

Admissible Significance Tests in Simultaneous Equation Models*

T. W. Anderson[†]
Stanford University

December 17, 2013

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. In particular, this likelihood ratio test dominates a test based on the Two-Stage Least Squares estimator.

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

JEL classification(s): C01, C12, C30

1 Introduction

There is a considerable literature on statistical inference concerning a single structural equation in a simultaneous equation model. Much of the literature concerns estimation of

*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 sponsored by the London School of Economics and University College, London, on October 29, 2009. This version was presented to the Haavelmo Centennial Symposium, Oslo, on December 14, 2011.

[†]Department of Statistics and Department of Economics, Stanford University, Stanford, CA 94305, USA; tw@stanford.edu

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 is known, the corresponding estimator is known as LIMLK. Anderson, Stein, and Zaman (1985) showed that the LIMLK estimator is admissible for a suitable loss function in a model corresponding to two simultaneous equations. They showed that the LIMLK estimator was the best estimator invariant under linear transformations that leave the model and loss function invariant. It follows that the LIMLK estimator is admissible in the class of all estimators including randomized estimators.

Anderson and Rubin (1949) also suggested a test of the null hypothesis, say, H_0 , that the vector of coefficients of the endogenous variables, say, β , is a specified vector, say, β_0 ; the alternative hypothesis, say H_2 , is that β is unrestricted. The test is admissible if the equation is *just identified*, but not if the equation is over-identified. Anderson and Kunitomo (2007) derived an alternative 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; see also Anderson and Kunitomo, 2009.)

Anderson (1976, 1984) 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. Moreira (2003) derived the test in more generality; he called the test the *conditional likelihood ratio test*.

The current paper treats the testing problem when the disturbances matrix is known and the number of endogenous variables in the single equation is two. When the disturbance covariance matrix is known, a sufficient statistic is the sample regression of the dependent variables on the independent variables, say \mathbf{P} , and the sample covariance of the independent variables, say \mathbf{A} . The likelihood ratio criterion is a function of $\mathbf{P}'\mathbf{A}\mathbf{P}$, the distribution of which depends only on a noncentrality parameter, say λ . It is shown that the likelihood ratio test is identical to a Bayes test of H_0 vs H_1 conditional on this parameter λ . Thus for each λ this conditional test of H_0 vs H_1 is the uniformly most powerful invariant test; that is, the (conditional) test is admissible among tests for a specific noncentrality parameter λ . Since this comparison does not depend on the value of λ , it holds for *every* λ .

Now consider the class of all tests of H_0 vs H_1 , not necessarily invariant, but including randomized tests. By a version of the “Hunt–Stein theorem” the likelihood ratio test is admissible among all tests. This means that there is no test with better significance level and/or better power. In particular, the Two-Stage Least Squares is inferior as an estimator and yields an inferior test procedure.

It should be noted that the admissibility properties in this paper are “exact,” that is, the results are not asymptotic or approximate. However, the admissibility property is a comparison of tests; it does not establish a distribution or significance point.

2 A simultaneous equation model

The observed data consist of a $T \times G$ matrix of endogenous or dependent or nonstochastic 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

$$\mathbf{Y} = \mathbf{Z}\mathbf{\Pi} + \mathbf{V}, \quad (2.1)$$

where $\mathbf{\Pi}$ is a $K \times G$ matrix of parameters and \mathbf{V} is a $T \times G$ matrix of unobservable disturbances. The rows of \mathbf{V} are assumed independent; each row has a normal distribution $N(\mathbf{0}, \mathbf{\Omega})$.

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

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

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

$$\mathbf{H} = (\mathbf{Y} - \mathbf{Z}\mathbf{P})'(\mathbf{Y} - \mathbf{Z}\mathbf{P}) = \mathbf{Y}'\mathbf{Y} - \mathbf{P}'\mathbf{A}\mathbf{P} \quad (2.3)$$

and $\mathbf{A} = \mathbf{Z}'\mathbf{Z}$. The matrices \mathbf{P} and \mathbf{H} constitute sufficient statistics for the model. See Haavelmo (1944).

A structural or behavioral equation may involve a $T \times G_1$ subset of the endogenous

variables \mathbf{Y}_1 , a $T \times K_1$ subset of the exogenous variables \mathbf{Z}_1 , and a $T \times G_1$ subset of disturbances \mathbf{V}_1 . The structural equation of interest is

$$\mathbf{Y}_1\boldsymbol{\beta}_1 = \mathbf{Z}_1\boldsymbol{\gamma}_1 + \mathbf{u}, \quad (2.4)$$

where $\mathbf{u} = \mathbf{V}_1\boldsymbol{\beta}_1$ and $\mathbf{V} = (\mathbf{V}_1, \mathbf{V}_2)$. A component of \mathbf{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

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

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

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

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

$$\begin{bmatrix} \boldsymbol{\gamma}_1 \\ \mathbf{0} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Pi}_{11} & \boldsymbol{\Pi}_{12} \\ \boldsymbol{\Pi}_{21} & \boldsymbol{\Pi}_{22} \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}_1 \\ \mathbf{0} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Pi}_{11}\boldsymbol{\beta}_1 \\ \boldsymbol{\Pi}_{21}\boldsymbol{\beta}_1 \end{bmatrix}. \quad (2.7)$$

The second submatrix of (2.7),

$$\boldsymbol{\Pi}_{21}\boldsymbol{\beta}_1 = \mathbf{0}, \quad (2.8)$$

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. See Section 6. 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 γ_1 is vacuous, that is $K_1 = 0$. Furthermore, we set $G_2 = 0$ so that $G = G_1$. Then the reduced form and structural equation are

$$\mathbf{Y} = \mathbf{Z}\mathbf{\Pi} + \mathbf{V}. \quad (2.9)$$

3 Invariance and normalization

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

$$\mathbf{Z}^+ = \mathbf{Z}\mathbf{C}, \quad \mathbf{\Pi}^+ = \mathbf{C}^{-1}\mathbf{\Pi} \quad (3.1)$$

for \mathbf{C} being nonsingular. Then

$$\mathbf{Z}^+\mathbf{\Pi}^+ = \mathbf{Z}\mathbf{\Pi}, \quad \mathbf{A}^+ = \mathbf{C}'\mathbf{A}\mathbf{C}, \quad \mathbf{P}^+ = \mathbf{C}^{-1}\mathbf{P}, \quad (3.2)$$

and

$$\mathbf{G}^+ = \mathbf{P}^{+'}\mathbf{A}^+\mathbf{P}^+ = \mathbf{P}'\mathbf{A}\mathbf{P} = \mathbf{G}, \quad \mathbf{H}^+ = \mathbf{Y}'\mathbf{Y} - \mathbf{P}^{+'}\mathbf{A}^+\mathbf{P}^+ = \mathbf{H}. \quad (3.3)$$

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

$$\boldsymbol{\beta}'\boldsymbol{\Omega}\boldsymbol{\beta} = 1, \quad (3.4)$$

which determines the constant except for sign. The model (2.1), $\mathbf{\Pi}\boldsymbol{\beta} = \mathbf{0}$, and (3.4) is invariant with respect to transformations

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

and

$$\boldsymbol{\Omega}^* = \boldsymbol{\Phi}'\boldsymbol{\Omega}\boldsymbol{\Phi}, \quad \boldsymbol{\beta}_0^* = \boldsymbol{\Phi}^{-1}\boldsymbol{\beta}_0, \quad (3.6)$$

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

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

and

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

Now we consider the model (2.1) and $\boldsymbol{\Pi}\boldsymbol{\beta} = \mathbf{0}$ when $\boldsymbol{\Omega}$ (the covariance matrix of a row of \boldsymbol{V}) is known except for a constant. In this case we can make a transformation (3.5) and (3.6) so $\boldsymbol{\Omega} = \sigma^2\boldsymbol{I}$. Then the first equation in (3.6) is

$$\boldsymbol{I} = \boldsymbol{O}'\boldsymbol{O}, \quad (3.9)$$

that is, the invariance with respect to transformations (3.7) and (3.8) 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

$$\begin{aligned} \boldsymbol{Y}^* &= \boldsymbol{Y}\boldsymbol{O}, \quad \boldsymbol{\Pi}^* = \boldsymbol{\Pi}\boldsymbol{O}, \quad \boldsymbol{\beta}^* = \boldsymbol{O}'\boldsymbol{\beta}, \quad \boldsymbol{V}^* = \boldsymbol{V}\boldsymbol{O}, \\ \boldsymbol{\beta}_0^* &= \boldsymbol{O}'\boldsymbol{\beta}_0, \quad \boldsymbol{\beta}^{*'}\boldsymbol{\beta}^* = \boldsymbol{\beta}'\boldsymbol{\beta} = 1. \end{aligned} \quad (3.10)$$

$$\begin{aligned} \boldsymbol{P}^* &= \boldsymbol{P}\boldsymbol{O}, \quad \boldsymbol{H}^* = \boldsymbol{O}'\boldsymbol{H}\boldsymbol{O}, \\ \boldsymbol{P}^{*'}\boldsymbol{A}\boldsymbol{P}^* &= \boldsymbol{O}'\boldsymbol{P}'\boldsymbol{A}\boldsymbol{P}\boldsymbol{O}. \end{aligned} \quad (3.11)$$

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

The reader's intuition can be helped by thinking of the case $G = 2$. A row of \boldsymbol{Y} is a point on a two-dimensional graph, say a map. The rotation by \boldsymbol{O} corresponds to rotating the map, that is, looking at the map from a different point of view. In this study there is no preferred coordinate system.

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

The main part of this paper concerns the model for $\mathbf{\Omega} = \sigma^2 \mathbf{I}_2$ and

$$G_1 = G = 2, \quad G_2 = 0, \quad K_1 = 0, \quad K_2 = K \geq 2. \quad (4.1)$$

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

$$\mathbf{\Pi}\boldsymbol{\beta} = \mathbf{0}, \quad \boldsymbol{\beta}'\boldsymbol{\beta} = 1. \quad (4.2)$$

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

$$\boldsymbol{\beta} = \begin{bmatrix} \cos \theta \\ \sin \theta \end{bmatrix}, \quad -\pi \leq \theta \leq \pi. \quad (4.3)$$

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

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

$$\mathbf{\Pi} = \boldsymbol{\gamma}\boldsymbol{\alpha}', \quad (4.4)$$

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

$$\boldsymbol{\alpha} = \begin{bmatrix} -\sin \theta \\ \cos \theta \end{bmatrix}. \quad (4.5)$$

Note that

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

is an orthogonal matrix. The model is identified when $\boldsymbol{\gamma} \neq \mathbf{0}$. Since $\mathbf{\Omega}$ is known, a sufficient statistic in the model is \mathbf{P} .

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

$$\mathbf{\Pi}^+ = \boldsymbol{\nu}\boldsymbol{\alpha}', \quad \mathbf{P}'\mathbf{A}\mathbf{P} = \mathbf{Q}'\mathbf{Q}, \quad \boldsymbol{\nu} = \mathbf{C}^{-1}\boldsymbol{\gamma}, \quad (4.7)$$

and $\boldsymbol{\alpha}'\boldsymbol{\alpha} = 1$. The model is

$$\mathbf{Q} = \boldsymbol{\nu}\boldsymbol{\alpha}' + \mathbf{W}. \quad (4.8)$$

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

$$\mathcal{E}(\mathbf{w}_1\mathbf{w}_1') = \mathcal{E}(\mathbf{w}_2\mathbf{w}_2') = \sigma^2\mathbf{I}_K, \quad \mathcal{E}(\mathbf{w}_1\mathbf{w}_2') = \mathbf{0}. \quad (4.9)$$

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 $\theta = \theta_0$.

Define λ by $\boldsymbol{\nu}'\boldsymbol{\nu} = \lambda^2$. Then $\boldsymbol{\nu} = \lambda\boldsymbol{\eta}$, where $\boldsymbol{\eta}'\boldsymbol{\eta} = 1$. We call $\lambda^2 = \text{tr}\mathcal{E}\mathbf{Q}\mathcal{E}\mathbf{Q}' = \text{tr}\boldsymbol{\nu}\boldsymbol{\alpha}'\boldsymbol{\alpha}\boldsymbol{\nu}' = \text{tr}\boldsymbol{\nu}\boldsymbol{\nu}' = \boldsymbol{\nu}'\boldsymbol{\nu}$ the *noncentrality* parameter.

The density of \mathbf{Q} is

$$\begin{aligned} \frac{1}{(2\pi\sigma^2)^K} e^{-\frac{1}{2}\text{tr}\mathbf{W}'\mathbf{W}/\sigma^2} &= \frac{1}{(2\pi\sigma^2)^K} e^{-\frac{1}{2}\text{tr}(\mathbf{Q}-\boldsymbol{\nu}\boldsymbol{\alpha}')'(\mathbf{Q}-\boldsymbol{\nu}\boldsymbol{\alpha}')/\sigma^2} \\ &= \frac{1}{(2\pi\sigma^2)^K} e^{-\frac{1}{2}\text{tr}(\mathbf{Q}-\lambda\boldsymbol{\eta}\boldsymbol{\alpha}')'(\mathbf{Q}-\lambda\boldsymbol{\eta}\boldsymbol{\alpha}')/\sigma^2} \\ &= \frac{1}{(2\pi\sigma^2)^K} e^{(-\frac{1}{2}\text{tr}\mathbf{Q}'\mathbf{Q}-\frac{1}{2}\lambda^2+\lambda\boldsymbol{\eta}'\mathbf{Q}\boldsymbol{\alpha}')/\sigma^2} \end{aligned} \quad (4.10)$$

since $\lambda^2\text{tr}(\boldsymbol{\alpha}\boldsymbol{\eta}'\boldsymbol{\eta}\boldsymbol{\alpha}') = \lambda^2$ and $\lambda\text{tr}(\boldsymbol{\alpha}\boldsymbol{\eta}'\mathbf{Q}) = \lambda\boldsymbol{\eta}'\mathbf{Q}\boldsymbol{\alpha}$.

5 The likelihood ratio test

In this section we derive the likelihood ratio test of $H_0 : \boldsymbol{\alpha} = \boldsymbol{\alpha}_0$ vs $H_1 : \mathcal{E}\mathbf{Q} = \boldsymbol{\nu}\boldsymbol{\alpha}'$ when \mathbf{Q} has the likelihood defined by (4.10). The derivative of the logarithm of (4.10) with respect to $\boldsymbol{\nu}$ set equal to $\mathbf{0}$ gives

$$\mathbf{Q}\boldsymbol{\alpha} = \boldsymbol{\nu}. \quad (5.1)$$

Then

$$\boldsymbol{\nu}'\mathbf{Q}\boldsymbol{\alpha} = \boldsymbol{\nu}'\boldsymbol{\nu} = \boldsymbol{\alpha}'\mathbf{Q}'\mathbf{Q}\boldsymbol{\alpha} = \boldsymbol{\alpha}'\mathbf{G}\boldsymbol{\alpha}, \quad (5.2)$$

where $\mathbf{G} = \mathbf{Q}'\mathbf{Q}$. The logarithm of the likelihood maximized with respect to $\boldsymbol{\nu}$ is $-K\log(2\pi)$ plus

$$-\frac{1}{2} [\text{tr}\mathbf{G} - \boldsymbol{\alpha}'\mathbf{G}\boldsymbol{\alpha}] = -\frac{1}{2}\text{tr}\mathbf{G} + \frac{1}{2}\boldsymbol{\alpha}'\mathbf{G}\boldsymbol{\alpha}. \quad (5.3)$$

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

$$\mathbf{R} = \begin{bmatrix} r_1 & 0 \\ 0 & r_2 \end{bmatrix}, \quad \mathbf{O}_t = \begin{bmatrix} \cos t & -\sin t \\ \sin t & \cos t \end{bmatrix} = (\boldsymbol{\beta}_t, \boldsymbol{\alpha}_t); \quad (5.4)$$

that is, $r_1 \leq r_2$ are the eigenvalues of \mathbf{G} and $(\cos t, \sin t)'$ and $(-\sin t, \cos t)'$ are the corresponding eigenvectors of \mathbf{G} . Under H_1 the likelihood is maximized by $\boldsymbol{\alpha} = (-\sin t, \cos t)'$ and (5.3) is $-\frac{1}{2}r_1$. Under H_0 the likelihood is maximized by $\boldsymbol{\alpha} = \boldsymbol{\alpha}_0 = (-\sin \theta_0, \cos \theta_0)'$. Then

$$\begin{aligned} \boldsymbol{\alpha}_0' \mathbf{G} \boldsymbol{\alpha}_0 &= \boldsymbol{\alpha}_0' \mathbf{O}_t \mathbf{R} \mathbf{O}_t' \boldsymbol{\alpha}_0 \\ &= r_1 \sin^2(t - \theta_0) + r_2 \cos^2(t - \theta_0) \\ &= (r_1 - r_2) \sin^2(t - \theta_0) + r_2. \end{aligned} \quad (5.5)$$

The likelihood ratio criterion for testing $\theta = \theta_0$ in the model (4.10) is

$$e^{-\frac{1}{2}(r_2 - r_1) \sin^2(t - \theta_0)}. \quad (5.6)$$

The acceptance region of the test is

$$(r_2 - r_1) \sin^2(t - \theta_0) \leq \text{function of } \lambda. \quad (5.7)$$

Note that the problem is invariant with respect to the group of transformations

$$\begin{aligned} \boldsymbol{\alpha} &\longrightarrow \mathbf{O}_a \boldsymbol{\alpha}, & \boldsymbol{\alpha}_0 &\longrightarrow \mathbf{O}_a \boldsymbol{\alpha}_0, & \boldsymbol{\eta} &\longrightarrow \mathbf{O}_b \boldsymbol{\eta}, \\ \theta &\longrightarrow \theta + a, & \theta_0 &\longrightarrow \theta_0 + a. \end{aligned} \quad (5.8)$$

The parameters that are invariant are the noncentrality parameter λ^2 and the difference in angles $\theta - \theta_0$. We shall consider testing $H_0 : \theta = \theta_0$ for each fixed λ . We want to separate the effect of the testing procedure from the effect of the noncentrality parameter.

6 Definition of admissibility of tests

Consider a family of densities $f(\mathbf{y}|\omega)$ defined over a sample space \mathcal{Y} and a parameter space Ω . The parameter space is partitioned into two disjoint sets Ω_0 representing the null hypothesis and Ω_1 representing the alternative. A set \mathcal{A} in the sample space represents the acceptance of the null hypothesis.

Definition 6.1. A test A is *as good* as B if

$$\Pr(\mathcal{A}|\omega) \geq \Pr(\mathcal{B}|\omega), \quad \omega \in \Omega_0, \quad (6.1)$$

$$\Pr(\mathcal{A}|\omega) \leq \Pr(\mathcal{B}|\omega), \quad \omega \in \Omega_1. \quad (6.2)$$

Definition 6.2. A is *better than* B if the equations above hold with strict inequality for at least one ω .

Definition 6.3. A is *admissible* if there is no B better than A.

See, for example, Anderson (2003, Def. 5.6.3) or Lehmann (1986, Sect. 1.8).

Admissibility of a likelihood ratio test asserts that

$$\Pr\{\text{Accept } H_0|H_0, \text{LR}\} \geq \Pr\{\text{Accept } H_0|H_0, \text{competing test}\}, \quad (6.3)$$

$$\Pr\{\text{Accept } H_0|H_1, \text{LR}, \lambda\} \leq \Pr\{\text{Accept } H_0|H_1, \text{competing test}, \lambda\}. \quad (6.4)$$

In the terms of economics the LR test is Pareto-optimal.

The inequality (6.3) says that LR is at least as good as the competitor with respect to significance level, that is, probability of acceptance of the null hypothesis, at each parameter point ω in the null hypothesis. The inequality (6.4) says that LR is at least as good as the competitor with respect to power. Note that the comparison of the two tests is made for each parameter point ω . In the testing problem considered in Sections 6 and 7, $\mathbf{Q} = \lambda\boldsymbol{\eta}\boldsymbol{\alpha}' + \mathbf{W}$, the invariant parameters are essentially the noncentrality parameter λ and the difference between the null hypothesis angle θ_0 and the model value of θ . Thus $\omega = (\lambda, \theta - \theta_0)$. The model is invariant with respect to transformations (5.8).

7 Function of \mathbf{G}

First we show that a function of \mathbf{Q} that is invariant with respect to transformations (5.8) is a function of $\mathbf{Q}'\mathbf{Q} = \mathbf{G}$.

Lemma 7.1. *A function of \mathbf{Q} that is invariant with respect to*

$$\mathbf{Q} \rightarrow \mathbf{O}_a \mathbf{Q}, \quad \mathbf{Q} \rightarrow \mathbf{Q} \mathbf{O}_b, \quad (7.1)$$

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

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

$$\mathbf{Q}'_1 \mathbf{Q}_1 = \mathbf{Q}'_2 \mathbf{Q}_2, \quad (7.2)$$

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

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

8 Density of \mathbf{G}

The matrix $\mathbf{G} = \mathbf{Q}'\mathbf{Q}$ has the noncentral Wishart distribution with K degrees of freedom, covariance $\sigma^2 \mathbf{I}_2$, and noncentrality matrix

$$(\lambda \boldsymbol{\eta} \boldsymbol{\alpha}')' (\lambda \boldsymbol{\eta} \boldsymbol{\alpha}') = \lambda^2 \boldsymbol{\alpha} \boldsymbol{\alpha}'. \quad (8.1)$$

See Anderson and Girshick (1944). The density or likelihood of \mathbf{G} is

$$\frac{e^{-\frac{1}{2}\lambda^2 - \frac{1}{2}\text{tr} \mathbf{G}} |\mathbf{G}|^{\frac{1}{2}(K-3)}}{2^{\frac{1}{2}K+1} \pi^{\frac{1}{2}} \Gamma\left[\frac{1}{2}(K-1)\right]} (\lambda^2 \boldsymbol{\alpha}' \mathbf{G} \boldsymbol{\alpha})^{-(K-2)/4} \mathbf{I}_{\frac{1}{2}(K-2)}\left(\lambda \sqrt{\boldsymbol{\alpha}' \mathbf{G} \boldsymbol{\alpha}}\right), \quad (8.2)$$

where

$$\mathbf{I}_{\frac{1}{2}(K-2)}(z) = \left(\frac{1}{2}z\right)^{\frac{1}{2}(K-2)} \sum_{j=0}^{\infty} \left(\frac{z^2}{4}\right)^j \frac{1}{j! \Gamma\left(\frac{1}{2}K + j\right)} \quad (8.3)$$

is the modified Bessel function of order $(K-2)/2$ (Abramowitz and Stegun, 1972, (9.6.10) on p. 375); see also Appendix B. The first factor in (8.2) is a constant times the central

Wishart density.

Transform \mathbf{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 ($-\pi \leq t \leq \pi$) is

$$\frac{(r_2 - r_1)e^{-\frac{1}{2}\lambda^2 - \frac{1}{2}(r_1+r_2)}(r_1r_2)^{\frac{1}{2}(K-3)}}{2^{\frac{1}{2}K+1}\pi^{\frac{1}{2}}\Gamma\left[\frac{1}{2}(K-1)\right]} \mathbf{I}_{\frac{1}{2}(K-2)}^*(\lambda^2c^2), \quad (8.4)$$

where

$$\begin{aligned} c^2 &= \boldsymbol{\alpha}'\mathbf{O}_t\mathbf{R}\mathbf{O}_t'\boldsymbol{\alpha} \\ &= \boldsymbol{\alpha}'_{\theta-t}\mathbf{R}\boldsymbol{\alpha}_{\theta-t} \\ &= r_1 \sin^2(t - \theta) + r_2 \cos^2(t - \theta) \\ &= r_2 - (r_2 - r_1) \sin^2(t - \theta), \end{aligned} \quad (8.5)$$

$$\begin{aligned} \mathbf{I}_{\frac{1}{2}(K-2)}^*(\lambda^2c^2) &= \left(\frac{\lambda c}{2}\right)^{-\frac{1}{2}(K-2)} \mathbf{I}_{\frac{1}{2}(K-2)}(\lambda c) \\ &= \sum_{j=0}^{\infty} \left(\frac{\lambda^2c^2}{4}\right)^j \frac{1}{j!\Gamma\left(\frac{1}{2}K+j\right)}. \end{aligned} \quad (8.6)$$

Let

$$n(r_1, r_2) = \frac{(r_2 - r_1)(r_1r_2)^{\frac{1}{2}(K-3)}e^{-(r_1+r_2)/2}}{2^{\frac{1}{2}K+1}\pi^{\frac{1}{2}}\Gamma\left[\frac{1}{2}(K-1)\right]}. \quad (8.7)$$

Then the density of r_1, r_2 , and t is

$$h(r_1, r_2, t|\theta, \lambda) = n(r_1, r_2)e^{-\frac{1}{2}\lambda^2} \mathbf{I}_{\frac{1}{2}(K-2)}^*(\lambda^2c^2). \quad (8.8)$$

9 Likelihood ratio criterion in terms of \mathbf{G}

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

$$H_0 : \theta = \theta_0 \quad (9.1)$$

is

$$\max_{H_0} \text{Lhd} = n(r_1, r_2)e^{-\lambda^2/2} \mathbf{I}_{\frac{1}{2}(K-2)}^*(\lambda^2c_0^2), \quad (9.2)$$

where

$$c_0^2 = r_1 \sin^2(t - \theta_0) + r_2 \cos^2(t - \theta_0) = r_2 - (r_2 - r_1) \sin^2(t - \theta_0). \quad (9.3)$$

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

$$H_1 : -\pi \leq \theta \leq \pi \quad (9.4)$$

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

$$\max_{H_1} \text{Lhd} = n(r_1, r_2) e^{-\lambda^2/2} I_{\frac{1}{2}(K-2)}^*(\lambda^2 r_2). \quad (9.5)$$

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

$$\begin{aligned} \text{LRC} &= \frac{\max_{H_0} \text{Lhd}}{\max_{H_1} \text{Lhd}} = \frac{I_{\frac{1}{2}(K-2)}^*(\lambda^2 c_0^2)}{I_{\frac{1}{2}(K-2)}^*(\lambda^2 r_2)} \\ &= \frac{I_{\frac{1}{2}(K-2)}^* \left\{ \lambda^2 [r_2 - (r_2 - r_1) \sin^2(t - \theta_0)] \right\}}{I_{\frac{1}{2}(K-2)}^*(\lambda^2 r_2)}. \end{aligned} \quad (9.6)$$

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

$$(r_2 - r_1) \sin^2(t - \theta_0) \leq \text{function of } r_1, r_2, \text{ and } \lambda. \quad (9.7)$$

Note that the likelihood ratio criterion (the left side of (9.7)) 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 β is $\hat{\beta} = \beta_{\hat{\theta}}$.

The likelihood ratio criterion when λ is considered as a parameter could be derived from the model (4.10); that is equivalent to the hypothesis $\beta = \beta_{\theta_0}$, when $\beta' \beta = 1$. The likelihood of (4.10) is maximized with respect to ν for fixed θ at $\hat{\nu} = \mathbf{Q}\alpha$ yielding a

maximized likelihood of

$$\begin{aligned} \frac{1}{(2\pi)^K} e^{-\frac{1}{2} \text{tr} \mathbf{Q}' \mathbf{Q} + \frac{1}{2} \boldsymbol{\alpha}' \mathbf{Q}' \mathbf{Q} \boldsymbol{\alpha}} &= \frac{1}{(2\pi)^K} e^{-\frac{1}{2} (\text{tr} \mathbf{G} - \boldsymbol{\alpha}' \mathbf{G} \boldsymbol{\alpha})} \\ &= \frac{1}{(2\pi)^K} e^{-\frac{1}{2} \text{tr} \mathbf{R} + \frac{1}{2} c^2}. \end{aligned} \quad (9.8)$$

Under the null hypothesis c^2 is

$$c_0^2 = r_2 - (r_2 - r_1) \sin^2(t - \theta_0). \quad (9.9)$$

Under the alternative H_1 the maximum of the likelihood (9.8) occurs at $\theta = 0$ and $c^2 = r_2$.

Then the likelihood ratio criterion for testing H_0 vs H_1 is

$$e^{-\frac{1}{2} (r_2 - r_1) \sin^2(t - \theta_0)}. \quad (9.10)$$

However, to carry out the admissibility argument requires explicit treatment for each value of λ . See Anderson and Kunitomo (2009).

10 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 $a = \textit{accept } H_0$ and $a = \textit{reject } H_0$. The *risk of a test* is the expected loss

$$R(\theta, \lambda, d) = \int_0^\infty \int_0^{r_2} \int_{-\pi}^\pi L[\theta, d(r_1, r_2, t)] h(r_1, r_2, t | \theta, \lambda) dt dr_1 dr_2 \quad (10.1)$$

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

$$R^* [P(\cdot), \lambda, d] = \int_{-\pi}^{\pi} R(\theta, \lambda, d) dP(\theta). \quad (10.2)$$

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

$$\begin{aligned} R^* [P(\cdot), \lambda, d] &= \Pr\{\theta = \theta_0\} \Pr\{\text{reject } H_0 | \theta_0, \lambda\} \\ &\quad + [1 - \Pr\{\theta = \theta_0\}] \int_{-\pi}^{\pi} \Pr\{\text{accept } H_0 | \theta, \lambda\} p(\theta) d\theta \\ &= \Pr\{\theta = \theta_0\} \int_R h(r_1, r_2, t | \theta_0, \lambda) dr_1 dr_2 dt \\ &\quad + [1 - \Pr\{\theta = \theta_0\}] \int_A \bar{h}(r_1, r_2, t | \lambda) dr_1 dr_2 dt, \end{aligned} \quad (10.3)$$

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

$$\bar{h}(r_1, r_2, t | \lambda) = \int_{-\pi}^{\pi} h(r_1, r_2, t | \theta, \lambda) p(\theta) d\theta \quad (10.4)$$

is a density. The average risk can be written as

$$\begin{aligned} R^* [P(\cdot), \lambda, d] &= \Pr\{\theta = \theta_0\} + \int_A \{[1 - \Pr\{\theta = \theta_0\}] \bar{h}(r_1, r_2, t | \lambda) \\ &\quad - \Pr\{\theta = \theta_0\} h(r_1, r_2, t | \theta_0, \lambda)\} dr_1 dr_2 dt. \end{aligned} \quad (10.5)$$

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

$$[1 - \Pr\{\theta = \theta_0\}] \bar{h}(r_1, r_2, t | \lambda) - \Pr\{\theta = \theta_0\} h(r_1, r_2, t | \theta_0, \lambda) \leq 0, \quad (10.6)$$

that is, the largest set A for which

$$A : \frac{h(r_1, r_2, t | \theta_0, \lambda)}{\bar{h}(r_1, r_2, t | \lambda)} \geq \frac{1 - \Pr\{\theta = \theta_0\}}{\Pr\{\theta = \theta_0\}}. \quad (10.7)$$

Theorem 10.1. *For each λ the Bayes test of $H_0 : \theta = \theta_0$ vs $H_1 : \theta \neq \theta_0$ 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*

$p(\theta)$ ($\theta \neq \theta_0$), has the acceptance set (10.7).

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

$$p(\theta) = \frac{1}{2\pi}, \quad -\pi \leq \theta \leq \pi, \quad (10.8)$$

the denominator of the left-hand side of (10.7) is

$$\begin{aligned} & \bar{h}(r_1, r_2, \lambda) \\ &= n(r_1, r_2) e^{-\frac{1}{2}\lambda^2} \frac{1}{2\pi} \int_{-\pi}^{\pi} \sum_{j=0}^{\infty} \left(\frac{\lambda^2}{4}\right)^j \frac{1}{j! \Gamma\left[\frac{1}{2}K + j\right]} [r_2 - (r_2 - r_1) \sin^2(t - \theta)]^j d\theta \\ &= n(r_1, r_2) e^{-\frac{1}{2}\lambda^2} \sum_{j=0}^{\infty} \left(\frac{\lambda^2}{4}\right)^j \frac{1}{j! \Gamma\left[\frac{1}{2}K + j\right]} \frac{1}{2\pi} \int_{t-\pi}^{t+\pi} [r_2 - (r_2 - r_1) \sin^2 x]^j dx \\ &= n(r_1, r_2) e^{-\frac{1}{2}\lambda^2} f_K(r_1, r_2, \lambda), \end{aligned} \quad (10.9)$$

where

$$f_K(r_1, r_2, \lambda) = \sum_{j=0}^{\infty} \left(\frac{\lambda^2}{4}\right)^j \frac{1}{j! \Gamma\left[\frac{1}{2}K + j\right]} \frac{1}{2\pi} \int_{-\pi}^{\pi} [r_2 - (r_2 - r_1) \sin^2 x]^j dx. \quad (10.10)$$

The integrand in (10.10) 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 (10.7) is

$$\frac{h(r_1, r_2, t|\theta_0, \lambda)}{\bar{h}(r_1, r_2, t|\lambda)} = \frac{I_{\frac{1}{2}(K-2)}^*(\lambda^2 c^2)}{f_K(r_1, r_2, \lambda)}. \quad (10.11)$$

The numerator of the LRC and the left-hand side of (10.7) are the same.

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 10.2. *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 $1/(2\pi)$.*

11 Admissibility of invariant tests

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 11.1. *The Bayes test with acceptance region (10.7) 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 (10.7), resulting in the average risk $R^*[p(\cdot), \lambda, d_B]$. If this test is not admissible, then there is a test d^* that is better than d_B , that is,

$$R^*[p(\cdot), \lambda, d^*] \leq R^*[p(\cdot), \lambda, d_B] \quad (11.1)$$

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 . \square

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.

12 Admissibility over all tests

12.1 General theorem

Now we consider admissibility with respect to all tests. We assert 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 invariant

tests is admissible within the class of invariant tests, the transformed family 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 of transformations defining invariance is finite or compact. See Zaman (1996, Sect. 7.9) or Lehmann (1986, Thm. 7 of Chap. 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

$$Q = \lambda \boldsymbol{\eta} \boldsymbol{\alpha}' + W \quad (12.1)$$

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

12.2 An example

To illustrate the Hunt–Stein theory, consider the model in which θ can take on a finite number of values, say

$$\theta = 0, \frac{1}{N}2\pi, \frac{2}{N}2\pi, \dots, \frac{N-1}{N}2\pi. \quad (12.2)$$

Note that $\boldsymbol{\alpha}' = (-\sin \theta, \cos \theta)$. Consider the group of transformations

$$\theta \longrightarrow \theta + \frac{j}{N}2\pi, \quad t \longrightarrow t + \frac{j}{N}2\pi, \quad j = 0, 1, \dots, N-1. \quad (12.3)$$

Let these values of θ be labeled as $\theta_0^*, \theta_1^*, \dots, \theta_{N-1}^*$. Each of them corresponds to a null hypothesis. 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

$$A_k^*(t - \theta_k^*, r_1, r_2) = A_j^*(t - \theta_j^*, r_1, r_2) \quad (12.4)$$

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

The LR test of the hypothesis $\theta = \theta_i^*$ against the alternative $\theta = \theta_j^*$ for some $j =$

$0, 1, \dots, N - 1$ is the Bayes solution for the hypothesis $\theta = \theta_i^*$ for prior probabilities

$$\Pr\{\theta = \theta_j^*\} = \frac{1}{N}, \quad j = 0, 1, \dots, N - 1. \quad (12.5)$$

Suppose the set of tests are not necessarily invariant; that is, (12.4) 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

$$\frac{1}{N} \sum_{k=0}^{N-1} A_k^*(t - \theta_k^* + \theta_i^*, r_1, r_2). \quad (12.6)$$

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

Lemma 12.1. *If a test with an invariant family of acceptance regions A_0, A_1, \dots, A_{N-1} is admissible in the set of invariant tests, it is admissible in the set of all tests.*

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

13 Comments

13.1 Invariance with respect to linear transformations of exogenous variables

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

Similarly, the transformation does not affect the equation $\mathbf{\Pi}\boldsymbol{\beta} = \mathbf{0}$, in particular the null hypothesis $\mathbf{\Pi}\boldsymbol{\beta}_0 = \mathbf{0}$. 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 $\mathbf{G} = \mathbf{P}'\mathbf{A}\mathbf{P}$.

13.2 Invariance with respect to orthogonal transformations of endogenous variables

When $\Omega = I$ is assumed, an orthogonal transformation of the disturbance $V \rightarrow VO$ and a corresponding transformation of β , $\beta \rightarrow O'\beta$ and of the null hypothesis $\beta_0 \rightarrow O'\beta_0$ do not affect the equations, $\beta = \beta_0$ and $\beta'\beta = 1$. In the G -space this transformation is a rotation of coordinates.

13.3 Conclusions

Theorem 12.1 states that the LR test of $H_0 : \alpha = \alpha_0$ vs $H_1 : \mathcal{E}Q = \lambda\eta\alpha'$ ($\eta'\eta = 1$, $\alpha'\alpha = 1$) is admissible in the class of all (randomized) tests. This implies that given any test of H_0 vs H_1 in the model $\mathcal{E}Q = \lambda\eta\alpha'$ there is an LR test that is better than that test. Thus the statistician need only consider (randomized) LR tests.

Note that the comparison of a given test with respect to α holds for each specified λ . The theorem does not indicate how to establish the significance level for a specific λ . The analysis assumes that the system is in equilibrium; that is, that (4.8) holds for all T .

14 A more general model

Instead of (2.9) consider (2.4) with the hypothesis $H_0 : \beta_1 = \beta_0$, where β_1 satisfies (2.8). Let

$$\mathbf{Z}_{2.1} = \mathbf{Z}_2 - \mathbf{Z}_1\mathbf{A}_{11}^{-1}\mathbf{A}_{12}, \quad (14.1)$$

where \mathbf{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

$$\mathbf{Y}_1 = \mathbf{Z}_1(\mathbf{\Pi}_{11} + \mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{\Pi}_{21}) + \mathbf{Z}_{2.1}\mathbf{\Pi}_{21} + \mathbf{V}_1. \quad (14.2)$$

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

$$\mathbf{A}_{22.1} = \mathbf{Z}'_{2.1}\mathbf{Z}_{2.1} = \mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12}, \quad (14.3)$$

and they are independent. The developments above proceed with \mathbf{Z} replaced by $\mathbf{Z}_{2,1}$, \mathbf{Y} by \mathbf{Y}_1 , etc.

References

- Abramowitz, Milton and Irene A. Stegun. 1972. *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. Dover Books on Mathematics. Dover Publications. URL <http://books.google.com/books?id=IHhkzgAACAAJ>.
- Anderson, T. W. 1976. “Estimation of linear functional relationships: approximate distributions and connections with simultaneous equations in econometrics.” *J. Roy Statist. Soc. Ser. B* 38 (1):1–36. With a discussion by P. Sprent, J. B. Copas, M. S. Bartlett, N. N. Chan, Mary E. Solari, David F. Hendry, A. M. Walker, E. B. Andersen, John Bibby, R. W. Farebrother, Ejnar Lyttkens, A. E. Maxwell, R. J. O’Brien, P. C. B. Phillips, D. A. Williams, E. J. Williams and W. M. Patefield and a reply by the author.
- . 1984. “The 1982 Wald Memorial Lectures: Estimating linear statistical relationships.” *Ann. Statist.* 12 (1):1–45. URL [10.1214/aos/1176346390](https://doi.org/10.1214/aos/1176346390).
- . 2003. *An Introduction to Multivariate Statistical Analysis*. Wiley Series in Probability and Statistics. Hoboken, NJ: Wiley-Interscience [John Wiley & Sons], third ed.
- Anderson, T. W. and M. A. Girshick. 1944. “Some extensions of the Wishart distribution.” *Ann. Math. Statistics* 15:345–357. (Correction: 1964, **35**, 923).
- Anderson, T. W. and Naoto Kunitomo. 2007. “On likelihood ratio tests of structural coefficients: Anderson–Rubin (1949) revisited.” URL <http://www.cirje.e.u-tokyo.ac.jp/research/dp/2007/2007cf499.pdf>. Discussion Paper CIRJE-F-499.
- . 2009. “Testing a hypothesis about a structural coefficient in a simultaneous equation model for known covariance matrix (In honor of Raul P. Mentz).” *Estadística* 61:7–16.
- Anderson, T. W. and Herman Rubin. 1949. “Estimation of the parameters of a single equation in a complete system of stochastic equations.” *Ann. Math. Statistics* 20:46–63.

- Anderson, T. W., Charles Stein, and Asad Zaman. 1985. “Best invariant estimation of a direction parameter.” *Ann. Statist.* 13 (2):526–533. URL 10.1214/aos/1176349536.
- Creasy, Monica A. 1956. “Confidence limits for the gradient in the linear functional relationship.” *J. Roy. Statist. Soc. Ser. B.* 18:65–69.
- Haavelmo, Trygve. 1944. “The probability approach in econometrics.” *Econometrica* 12 Supplement:118.
- Lehmann, Erich L. 1986. *Testing Statistical Hypotheses*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. New York: John Wiley & Sons Inc., second ed.
- Moreira, Marcelo J. 2003. “A conditional likelihood ratio test for structural models.” *Econometrica* 71 (4):1027–1048. URL 10.1111/1468-0262.00438.
- Zaman, Asad. 1996. *Statistical Foundations for Econometric Techniques*. Economic Theory, Econometrics, and Mathematical Economics. San Diego, CA: Academic Press Inc.

A Jacobian

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

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

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

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

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

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

where \mathbf{Q}_2 and \mathbf{W}_2 are $(K-1) \times G$, \mathbf{Q}_1 and \mathbf{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 \mathbf{W} are independently normally distributed with means $\mathbf{0}$ and covariance matrix \mathbf{I}_G . Then $\mathbf{Q}'_2\mathbf{Q}_2 = \mathbf{G}_2$ has a (central) Wishart distribution $W(\mathbf{I}_G, K-1)$ with density

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

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

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

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

The joint density of $\mathbf{G} = \mathbf{q}_1\mathbf{q}'_1 + \mathbf{G}_2$ and \mathbf{q}_1 is

$$\begin{aligned} & \frac{|\mathbf{G} - \mathbf{q}_1\mathbf{q}'_1|^{\frac{1}{2}(K-G-2)} e^{-\frac{1}{2}\text{tr}\mathbf{G} + \lambda q_{11} - \lambda^2/2}}{2^{\frac{1}{2}KG} \pi^{G(G+1)/4} \prod_{i=1}^G \Gamma\left[\frac{1}{2}(K-i)\right]} \\ &= \frac{|\mathbf{G}|^{\frac{1}{2}(K-G-2)} (1 - \mathbf{q}'_1\mathbf{G}^{-1}\mathbf{q}_1)^{\frac{1}{2}(K-G-2)} e^{-\frac{1}{2}\text{tr}\mathbf{G} + \lambda q_{11} - \lambda^2/2}}{2^{\frac{1}{2}KG} \pi^{G(G+1)/4} \prod_{i=1}^G \Gamma\left[\frac{1}{2}(K-i)\right]}. \end{aligned} \quad (\text{B.4})$$

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

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

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

$(\mathbf{1}, \mathbf{0})'$, is

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

where

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

C Neyman–Pearson fundamental lemma

Let $p_0(x)$ and $p_1(x)$ be two densities defined for x in some (finite) Euclidean space. Consider testing the null hypothesis that the density of x is $p_0(x)$ against the alternative that the density is $p_1(x)$. The significance level of the test is defined as

$$\alpha = \Pr\{\text{rejecting } H_0 | \text{sampling } p_0(x)\}.$$

The power of the test is defined as

$$\Pr\{\text{rejecting } H_0 | \text{sampling } p_1(x)\}.$$

The most powerful test of H_0 given the significance level α is defined by the rejection region

$$\frac{p_0(x)}{p_1(x)} \leq k$$

for the minimum k . See Problem 6.4 of Anderson (2003), for example.