsymbol{Q}+\boldsymbol{\nu}\boldsymbol{\nu}'-2\boldsymbol{\alpha}_\theta\boldsymbol{\nu}'\boldsymbol{Q})} = \frac{1}{(2\pi)^K} e^{-\frac{1}{2}\mathrm{tr}(\boldsymbol{Q}'\boldsymbol{Q})-\frac{1}{2}\boldsymbol{\nu}'\boldsymbol{\nu}+\boldsymbol{\nu}'\boldsymbol{Q}\boldsymbol{\alpha}_\theta}.$$

Let

$$(5.2)\qquad \boldsymbol{\nu}'\boldsymbol{\nu} = \lambda^2, \qquad \boldsymbol{\nu} = \lambda\boldsymbol{\eta},$$

where $\boldsymbol{\eta}'\boldsymbol{\eta} = 1$. Then the density of $\boldsymbol{Q}$ is

$$(5.3)\qquad \frac{1}{(2\pi)^K} e^{-\frac{1}{2}\mathrm{tr}(\boldsymbol{Q}'\boldsymbol{Q})-\frac{1}{2}\lambda^2+\lambda\boldsymbol{\eta}'\boldsymbol{Q}\boldsymbol{\alpha}_\theta}.$$

We shall find the best test of $\theta = \theta_0$ that is invariant with respect to the group of transformations

$$(5.4)\qquad \boldsymbol{\alpha}_\theta \to \boldsymbol{O}_a\boldsymbol{\alpha}_\theta, \quad \boldsymbol{\alpha}_{\theta_0} \to \boldsymbol{O}_a\boldsymbol{\alpha}_{\theta_0}, \quad \boldsymbol{\eta} \to \boldsymbol{O}_b\boldsymbol{\eta}.$$

An explicit expression for the polar coordinates in K dimensions is given in Problem 7.1 of Anderson (2003).

6 Reduction to $\boldsymbol{G}$

First we show that a function of $\boldsymbol{Q}$ that is invariant with respect to transformations (5.4) is a function of $\boldsymbol{Q}'\boldsymbol{Q} = \boldsymbol{G}$.

Lemma 6.1. *A function of $\boldsymbol{Q}$ that is invariant with respect to*

$$(6.1)\qquad \boldsymbol{Q} \to \boldsymbol{O}_a\boldsymbol{Q}, \qquad \boldsymbol{Q} \to \boldsymbol{Q}\boldsymbol{O}_b,$$

is a function of $\boldsymbol{G} = \boldsymbol{Q}'\boldsymbol{Q}$.

Proof. (i) $\boldsymbol{G}$ is a function of $\boldsymbol{Q}$ that is invariant. (ii) If there are $\boldsymbol{Q}_1$ and $\boldsymbol{Q}_2$ such that

$$(6.2)\qquad \boldsymbol{Q}_1'\boldsymbol{Q}_1 = \boldsymbol{Q}_2'\boldsymbol{Q}_2,$$

then there exists an orthogonal matrix $\boldsymbol{O}_c$ such that $\boldsymbol{Q}_1 = \boldsymbol{Q}_c\boldsymbol{Q}_2$. □

Invariant tests of $H_0 : \theta = \theta_0$ can be based on $\boldsymbol{G} = \boldsymbol{Q}'\boldsymbol{Q}$.

7 Density of $\boldsymbol{G}$

The matrix $\boldsymbol{G} = \boldsymbol{Q}'\boldsymbol{Q}$ has the noncentral Wishart distribution with K degrees of freedom, covariance $\boldsymbol{I}_2$, and noncentrality matrix

$$(7.1)\qquad (\lambda\boldsymbol{\eta}\boldsymbol{\alpha}_\theta')'(\lambda\boldsymbol{\eta}\boldsymbol{\alpha}_\theta') = \lambda^2\boldsymbol{\alpha}_\theta\boldsymbol{\alpha}_\theta'.$$

See Anderson and Girshick (1944): "Some extensions of the Wishart distribution," *Annals of Mathematical Statistics*. (Corrections, 1964). The density or likelihood of $\boldsymbol{G}$ is

$$\frac{e^{-\frac{1}{2}\lambda^2-\frac{1}{2}\mathrm{tr}\mathbf{G}}|\boldsymbol{G}|^{\frac{1}{2}(K-3)}}{2^{\frac{1}{2}K+1}\pi^{\frac{1}{2}}\Gamma\left[\frac{1}{2}(K-1)\right]}(\lambda^2\boldsymbol{\alpha}_\theta'\mathbf{G}\boldsymbol{\alpha}_\theta)^{-(K-2)/4}\mathrm{I}_{\frac{1}{2}(K-2)}\left(\lambda\sqrt{\boldsymbol{\alpha}_\theta'\mathbf{G}\boldsymbol{\alpha}_\theta}\right), \tag{7.2}$$

where

$$\mathrm{I}_{\frac{1}{2}(K-2)}(z)=\left(\frac{1}{2}z\right)^{\frac{1}{2}(K-2)}\sum_{j=0}^{\infty}\left(\frac{z^2}{4}\right)^j\frac{1}{j!\Gamma(\frac{1}{2}K+j)} \tag{7.3}$$

is the modified Bessel function of order $(K-2)/2$. (Abramowitz and Stegun, 1972, (9.6.10) on p. 375); see also Appendix B.

Let $\boldsymbol{G}=\boldsymbol{O}_t\boldsymbol{R}\boldsymbol{O}_t'$, where

$$\boldsymbol{R}=\begin{bmatrix} r_1 & 0\\ 0 & r_2\end{bmatrix}, \tag{7.4}$$

$$\boldsymbol{O}_t=\begin{bmatrix}\cos t & -\sin t\\ \sin t & \cos t\end{bmatrix}=(\boldsymbol{\beta}_t,\boldsymbol{\alpha}_t)\,. \tag{7.5}$$

The diagonal elements of $\boldsymbol{R}$ are the eigenvalues of $\boldsymbol{G}$ ($0\le r_1\le r_2<\infty$), and $\boldsymbol{\beta}_t$ and $\boldsymbol{\alpha}_t$ are the corresponding eigenvectors; that is,

$$\boldsymbol{G}(\boldsymbol{\beta}_t,\quad \boldsymbol{\alpha}_t)=(\boldsymbol{\beta}_t,\quad \boldsymbol{\alpha}_t)\,\boldsymbol{R}. \tag{7.6}$$

Transform $\boldsymbol{G}$ (2×2) to (r_1,r_2,t), The Jacobian of the transformation is r_2-r_1; see Appendix A.

The density of r_1,r_2 and t ($-\pi\le t\le\pi$) is

$$\frac{(r_2-r_1)e^{-\frac{1}{2}\lambda^2-\frac{1}{2}(r_1+r_2)}(r_1r_2)^{\frac{1}{2}(K-3)}}{2^{\frac{1}{2}K+1}\pi^{\frac{1}{2}}\Gamma\left[\frac{1}{2}(K-1)\right]}\mathrm{I}^*_{\frac{1}{2}(K-2)}(\lambda^2c^2), \tag{7.7}$$

where

$$\begin{aligned} c^2 &= \boldsymbol{\alpha}_\theta'\boldsymbol{O}_t\boldsymbol{R}\boldsymbol{O}_t'\boldsymbol{\alpha}_\theta\\ &= \boldsymbol{\alpha}_{\theta-t}'\boldsymbol{R}\boldsymbol{\alpha}_{\theta-t}\\ &= r_1\sin^2(t-\theta)+r_2\cos^2(t-\theta)\\ &= r_2-(r_2-r_1)\sin^2(t-\theta), \end{aligned} \tag{7.8}$$

$$\begin{aligned} \mathrm{I}^*_{\frac{1}{2}(K-2)}(\lambda^2c^2) &= \left(\frac{\lambda c}{2}\right)^{-\frac{1}{2}(K-2)}\mathrm{I}_{\frac{1}{2}(K-2)}(\lambda c)\\ &= \sum_{j=0}^{\infty}\left(\frac{\lambda^2c^2}{4}\right)^j\frac{1}{j!\Gamma(\frac{1}{2}K+j)}. \end{aligned} \tag{7.9}$$

Let

$$n(r_1,r_2)=\frac{(r_2-r_1)(r_1r_2)^{\frac{1}{2}(K-3)}e^{-(r_1+r_2)/2}}{2^{\frac{1}{2}K+1}\pi^{\frac{1}{2}}\Gamma\left[\frac{1}{2}(K-1)\right]}. \tag{7.10}$$

Then the density of r_1,r_2, and t is

$$h(r_1,r_2,t|\theta,\lambda)=n(r_1,r_2)e^{-\frac{1}{2}\lambda^2}\mathrm{I}^*_{\frac{1}{2}(K-2)}(\lambda^2c^2). \tag{7.11}$$

8 Likelihood ratio criterion

The density (i.e., likelihood) of $r_1,\ r_2$, and t given λ and

$$H_0 : \theta = \theta_0 \tag{8.1}$$

is

$$\max_{H_0} \mathrm{Lhd} = n(r_1, r_2) e^{-\lambda^2/2} \mathrm{I}^*_{\frac{1}{2}(K-2)}(\lambda^2 c_0^2), \tag{8.2}$$

where

$$c_0^2 = r_1 \sin^2(t - \theta_0) + r_2 \cos^2(t - \theta_0) = r_2 - (r_2 - r_1)\sin^2(t - \theta_0). \tag{8.3}$$

The likelihood is maximized with respect to θ (given λ) for

$$H_1 : -\pi \leq \theta \leq \pi \tag{8.4}$$

at $\hat{\theta} = t$. Then

$$\max_{H_1} \mathrm{Lhd} = n(r_1, r_2) e^{-\lambda^2/2} \mathrm{I}^*_{\frac{1}{2}(K-2)}(\lambda^2 r_2). \tag{8.5}$$

The likelihood ratio criterion for testing $H_0 : \theta = \theta_0$ against the alternative $H_1 : -\pi \leq \theta \leq \pi$ given λ is

$$\begin{aligned} \mathrm{LRC} = \frac{\max_{H_0} \mathrm{Lhd}}{\max_{H_1} \mathrm{Lhd}} &= \frac{\mathrm{I}^*_{\frac{1}{2}(K-2)}(\lambda^2 c_0^2)}{\mathrm{I}^*_{\frac{1}{2}(K-2)}(\lambda^2 r_2)} \\ &= \frac{\mathrm{I}^*_{\frac{1}{2}(K-2)}\left\{\lambda^2\left[r_2 - (r_2 - r_1)\sin^2(t - \theta_0)\right]\right\}}{\mathrm{I}^*_{\frac{1}{2}(K-2)}(\lambda^2 r_2)} \\ &= \frac{\mathrm{I}^*_{\frac{1}{2}(K-2)}\left[s_2 - (s_2 - s_1)\sin^2(t - \theta_0)\right]}{\mathrm{I}^*_{\frac{1}{2}(K-2)}[s_2]}, \end{aligned} \tag{8.6}$$

where $s_1 = \lambda^2 r_1$ and $s_2 = \lambda^2 r_2$.

The function $\mathrm{I}^*_{\frac{1}{2}(K-2)}(\lambda^2 c_0)$ is an increasing function of $\lambda^2 c_0$, and c_0^2 is an increasing function of $(r_2 - r_1)\sin^2(t - \theta_0)$, hence $\mathrm{I}^*_{\frac{1}{2}(K-2)}(\lambda^2 c_0)$ is a decreasing function of $(r_2 - r_1)\sin^2(t - \theta_0)$. The acceptance region of the likelihood ratio test of $H_0 : \theta = \theta_0$ given λ can be written

$$(r_2 - r_1)\sin^2(t - \theta_0) \leq \text{function of } r_1,\ r_2 \text{ and } \lambda. \tag{8.7}$$

Note that the likelihood ratio criterion does not depend on the parameter λ. However, the probability of acceptance does depend on λ. When the null hypothesis is true, the distribution of the LRC does not depend on θ_0; that is, the distribution is invariant with respect to transformation (3.10). The maximum likelihood estimator of θ is $\hat{\theta} = t$; the maximum likelihood estimator of $\boldsymbol{\beta}$ is $\hat{\boldsymbol{\beta}} = \boldsymbol{\beta}_{\hat{\theta}}$.

The likelihood ratio criterion when λ is considered as a parameter could be derived from the model (5.1); that is equivalent to the hypothesis $\boldsymbol{\beta} = \boldsymbol{\beta}_{\theta_0}$, when $\boldsymbol{\beta}'\boldsymbol{\beta} = 1$. The likelihood of (5.1) is maximized with respect to $\boldsymbol{\nu}$ for fixed θ at $\hat{\boldsymbol{\nu}} = \boldsymbol{Q}\boldsymbol{\alpha}_\theta$ yielding a maximized likelihood of

$$\frac{1}{(2\pi)^K} e^{-\frac{1}{2}\mathrm{tr}\boldsymbol{Q}'\boldsymbol{Q}+\frac{1}{2}\boldsymbol{\alpha}_\theta'\boldsymbol{Q}'\boldsymbol{Q}\boldsymbol{\alpha}_\theta} = \frac{1}{(2\pi)^K} e^{-\frac{1}{2}(\mathrm{tr}\boldsymbol{G}-\boldsymbol{\alpha}_\theta'\boldsymbol{G}\boldsymbol{\alpha}_\theta)} = \frac{1}{(2\pi)^K} e^{-\frac{1}{2}\mathrm{tr}\boldsymbol{R}+\frac{1}{2}c^2}. \tag{8.8}$$

Under the null hypothesis c^2 is

$$c_0^2 = r_2 - (r_2 - r_1)\sin^2(t - \theta_0). \tag{8.9}$$

Under the alternative H_1, the maximum of the likelihood (8.8) occurs at $\theta = 0$ and $c^2 = r_2$. Then the likelihood ratio criterion for testing H_0 vs. H_1 is

$$e^{-\frac{1}{2}(r_2-r_1)\sin^2(t-\theta_0)}. \tag{8.10}$$

However, to carry out the admissibility argument requires explicit treatment for each value of λ.

9 Bayes test

We now formulate the testing problem as a 2-decision problem : $\theta = \theta_0$ vs $\theta \neq \theta_0$ with the loss function $L(\theta, a)$, where the action a is *Accept* H_0 or *Reject* H_0.

$L(\theta, a)$	Action	
	Accept H_0	Reject H_0
Parameter $\theta = \theta_0$	0	1
$\theta \neq \theta_0$	1	0

A test (or decision rule) is a function $d(r_1, r_2, t)$ taking values $a =$ *accept* H_0 and $a =$ *reject* H_0. The *risk* of a test is the expected loss

$$R(\theta, \lambda, d) = \int_0^\infty \int_0^{r_2} \int_{-\pi}^{\pi} L[\theta, d(r_1, r_2, t)] h(r_1, r_2, t|\theta, \lambda)\, dt\, dr_1\, dr_2 \tag{9.1}$$

as a function of θ and λ. The *average risk* of a procedure with prior distribution $P(\theta)$ is

$$R^*[P(\cdot), \lambda, d] = \int_{-\pi}^{\pi} R(\theta, \lambda, d)\, dP(\theta). \tag{9.2}$$

We suppose the distribution $P(\theta)$ has a jump of $\Pr\{\theta = \theta_0\}$ at θ_0 and a density $[1 - \Pr\{\theta = \theta_0\}]p(\theta)$ for $\theta \neq \theta_0$. Then

$$\begin{aligned} R^*[P(\cdot), \lambda, d] &= \Pr\{\theta = \theta_0\} \Pr\{\text{reject } H_0|\theta_0, \lambda\} \\ &\quad + [1 - \Pr\{\theta = \theta_0\}] \int_{-\pi}^{\pi} \Pr\{\text{accept } H_0|\theta, \lambda\} p(\theta)\, d\theta \\ &= \Pr\{\theta = \theta_0\} \int_R h(r_1, r_2, t|\theta_0, \lambda)\, dr_1\, dr_2\, dt \\ &\quad + [1 - \Pr\{\theta = \theta_0\}] \int_A \bar{h}(r_1, r_2, t|\lambda)\, dr_1\, dr_2\, dt, \end{aligned} \tag{9.3}$$

where R is the rejection set of r_1, r_2, t and A is the acceptance set, and

$$\bar{h}(r_1, r_2, t|\lambda) = \int_{-\pi}^{\pi} h(r_1, r_2, t|\theta, \lambda) p(\theta)\, d\theta \tag{9.4}$$

is a density. The average risk can be written as

$$\begin{aligned} R^*[P(\cdot), \lambda, d] = \Pr\{\theta = \theta_0\} + \int_A \{[1 - \Pr\{\theta = \theta_0\}]\bar{h}(r_1, r_2, t|\lambda) \\ - \Pr\{\theta = \theta_0\} h(r_1, r_2, t|\theta_0, \lambda)\}\, dr_1\, dr_2\, dt. \end{aligned} \tag{9.5}$$

The average risk $R^*[P(\,\cdot\,), \lambda, d]$ is minimized by the set A for which

$$[1 - \Pr\{\theta = \theta_0\}]\bar{h}(r_1, r_2, t|\lambda) - \Pr\{\theta = \theta_0\} h(r_1, r_2, t|\theta_0, \lambda) \le 0, \tag{9.6}$$

that is

$$A: \ \frac{h(r_1, r_2, t|\theta_0, \lambda)}{\bar{h}(r_1, r_2, t|\lambda)} \ge \frac{1 - \Pr\{\theta = \theta_0\}}{\Pr\{\theta = \theta_0\}}. \tag{9.7}$$

Theorem 9.1. *For each λ the Bayes test of $H_0 : \theta = \theta_0$ vs $H_1 : \theta \ne \theta_0$ when H_0 has the prior probability $[1 - \Pr\{\theta = \theta_0\}]$ and H_1 has the prior probability $\Pr\{\theta = \theta_0\}$ with density $p(\theta)$ $(\theta \ne \theta_0)$, has the acceptance set (9.7).*

The Bayes test is essentially obtained by applying the Neyman–Pearson Fundamental Lemma to $h(r_1, r_2, t|\theta_0, \lambda)$ and $\bar{h}(r_1, r_2, t|\lambda)$.

When

$$p(\theta) = \frac{1}{2\pi}, \qquad -\pi \le \theta \le \pi, \tag{9.8}$$

the denominator of the left-hand side of (9.7) is

$$\begin{aligned} &\bar{h}(r_1, r_2, \lambda) \\ &= n(r_1, r_2) e^{-\frac{1}{2}\lambda^2} \frac{1}{2\pi} \int_{-\pi}^{\pi} \sum_{j=0}^{\infty} \left(\frac{\lambda^2}{4}\right)^j \frac{1}{j!\Gamma[\frac{1}{2}K + j]} \left[r_2 - (r_2 - r_1)\sin^2(t - \theta)\right]^j d\theta \\ &= n(r_1, r_2) e^{-\frac{1}{2}\lambda^2} \sum_{j=0}^{\infty} \left(\frac{\lambda^2}{4}\right)^j \frac{1}{j!\Gamma[\frac{1}{2}K + j]} \frac{1}{2\pi} \int_{t-\pi}^{t+\pi} \left[r_2 - (r_2 - r_1)\sin^2 x\right]^j dx \\ &= n(r_1, r_2) e^{-\frac{1}{2}\lambda^2} \sum_{j=0}^{\infty} \left(\frac{\lambda^2}{4}\right)^j \frac{1}{j!\Gamma[\frac{1}{2}K + j]} \frac{1}{2\pi} \int_{-\pi}^{\pi} \left[r_2 - (r_2 - r_1)\sin^2 x\right]^j dx \\ &= n(r_1, r_2) e^{-\frac{1}{2}\lambda^2} f_K(r_1, r_2, \lambda), \end{aligned} \tag{9.9}$$

say. The integrand in the fourth line of (9.9) is nonnegative and less than r_2^j; hence, the sum converges and $f_K(r_1, r_2, \lambda)$ is well-defined. Then the left-hand side of (9.7) is

$$\frac{h(r_1, r_2, t|\theta_0, \lambda)}{\bar{h}(r_1, r_2, t|\lambda)} = \frac{\mathrm{I}^*_{\frac{1}{2}(K-2)}(\lambda^2 c^2)}{f_K(r_1, r_2, \lambda)}. \tag{9.10}$$

Let $\lambda^2 r_1 = s_1$ and $\lambda^2 r_2 = s_2$. Then the LRC can be written as

$$\mathrm{LRC} = \frac{\mathrm{I}^*_{\frac{1}{2}(K-2)}\{s_2 - (s_2 - s_1)\sin^2(t - \theta_0)\}}{\mathrm{I}^*_{\frac{1}{2}(K-2)}\{s_2\}} \tag{9.11}$$

and the left-hand side of (9.7) is

$$\frac{\mathrm{I}^*_{\frac{1}{2}(K-2)}\{s_2 - (s_2 - s_1)\sin^2(t-\theta_0)\}}{\int_{-\pi}^{\pi} \mathrm{I}^*_{\frac{1}{2}(K-2)}\{s_2 - (s_2 - s_1)x\}dx}. \tag{9.12}$$

The numerator of the LRC and the left-hand side of (9.7) are the same; the denominator of the LRC and the left-hand side of (9.7) defining the Bayes test are functions of s_1 and s_2.

The conclusion is that a LR test can be expressed as a Bayes test for a prior of the uniform distribution for the parameter θ.

Theorem 9.2. *The likelihood ratio test for $H_0 : \theta = \theta_0$ vs $H_1 : \theta \neq \theta_0$ is a Bayes test for a prior density $1/(2\pi)$.*

10 Admissibility of invariant tests

Consider a family of densities $f(\boldsymbol{y}|\omega)$ defined over a sample space $\mathcal{Y}$ and a parameter space $\boldsymbol{\Omega}$. The parameter space is partitioned into disjoint sets $\boldsymbol{\Omega}_0$ representing the null hypothesis and $\boldsymbol{\Omega}_1$ representing the alternative. A set $\mathcal{A}$ or $\mathcal{B}$ in the sample space represents the acceptance of the null hypothesis.

Definition 10.1. *A test A is* as good as *a test B if*

$$\Pr(\mathcal{A}|\omega) \geq \Pr(\mathcal{B}|\omega), \qquad \omega \in \boldsymbol{\Omega}_0, \tag{10.1}$$

$$\Pr(\mathcal{A}|\omega) \leq \Pr(\mathcal{B}|\omega), \qquad \omega \in \boldsymbol{\Omega}_1. \tag{10.2}$$

Definition 10.2. *A test A is* better than *B if (10.1) and (10.2) hold with strict inequality for at least one ω.*

Definition 10.3. *A test A is* admissible *if there is no test B that is better than A.*

See Section 5.6.2 of Anderson (2003), for example. If the sets $\mathcal{A}$ and $\mathcal{B}$ are invariant with respect to a group of transformations, the test with acceptance set $\mathcal{A}$ is known as an *admissible invariant test.*

Theorem 10.1. *The Bayes test with acceptance region (9.7) is an admissible invariant test of H_0 vs. H_1.*

Proof. Let the Bayes test for $\Pr\{\theta = \theta_0\}$ and the density $p(\cdot)$ be given by (9.7), resulting in the average risk $\mathrm{R}^*[p(\cdot), \lambda, d_B]$. If this test is not admissible, then there is a test d^* that is better than d_B, that is,

$$\mathrm{R}^*[p(\cdot), \lambda, d^*] \leq \mathrm{R}^*[p(\cdot), \lambda, d_B] \tag{10.3}$$

for all θ and λ with strict inequality for some θ and λ. However, this assertion contradicts the construction of the Bayes test d_B. □

The conclusion is that the LR test is an admissible invariant test.

The invariance involved here is with respect to certain linear transformations. This consideration is a generalization of the notion that the questions at issue do not depend on the unit of measurement; for example, inches vs. feet vs. meters or pounds vs. kilograms or radians vs. degrees. The linear transformations do not affect the inference problems for which the model is used.

11 Admissibility over all tests

11.1 General theorem

Now we consider admissibility with respect to all tests. We want to show that the best invariant test of $\theta = \theta_0$ is admissible within the class of all tests; in particular, a LR test is admissible within the class of all tests. The idea is that a family of tests -invariant or not- can be transformed to a family of *randomized invariant* tests; if the original family of (randomized) invariant tests is admissible within the class of all tests.

We apply the so-called Hunt–Stein theorem to the effect that the best invariant test is admissible in the class of all tests if the group transformations defining invariance is finite or compact. See Zaman (1996), Section 7.9, or Lehmann (1986), Theorem 7 of Chapter 3. The proofs of such theorems are based on the argument that the randomization of the noninvariant tests yields an invariant test that is as good as the noninvariant test.

In the model

$$\boldsymbol{Q} = \lambda \boldsymbol{\eta} \boldsymbol{\alpha}' + \boldsymbol{W} \tag{11.1}$$

for fixed λ, each parameter vector $\boldsymbol{\eta}$ and $\boldsymbol{\alpha}$ take values in closed sets $\boldsymbol{\eta}'\boldsymbol{\eta} = 1$ and $\boldsymbol{\alpha}'\boldsymbol{\alpha} = 1$,which are therefore compact and satisfy the *Hunt–Stein* conditions.

Theorem 11.1. *The LR test of $\theta = \theta_0$ is admissible in the set of all tests.*

11.2 An example

Cosider the model in which θ can take on a finite number of values.

The possible parameter values are

$$\theta = 0, \frac{1}{N}2\pi, \frac{2}{N}2\pi, \ldots, \frac{N-1}{N}2\pi. \tag{11.2}$$

Consider the group of transformations

$$\theta \longrightarrow \theta + \frac{j}{N}2\pi, \quad t \longrightarrow t + \frac{j}{N}2\pi, \qquad j = 0, 1, \ldots, N-1. \tag{11.3}$$

Let these values of θ be labelled as $\theta_0^*, \theta_1^*, \ldots, \theta_{N-1}^*$. Each of them corresponds to a null hpothesis. Define a test of the hypothesis $\theta = \theta_k^*$ by the acceptance region $A_k^* = A_k^*(t, r_1, r_2)$ in the space of t, r_1, r_2. The set of tests is an *invariant* set if

$$A_k^*(t - \theta_k^*, r_1, r_2) = A_j^*(t - \theta_k^*, r_1, r_2) \tag{11.4}$$

for $k, j = 0, 1, \cdots, N-1$.

The LR test of the hypothesis $\theta = \theta_i^*$ against the alternative $\theta = \theta_j^*$ for some $j = 0, 1, \ldots, N-1$ is the Bayes solution for the hypothesis $\theta = \theta_i^*$ for prior probabilities

$$\Pr\{\theta = \theta_j^*\} = \frac{1}{N}, \qquad j = 0, 1, \ldots, N-1. \tag{11.5}$$

Non-invariant tests. Suppose the set of tests are not necessarily invariant; that is, (11.3) does not necessarily hold. We can randomize these N tests by defining an invariant randomized test.

The acceptance region $A_k^*(t, r_1, r_2)$ can be adapted to test $\theta = \theta_i^*$ by subtracting θ_k^* from $A_k^*(t, r_1, r_2)$ and adding θ_i^*, which is the region $A_k^*(t - \theta_k^* + \theta_i^*, r_1, r_2)$. A randomized test for the null hypothesis $\theta = \theta_i^*$ has acceptance region

$$\frac{1}{N} \sum_{k=0}^{N-1} A_k^*(t - \theta_k^* + \theta_i^*, r_1, r_2). \tag{11.6}$$

The set of such tests for $\theta_i^*,\ i = 0, 1, \ldots, N-1$ is an invariant set.

Lemma 11.1. *If a test with an invariant family of acceptance regions $A_0, A_1, \ldots, A_{N-1}$ is admissible in the set of invariant tests, it is admissible in the set of all tests.*

Proof by contradiction. Suppose $\bar{A}_0, \ldots, \bar{A}_{N-1}$ is a family of better tests (not necessarily invariant). Then the invariant randomized tests based on $\bar{A}_0, \ldots, \bar{A}_{N-1}$ is better than the family of $A_0, \ldots, A_{N-1}$. But this contradicts the assumption that $A_0, \ldots, A_{N-1}$ is admissible in the set of invariant tests. □

12 Comments

12.1 Invariance with respect to linear transformations of exogenous variables

In the model (2.1) $\boldsymbol{Y} = \boldsymbol{Z\Pi} + \boldsymbol{V}$ a linear transformation of $\boldsymbol{Z}$ and $\boldsymbol{\Pi}$ ($\boldsymbol{Z}^+ = \boldsymbol{ZC}$ and $\boldsymbol{\Pi}^+ = \boldsymbol{C}^{-1}\boldsymbol{\Pi}$) leaves $\boldsymbol{Z\Pi}$ invariannt and hence does not affect the model.

Similarly, the transformation does not affect the equation $\boldsymbol{\Pi\beta} = \boldsymbol{0}$, in particular the null hypothesis $\boldsymbol{\Pi\beta}_0 = \boldsymbol{0}$. This property is a generalization of the idea that the model and the problem do not depend on the units of measurement. This property implies that a test can be based on $\boldsymbol{G} = \boldsymbol{P}'\boldsymbol{AP}$.

12.2 Invariance with respect to orthogonal transformations of endogenous variables

When $\boldsymbol{\Omega} = \boldsymbol{I}$ is assumed, an orthogonal transformation of the disturbance $\boldsymbol{V} \to \boldsymbol{VO}$ and a corresponding transformation of $\boldsymbol{\beta}$, $\boldsymbol{\beta} \to \boldsymbol{O}'\boldsymbol{\beta}$ and of the null hypothesis $\boldsymbol{\beta}_0 \to \boldsymbol{O}'\boldsymbol{\beta}_0$ do not affect the equations, $\boldsymbol{\beta} = \boldsymbol{\beta}_0$ and $\boldsymbol{\beta}'\boldsymbol{\beta} = 1$. In the $G-$space this transformation is a rotation of coordinates.

12.3 Conventional normalization

The *conventional normalization* of $\boldsymbol{\beta}$ which satisfies $\boldsymbol{\Pi\beta} = \boldsymbol{0}$ is to set one coefficient of $\boldsymbol{\beta}$, say the first component equal to 1; that is

$$\boldsymbol{\beta} = \begin{bmatrix} 1 \\ -\boldsymbol{\beta}_2 \end{bmatrix}. \tag{12.1}$$

When $\mathcal{E}\boldsymbol{Y} = \boldsymbol{Z\Pi}$ is replaced by $\mathcal{E}\boldsymbol{Q} = \lambda\boldsymbol{\eta}\boldsymbol{\alpha}'$ and $G = 2$, the condition $\boldsymbol{\Pi\beta} = \boldsymbol{0}$ is replaced by

$$0 = \boldsymbol{\alpha}'\boldsymbol{\beta} = \alpha_1 - \alpha_2\beta_2. \tag{12.2}$$

The null hypothesis $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0$ is $H_0 : \beta_2 = \beta_2^0$, that is,

$$H_0 : 0 = \alpha_1 - \alpha_2\beta_2^0 = -\sin\theta_0 - \beta_2^0\cos\theta_0, \tag{12.3}$$

which is $H_0 : \tan\theta_0 = -\beta_2^0$. Thus the admissibility of the LR test given λ shows that the LR test dominates a test based on the Two-Stage Least Squares estimator.

13 Conclusions

13.1 Estimation

Anderson, Stein, and Zaman (1985) considered the estimation of $\boldsymbol{\eta}$ and $\boldsymbol{\alpha}$ when the loss of estimation of $\boldsymbol{\alpha}$ by $\hat{\boldsymbol{\alpha}}$ was defined as

$$L(\boldsymbol{\alpha}, \hat{\boldsymbol{\alpha}}) = 1 - (\boldsymbol{\alpha}'\hat{\boldsymbol{\alpha}})^2 = \sin^2(\hat{\theta} - \theta). \tag{13.1}$$

The loss function is invariant with respect to transformations (3.1) and (3.10). When $G = 2$, this is the model treated here. The estimator t of θ is the LIMLK estimator. Corollary 1 of Anderson, Stein, and Zaman (1985) states that the LIMLK estimator is admissible for the loss function (13.1) and every fixed λ and hence for all λ.

The risk of an estimator is $\mathcal{E}\sin^2(\hat{\theta} - \theta)$ which is a function of $\lambda, \boldsymbol{\eta}$, and $\boldsymbol{\alpha}$. Admissibility of the LIMLK estimator means that there is no estimator for which $\mathcal{E}\sin^2(\hat{\theta} - \theta)$ is as small or smaller than for LIMLK for all $\lambda, \boldsymbol{\eta}$, and $\boldsymbol{\alpha}$.

With the normalization $\boldsymbol{\alpha}'\boldsymbol{\alpha} = 1 = \boldsymbol{\beta}'\boldsymbol{\beta}$ an estimator of $\boldsymbol{\alpha}$ or $\boldsymbol{\beta}$ implies an estimator of θ. Since $\sin x = x - x^3/3! + \cdots$, for small $\hat{\theta} - \theta$ the loss function is approximately $(\hat{\theta} - \theta)^2$. It is periodic with a period of 2π, which is appropriate for an undirected line.

13.2 Testing

The acceptance set (in terms of r_1, r_2, t) of the LR test can be written

$$\mathcal{A}: \ \sin^2(t - \theta_0) \leq a(r_1, r_2, \lambda). \tag{13.2}$$

Let the acceptance set of a competing invariant test be

$$\mathcal{B}: \ t - \theta_0 \in b(r_1, r_2, \lambda). \tag{13.3}$$

Theorem 10.1 states that

$$\Pr\{\mathcal{A}|\theta_0, \lambda\} \geq \Pr\{\mathcal{B}|\theta_0, \lambda\}, \tag{13.4}$$

$$\Pr\{\mathcal{B}|\theta, \lambda\} \leq \Pr\{\mathcal{B}|\theta, \lambda\}, \qquad \theta \neq \theta_0, \tag{13.5}$$

with a strict inequality for some θ and λ.

13.3 A more general model

Instead of (2.9) consider (2.4) with the hypothesis $H_0 : \boldsymbol{\beta}_1 = \boldsymbol{\beta}_0$, where $\boldsymbol{\beta}_1$ satisfies (2.8). Let

$$\boldsymbol{Z}_{2.1} = \boldsymbol{Z}_2 - \boldsymbol{Z}_1\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12}, \tag{13.6}$$

where $\boldsymbol{A}$ has been partitioned into K_1 and K_2 rows and columns. Then the relevant part of the reduced form (2.6) can be written

$$\boldsymbol{Y}_1 = \boldsymbol{Z}_1 \left(\boldsymbol{\Pi}_{11} + \boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12}\boldsymbol{\Pi}_{21}\right) + \boldsymbol{Z}_{2.1}\boldsymbol{\Pi}_{21} + \boldsymbol{V}_1. \tag{13.7}$$

The sufficient statistics are $\boldsymbol{A}_{11}^{-1}\boldsymbol{Z}_1'\boldsymbol{Y}_1$ and $\boldsymbol{P}_2 = \boldsymbol{A}_{22.1}^{-1}\boldsymbol{Z}_{2.1}'\boldsymbol{Y}_1$, where

$$\boldsymbol{A}_{22.1} = \boldsymbol{Z}_{2.1}'\boldsymbol{Z}_{2.1} = \boldsymbol{A}_{22} - \boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12}, \tag{13.8}$$

and they are independent. The developments above proceed with $\boldsymbol{Z}$ replaced by $\boldsymbol{Z}_{2.1}$, $\boldsymbol{Y}$ by $\boldsymbol{Y}_1$, etc.

References

[1] Abramowitz, M. and I. Stegun (1972), *Handbook of Mathematical Functions*, Dover, New York.

[2] Anderson, T.W. (1976), "Estimation of Linear Functional Relationships : Approximate Distributions and Connections to Simultaneous Equations in Econometrics," *Journal of the Royal Statistical Society Series B* **38**, 1–36.

[3] Anderson, T.W. (2003), *An Introduction to Multivariate Statistical Analysis*, 3rd Edition, John Wiley and Sons, Inc., New York.

[4] Anderson, T.W. and M.A. Girshick (1944), "Some extensions of the Wishart distribution," *Annals of Mathematical Statistics* **15**, 345–357. (Corrections : (1964) **35**, 923–925).

[5] 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. `www.e.u-tokyo.ac.jp/cirje/research/dp/2007`

[6] Anderson, T.W. and N. Kunitomo (2009), "Testing a Hypothesis about a Structural Coefficient in a Simultaneous Equation Model for Known Covariance Matrix" (In honor of Raul P. Mentz), *Estadistica* **61**, 7–16.

[7] 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**, 46–63.

[8] 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* **21**, 570–582.

[9] Anderson, T.W. and T. Sawa (1977), "Two-Stage Least Squares : in which direction should the residuals be minimized?" *Journal of the American Statistical Association* **72**, 187–191.

[10] Anderson, T.W., C. Stein and A. Zaman (1985), "Best Invariant Estimation of a Direction Parameter," *Annals of Statistics* **13**, 526–533.

[11] Andrews, D.W.H., M.J. Moreira, and J.H. Stock (2006), "Optimal Two-Sided Invariant Similar Tests for Instrumental Variables Regression," *Econometrica* **74**, 715–752.

[12] Chamberlain, G. (2007), "Decision Theory Applied to an Instrumental Variables Model," *Econometrica* **75-3**, 609–652.

[13] Creasy, M.A. (1956), "Confidence Limits for the Gradient in the Linear Functional Relationship," *Journal of the Royal Statistical Society Series B* **18**, 65–69.

[14] Lehmann, E. (1986), *Testing Statistical Hypothesis*, 2nd Edition, Springer.

[15] Moreira, M. (2003), "A Conditional Likelihood Ratio Test for Structural Models," *Econometrica* **71**, 1027–1048.

[16] Zaman, A. (1996), *Statistical Foundations for Econometric Techniques*, Adademic Press.

A Jacobian

The representation of $\boldsymbol{G} = \boldsymbol{O}_t \boldsymbol{R} \boldsymbol{O}_t'$ in components is

$$\text{(A.1)} \qquad \begin{bmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{bmatrix} = \begin{bmatrix} r_1 \cos^2 t + r_2 \sin^2 t & (r_1 - r_2)\cos t \sin t \\ (r_1 - r_2)\cos t \sin t & r_1 \sin^2 t + r_2 \cos^2 t \end{bmatrix}.$$

The matrix of partial derivatives of g_{11}, g_{22}, g_{12} with respect to r_1, r_2 and t is

$$\text{(A.2)} \qquad \begin{bmatrix} \cos^2 t & \sin^2 t & -2(r_1 - r_2)\cos t \sin t \\ \sin^2 t & \cos^2 t & 2(r_1 - r_2)\cos t \sin t \\ \cos t \sin t & -\cos t \sin t & (r_1 - r_2)(\cos^2 t - \sin^2 t) \end{bmatrix}.$$

The Jacobian of the transformation is the absolute value of the determinant of (A.2) which is $r_2 - r_1$.

B The noncentral Wishart distribution

Let $\boldsymbol{Q} = \lambda \boldsymbol{\eta} \boldsymbol{\alpha}' + \boldsymbol{W}$, and

$$\text{(B.1)} \qquad \boldsymbol{Q} = \begin{bmatrix} \boldsymbol{q}_1' \\ \boldsymbol{Q}_2 \end{bmatrix}, \quad \boldsymbol{\eta} = \begin{bmatrix} 1 \\ \mathbf{0} \end{bmatrix}, \quad \boldsymbol{\alpha} = \begin{bmatrix} 1 \\ \mathbf{0} \end{bmatrix}, \quad \boldsymbol{W} = \begin{bmatrix} \boldsymbol{w}_1' \\ \boldsymbol{W}_2 \end{bmatrix},$$

where $\boldsymbol{Q}_2$ and $\boldsymbol{W}_2$ are $(K-1) \times G$, $\boldsymbol{q}_1$ and $\boldsymbol{w}_1$ are $G \times 1$, $\boldsymbol{\eta}$ is $K \times 1$ and $\boldsymbol{\alpha}$ is $G \times 1$. Note that $\boldsymbol{\eta}'\boldsymbol{\eta} = 1 = \boldsymbol{\alpha}'\boldsymbol{\alpha}$. The rows of $\boldsymbol{W}$ are independently normally distributed with means $\mathbf{0}$ and covariance matrix $\boldsymbol{I}_G$. Then $\boldsymbol{Q}_2'\boldsymbol{Q}_2 = \boldsymbol{G}_2$ has a (central) Wishart distribution $W(\boldsymbol{I}_G, K-1)$ with density

$$\text{(B.2)} \qquad \frac{|\boldsymbol{G}_2|^{\frac{1}{2}(K-G-2)} e^{-\frac{1}{2}\mathrm{tr}\boldsymbol{G}_2}}{2^{\frac{1}{2}(K-1)G} \pi^{G(G-1)/4} \prod_{i=1}^{G} \Gamma[\frac{1}{2}(K-i)]}$$

(Anderson, 2003, Th. 7.2.2). The vector $\boldsymbol{q}_1' = (q_{11}, \boldsymbol{q}_{12}')$ has the density

$$\text{(B.3)} \qquad \frac{1}{(2\pi)^{\frac{1}{2}G}} e^{-\frac{1}{2}(q_{11}-\lambda)^2 - \frac{1}{2}\boldsymbol{q}_{12}'\boldsymbol{q}_{12}}.$$

The joint density of the matrix $\boldsymbol{G}_2$ and the vector $\boldsymbol{q}_1'$ is the product of (B.2) and (B.3). The joint density of $\boldsymbol{G} = \boldsymbol{q}_1\boldsymbol{q}_1' + \boldsymbol{G}_2$ and $\boldsymbol{q}_1$ is

$$\text{(B.4)} \qquad \frac{|\boldsymbol{G} - \boldsymbol{q}_1\boldsymbol{q}_1'|^{\frac{1}{2}(K-G-2)} e^{-\frac{1}{2}\mathrm{tr}\boldsymbol{G} + \lambda q_{11} - \lambda^2/2}}{2^{\frac{1}{2}KG}\pi^{G(G+1)/4}\prod_{i=1}^{G}\Gamma[\frac{1}{2}(K-i)]} = \frac{|\boldsymbol{G}|^{\frac{1}{2}(K-G-2)}(1 - \boldsymbol{q}_1'\boldsymbol{G}^{-1}\boldsymbol{q}_1)^{\frac{1}{2}(K-G-2)} e^{-\frac{1}{2}\mathrm{tr}\boldsymbol{G} + \lambda q_{11} - \lambda^2/2}}{2^{\frac{1}{2}KG}\pi^{G(G+1)/4}\prod_{i=1}^{G}\Gamma[\frac{1}{2}(K-i)]}.$$

See Corollary A.3.1 of Anderson (2003), for example.

The noncentral Wishart density of $\boldsymbol{G}$ is the integral of (B.4) with respect to the vector $\boldsymbol{q}_1' = (q_{11}, \boldsymbol{q}_{12}')$ over the range for which $1 - \boldsymbol{q}_1'\boldsymbol{G}^{-1}\boldsymbol{q}_1$ is positive. Anderson and Girshick (1944) carried out the algebraic details of this integration.

Theorem B.1. *The density of* $\boldsymbol{G} = \boldsymbol{Q}'\boldsymbol{Q}$, *where* $\boldsymbol{Q} = \lambda\boldsymbol{\eta}\boldsymbol{\alpha}' + \boldsymbol{W}$, $\boldsymbol{\eta} = (1, \boldsymbol{0})'$ *and* $\boldsymbol{\alpha} = (1, \boldsymbol{0})'$, *is*

$$\text{(B.5)} \qquad \frac{e^{-(1/2)\lambda^2 - (1/2)tr\boldsymbol{G}}}{2^{(1/2)KG - (1/2)(K-2)}\pi^{G(G-1)/4}\prod_{i=1}^{G-1}\Gamma[(1/2)(K-i)]}|\boldsymbol{G}|^{\frac{1}{2}(K-G-1)} I^*_{\frac{1}{2}(K-2)}(\lambda^2 g_{11})$$

where

$$\text{(B.6)} \qquad I^*_{\frac{1}{2}(K-2)}(z^2) = \sum_{j=0}^{\infty}\left(\frac{z^2}{4}\right)^j \frac{1}{j!\Gamma(\frac{K}{2} + j)}.$$