Admissible Significance Tests in Simultaneous Equation Models*

T. W. Anderson†
Stanford University

January 18, 2012

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

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 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, $\boldsymbol{\beta}$, is a specified vector, say, $\boldsymbol{\beta}_0$;

*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; `twa@stanford.edu`

the alternative hypothesis, say H_2, is that $\boldsymbol{\beta}$ 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 is assumed to be proportional to $\boldsymbol{I}$ and the number of endogenous variables in the single equation is two. In this case it is convenient to use polar coordinates for the vector $\boldsymbol{\beta}$.

When the disturbance covariance matrix is known, a sufficient statistic is the sample regression of the dependent variables on the independent variables, say $\boldsymbol{P}$, and the sample covariance of the independent variables, say $\boldsymbol{A}$. The likelihood ratio criterion is a function of $\boldsymbol{P}'\boldsymbol{A}\boldsymbol{P}$, the distribution of which depends on a noncentrality parameter, say λ. It is shown that if λ is known, the likelihood ratio criterion 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. 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 consist of a $T \times G$ matrix of endogenous or dependent or nonstochastic variables $\boldsymbol{Y}$ and a $T \times K$ matrix of exogenous or independent variables $\boldsymbol{Z}$ $(G < K)$. 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

$$\boldsymbol{H} = (\boldsymbol{Y} - \boldsymbol{ZP})'(\boldsymbol{Y} - \boldsymbol{ZP}) = \boldsymbol{Y}'\boldsymbol{Y} - \boldsymbol{P}'\boldsymbol{AP} \tag{2.3}$$

and $\boldsymbol{A} = \boldsymbol{Z}'\boldsymbol{Z}$. The matrices $\boldsymbol{P}$ and $\boldsymbol{H}$ 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 $\boldsymbol{Y}_1$, 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

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

The second submatrix of (2.7),

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

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

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

3 Invariance and normalization

Exogenous variables The model (2.1) 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

$$\boldsymbol{G}^+ = \boldsymbol{P}^{+'}\boldsymbol{A}^+\boldsymbol{P}^+ = \boldsymbol{P}'\boldsymbol{A}\boldsymbol{P} = \boldsymbol{G}, \qquad \boldsymbol{H}^+ = \boldsymbol{Y}'\boldsymbol{Y} - \boldsymbol{P}^{+'}\boldsymbol{A}^+\boldsymbol{P}^+ = \boldsymbol{H}. \tag{3.3}$$

Endogenous variables 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), $\boldsymbol{\Pi}\boldsymbol{\beta} = \boldsymbol{0}$, 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

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

and

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

Now we consider the model (2.1) and $\boldsymbol{\Pi}\boldsymbol{\beta} = \boldsymbol{0}$ 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 (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

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

$$\begin{gathered} \boldsymbol{P}^* = \boldsymbol{P}\boldsymbol{O}, \qquad \boldsymbol{H}^* = \boldsymbol{O}'\boldsymbol{H}\boldsymbol{O}, \\ \boldsymbol{P}^{*'}\boldsymbol{A}\boldsymbol{P}^* = \boldsymbol{O}'\boldsymbol{P}'\boldsymbol{A}\boldsymbol{P}\boldsymbol{O}. \end{gathered} \tag{3.11}$$

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 $\mathbf{\Omega} = \boldsymbol{I}_2$ 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

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

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

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

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

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

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

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

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

Note that

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

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

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

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

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

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

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 $\theta = \theta_0$.

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

The density of $\boldsymbol{Q}$ is

$$\begin{aligned} \frac{1}{(2\pi)^K} e^{-\frac{1}{2}\operatorname{tr}\boldsymbol{W}'\boldsymbol{W}} &= \frac{1}{(2\pi)^K} e^{-\frac{1}{2}\operatorname{tr}(\boldsymbol{Q}-\lambda\boldsymbol{\eta}\boldsymbol{\alpha}')'(\boldsymbol{Q}-\lambda\boldsymbol{\eta}\boldsymbol{\alpha}')} \\ &= \frac{1}{(2\pi)^K} e^{-\frac{1}{2}\operatorname{tr}\boldsymbol{Q}'\boldsymbol{Q}-\frac{1}{2}\lambda^2+\lambda\boldsymbol{\eta}'\boldsymbol{Q}\boldsymbol{\alpha}} \end{aligned} \tag{4.10}$$

since $\lambda^2 \operatorname{tr}(\boldsymbol{\alpha}\boldsymbol{\eta}'\boldsymbol{\eta}\boldsymbol{\alpha}') = \lambda^2$ and $\lambda \operatorname{tr}(\boldsymbol{\alpha}\boldsymbol{\eta}'\boldsymbol{Q}) = \lambda\boldsymbol{\eta}'\boldsymbol{Q}\boldsymbol{\alpha}$. We shall find the *best invariant test* of $\theta = \theta_0$. Invariance is with respect to the group of transformations

$$\begin{gathered} \boldsymbol{\alpha} \longrightarrow \boldsymbol{O}_a\boldsymbol{\alpha}, \quad \boldsymbol{\alpha}_0 \longrightarrow \boldsymbol{O}_a\boldsymbol{\alpha}_0, \quad \boldsymbol{\eta} \longrightarrow \boldsymbol{O}_b\boldsymbol{\eta}, \\ \theta \longrightarrow \theta + a, \qquad \theta_0 \longrightarrow \theta_0 + a. \end{gathered} \tag{4.11}$$

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.

5 Admissibility of 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 $\mathcal{A}$ in the sample space represents the acceptance of the null hypothesis.

Definition 5.1. A test A is as good as B if

$$P(\mathcal{A}|\omega) \geq P(\mathcal{B}|\omega), \quad \omega \in \boldsymbol{\Omega}_0, \tag{5.1}$$

$$P(\mathcal{A}|\omega) \leq P(\mathcal{B}|\omega), \quad \omega \in \boldsymbol{\Omega}_1. \tag{5.2}$$

Definition 5.2. A is *better than* B if the equations above hold with strict inequality for at least one ω.

Definition 5.3. 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).

The inequality (5.1) says that test A is better than B 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 (5.2) says that test A is better than test B 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, $\boldsymbol{Q} = \lambda\boldsymbol{\eta}\boldsymbol{\alpha}' + \boldsymbol{W}$, 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 (4.11).

6 The likelihood ratio test of $\boldsymbol{\beta} = \boldsymbol{\beta}_0$

Let L be the likelihood function stated as the density of $\boldsymbol{Q}$ conditional on λ. Then

$$\log L = -\frac{1}{2} \operatorname{tr} \boldsymbol{Q}'\boldsymbol{Q} - \frac{1}{2}\lambda^2 + \lambda\boldsymbol{\eta}'\boldsymbol{Q}\boldsymbol{\alpha} - K \log 2\pi, \tag{6.1}$$

where $\boldsymbol{\eta}'\boldsymbol{\eta} = 1$ and $\boldsymbol{\alpha}'\boldsymbol{\alpha} = 1$. The hypothesis $\boldsymbol{\beta} = \boldsymbol{\beta}_0 = (\cos\theta_0, \sin\theta_0)'$ is equivalent to the hypothesis $\boldsymbol{\alpha} = \boldsymbol{\alpha}_0 = (-\sin\theta_0, \cos\theta_0)'$ for the normalization $\boldsymbol{\alpha}'\boldsymbol{\alpha} = 1 = \boldsymbol{\beta}'\boldsymbol{\beta}$. The derivatives of $\log L$ with respect to the vectors $\boldsymbol{\eta}$ and $\boldsymbol{\alpha}$ using Lagrange multipliers are

$$\begin{aligned} \frac{\partial}{\partial\boldsymbol{\eta}'}\left[\log L + u\left(\boldsymbol{\eta}'\boldsymbol{\eta} - 1\right)\right] &= \lambda\boldsymbol{Q}\boldsymbol{\alpha} + 2u\boldsymbol{\eta} = \mathbf{0}, \\ \frac{\partial}{\partial\boldsymbol{\alpha}}\left[\log L + v\left(\boldsymbol{\alpha}'\boldsymbol{\alpha} - 1\right)\right] &= \lambda\boldsymbol{\eta}'\boldsymbol{Q} + 2v\boldsymbol{\alpha}' = \mathbf{0}. \end{aligned} \tag{6.2}$$

These equations imply

$$\begin{aligned}\lambda\boldsymbol{\eta}'\boldsymbol{Q}\boldsymbol{\alpha} + 2u\boldsymbol{\eta}'\boldsymbol{\eta} &= \lambda\boldsymbol{\eta}'\boldsymbol{Q}\boldsymbol{\alpha} + 2u \ = 0,\\ \lambda\boldsymbol{\eta}'\boldsymbol{Q}\boldsymbol{\alpha} + 2v\boldsymbol{\alpha}'\boldsymbol{\alpha} &= \lambda\boldsymbol{\eta}'\boldsymbol{Q}\boldsymbol{\alpha} + 2v = 0.\end{aligned} \tag{6.3}$$

These imply $u = v$. Then

$$\mathbf{0} = \lambda\boldsymbol{Q}'\boldsymbol{Q}\boldsymbol{\alpha} + 2u\boldsymbol{Q}'\boldsymbol{\eta} = \lambda\boldsymbol{Q}'\boldsymbol{Q}\boldsymbol{\alpha} - \frac{4u^2}{\lambda}\boldsymbol{\alpha}. \tag{6.4}$$

Since $\boldsymbol{Q}'\boldsymbol{Q} = \boldsymbol{P}'\boldsymbol{A}\boldsymbol{P} = \boldsymbol{G}$,

$$\boldsymbol{G}\boldsymbol{\alpha} = \frac{4u^2}{\lambda^2}\boldsymbol{\alpha}. \tag{6.5}$$

Thus $\boldsymbol{\alpha}$ is an eigenvector of $\boldsymbol{G}$ corresponding to the eigenvalue $\frac{4u^2}{\lambda^2}$. Similarly,

$$\mathbf{0} = \lambda\boldsymbol{\eta}'\boldsymbol{Q}\boldsymbol{Q}' + 2v\boldsymbol{\alpha}'\boldsymbol{Q}' = \lambda\boldsymbol{\eta}'\boldsymbol{Q}\boldsymbol{Q}' - \frac{4u^2}{\lambda}\boldsymbol{\eta}'. \tag{6.6}$$

Thus $\boldsymbol{\eta}$ is a left eigenvector of $\boldsymbol{Q}\boldsymbol{Q}'$ corresponding to the eigenvalue $\frac{4u^2}{\lambda^2}$. The larger eigenvalue, say r_2, maximizes $\lambda\boldsymbol{\eta}'\boldsymbol{Q}\boldsymbol{\alpha}$ and hence the likelihood L.

When $\boldsymbol{\alpha} = \boldsymbol{\alpha}_0$, the derivative with respect to $\boldsymbol{\eta}$ is

$$\lambda\boldsymbol{Q}\boldsymbol{\alpha}_0 + 2u^*\boldsymbol{\eta} = \mathbf{0} \tag{6.7}$$

and

$$\boldsymbol{\alpha}_0'\boldsymbol{Q}'\boldsymbol{Q}\boldsymbol{\alpha}_0 = \frac{4u^{*2}}{\lambda^2}\boldsymbol{\eta}'\boldsymbol{\eta} = \frac{4u^{*2}}{\lambda^2}. \tag{6.8}$$

The maximized likelihoods are

$$\begin{aligned}\max_{H_1} L &= \frac{1}{(2\pi)^K}e^{-\frac{1}{2}\operatorname{tr}\boldsymbol{Q}'\boldsymbol{Q}-\frac{1}{2}\lambda^2-2u}\\ &= \frac{1}{(2\pi)^K}e^{-\frac{1}{2}\operatorname{tr}\boldsymbol{Q}'\boldsymbol{Q}-\frac{1}{2}\lambda^2+\lambda\sqrt{r_2}}\\ &= \frac{1}{(2\pi)^K}e^{-\frac{1}{2}\operatorname{tr}\boldsymbol{G}-\frac{1}{2}\lambda^2+\lambda\sqrt{r_2}}\end{aligned} \tag{6.9}$$

$$\begin{aligned}\max_{H_0} L &= \frac{1}{(2\pi)^K}e^{-\frac{1}{2}\operatorname{tr}\boldsymbol{Q}'\boldsymbol{Q}-\frac{1}{2}\lambda^2-2u^*}\\ &= \frac{1}{(2\pi)^K}e^{-\frac{1}{2}\operatorname{tr}\boldsymbol{Q}'\boldsymbol{Q}-\frac{1}{2}\lambda^2+\lambda\sqrt{\boldsymbol{\alpha}_0'\boldsymbol{Q}'\boldsymbol{Q}\boldsymbol{\alpha}_0}}\\ &= \frac{1}{(2\pi)^K}e^{-\frac{1}{2}\operatorname{tr}\boldsymbol{G}-\frac{1}{2}\lambda^2+\lambda\sqrt{\boldsymbol{\alpha}_0'\boldsymbol{G}\boldsymbol{\alpha}_0}}.\end{aligned} \tag{6.10}$$

The LR criterion for λ known is the ratio

$$\frac{\max_{H_0} L}{\max_{H_1} L} = e^{\lambda\sqrt{\boldsymbol{\alpha}_0'\boldsymbol{G}\boldsymbol{\alpha}_0}-\lambda\sqrt{r_2}}. \tag{6.11}$$

Accept H_0 if this ratio is large.

If λ is treated as a parameter and $\max_{H_1} L$ and $\max_{H_0} L$ are maximized with respect to λ, the maxima are

$$(2\pi)^{-K}e^{-\frac{1}{2}r_1} \quad \text{and} \quad (2\pi)^{-K}e^{-\frac{1}{2}(r_1+r_2)+\frac{1}{2}\boldsymbol{\alpha}_0'\boldsymbol{G}\boldsymbol{\alpha}_0}, \tag{6.12}$$

respectively. The likelihood ratio criterion is $e^{\frac{1}{2}\boldsymbol{\alpha}_0'\boldsymbol{G}\boldsymbol{\alpha}_0-\frac{1}{2}r_2}$ (Anderson and Kunitomo, 2009).

7 Bayes test

For testing $H_0 : \theta = \theta_0$ vs $H_1 : \theta \neq \theta_0$ the loss function is given in the following table:

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

The *a priori* probability of H_0 is

$$\Pr\{\theta = \theta_0\} = \pi_0. \tag{7.1}$$

The *a priori* probability of H_1 is

$$\Pr\{\theta \neq \theta_0\} = 1 - \pi_0. \tag{7.2}$$

The prior distribution of $\boldsymbol{\eta}$ in H_0 is the uniform spherical distribution $U_K(\boldsymbol{\eta})$. In H_1, $\boldsymbol{\eta}$ and $\boldsymbol{\alpha}$ are uniformly and independently distributed according to $U_K(\boldsymbol{\eta})$ and $U_2(\boldsymbol{\alpha})$, respectively.

A test procedure is defined by an acceptance region A in the space of $\boldsymbol{Q}$ and a rejection region R. The expected loss for given λ is

$$\begin{aligned}
&\pi_0 e^{-\frac{1}{2}\lambda^2} \int_R \frac{1}{(2\pi)^K} \int e^{-\frac{1}{2}\operatorname{tr}\boldsymbol{Q}'\boldsymbol{Q}+\lambda\boldsymbol{\eta}'\boldsymbol{Q}\boldsymbol{\alpha}_0}\, dU_K(\boldsymbol{\eta})\, d\boldsymbol{Q} \\
&\qquad + (1-\pi_0)e^{-\frac{1}{2}\lambda^2} \frac{1}{(2\pi)^K} \int_A e^{-\frac{1}{2}\operatorname{tr}\boldsymbol{Q}'\boldsymbol{Q}+\lambda\boldsymbol{\eta}'\boldsymbol{Q}\boldsymbol{\alpha}}\, dU_K(\boldsymbol{\eta})\, dU_2(\boldsymbol{\alpha})\, d\boldsymbol{Q} \\
&= e^{-\frac{1}{2}\lambda^2}\left\{\pi_0 + \frac{1}{(2\pi)^K}\int_A e^{-\frac{1}{2}\operatorname{tr}\boldsymbol{Q}'\boldsymbol{Q}} \int \left[(1-\pi_0)\int e^{\lambda\boldsymbol{\eta}'\boldsymbol{Q}\boldsymbol{\alpha}}\, dU_2(\boldsymbol{\alpha}) - \pi_0 e^{\lambda\boldsymbol{\eta}'\boldsymbol{Q}\boldsymbol{\alpha}_0}\right] dU_K(\boldsymbol{\eta})\, d\boldsymbol{Q}\right\}.
\end{aligned} \tag{7.3}$$

The Bayes test is defined by the acceptance and rejection regions that minimize the expected loss. The set A that minimizes the expected loss is the largest set of $\boldsymbol{Q}$ such that

$$A : (1-\pi_0)\int e^{\lambda\boldsymbol{\eta}'\boldsymbol{Q}\boldsymbol{\alpha}}\, dU_2(\boldsymbol{\alpha})\, dU_K(\boldsymbol{\eta}) - \pi_0 \int e^{\lambda\boldsymbol{\eta}'\boldsymbol{Q}\boldsymbol{\alpha}_0}\, dU_K(\boldsymbol{\eta}) \leq 0 \tag{7.4}$$

by the Neyman–Pearson Fundamental Lemma.

Lemma 7.1.

$$\int_{\boldsymbol{\eta}'\boldsymbol{\eta}=1} f(\boldsymbol{\eta}'\boldsymbol{x})\, dU_K(\boldsymbol{\eta}) = f\left(\sqrt{\boldsymbol{x}'\boldsymbol{x}}\right). \tag{7.5}$$

Proof. The integral is

$$\int_{\boldsymbol{\eta}'\boldsymbol{\eta}=1} f(\boldsymbol{\eta}'\boldsymbol{O}_K\boldsymbol{O}_K'\boldsymbol{x})\, dU_K(\boldsymbol{\eta}) \tag{7.6}$$

for any orthogonal $\boldsymbol{O}_K$. Define $\boldsymbol{O}_K$ by $\boldsymbol{O}_K'\boldsymbol{x} = \left(\sqrt{\boldsymbol{x}'\boldsymbol{x}}, 0, \ldots, 0\right)'$ and $\boldsymbol{\mu}$ by $\boldsymbol{\mu}' = \boldsymbol{\eta}'\boldsymbol{O}_K$. The integral is

$$\int_{\boldsymbol{\mu}'\boldsymbol{\mu}=1} f\left[\left(\sqrt{\boldsymbol{x}'\boldsymbol{x}}, 0, \ldots, 0\right)\boldsymbol{\mu}\right] dU_K(\boldsymbol{\mu}) = f\left(\sqrt{\boldsymbol{x}'\boldsymbol{x}}\right) \tag{7.7}$$

since $dU_1(\boldsymbol{\mu}) = 1$. □

The minimum expected loss (7.3) is defined by the largest A for which

$$e^{-\frac{1}{2}\lambda^2}\left\{\pi_0 + \frac{1}{(2\pi)^K}\int_A e^{-\frac{1}{2}\operatorname{tr}\boldsymbol{G}}\left[(1-\pi_0)\int e^{\lambda\sqrt{\boldsymbol{\alpha}'\boldsymbol{G}\boldsymbol{\alpha}}}\, dU_2(\boldsymbol{\alpha}) - \pi_0 e^{\lambda\sqrt{\boldsymbol{\alpha}_0'\boldsymbol{G}\boldsymbol{\alpha}_0}}\right] d\boldsymbol{G}\right\} \leq 0. \tag{7.8}$$

The Bayes acceptance set A is equivalently defined by the inequality

$$A : (1-\pi_0)\int e^{\lambda\sqrt{\boldsymbol{\alpha}'\boldsymbol{Q}'\boldsymbol{Q}\boldsymbol{\alpha}}}\, dU_2(\boldsymbol{\alpha}) \leq \pi_0 e^{\lambda\sqrt{\boldsymbol{\alpha}_0'\boldsymbol{Q}'\boldsymbol{Q}\boldsymbol{\alpha}_0}}, \tag{7.9}$$

that is,

$$A : \frac{\int e^{\lambda\sqrt{\boldsymbol{\alpha}'\boldsymbol{G}\boldsymbol{\alpha}}}\, dU_2(\boldsymbol{\alpha})}{e^{\lambda\sqrt{\boldsymbol{\alpha}_0'\boldsymbol{G}\boldsymbol{\alpha}_0}}} \leq \frac{\pi_0}{1-\pi_0}. \tag{7.10}$$

The largest set A for which (7.10) holds is the Bayes acceptance set. Since $\boldsymbol{\alpha}'\boldsymbol{G}\boldsymbol{\alpha} \leq r_2\boldsymbol{\alpha}'\boldsymbol{\alpha}$, the above inequality on $\boldsymbol{G}$ is

$$A : \frac{e^{\lambda\sqrt{r_2}}}{e^{\lambda\sqrt{\boldsymbol{\alpha}_0'\boldsymbol{G}\boldsymbol{\alpha}_0}}} \leq \frac{\pi_0}{1-\pi_0}. \tag{7.11}$$

Then (7.11) agrees with (6.11). This shows that the LR Test is a Bayes test. This is the LR Criterion (LRC) for given λ.

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

$$\boldsymbol{R} = \begin{bmatrix} r_1 & 0 \\ 0 & r_2 \end{bmatrix}, \qquad \boldsymbol{O}_t = \begin{bmatrix} \cos t & -\sin t \\ \sin t & \cos t \end{bmatrix} = (\boldsymbol{\beta}_t, \boldsymbol{\alpha}_t). \tag{7.12}$$

Then

$$\begin{aligned} \boldsymbol{\alpha}'\boldsymbol{G}\boldsymbol{\alpha} &= \boldsymbol{\alpha}_{t-\theta}'\boldsymbol{R}\boldsymbol{\alpha}_{t-\theta} \\ &= r_1\sin^2(t-\theta) + r_2\cos^2(t-\theta) \\ &= r_2 - (r_2 - r_1)\sin^2(t-\theta). \end{aligned} \tag{7.13}$$

The acceptance region A for given λ is defined by

$$(r_2 - r_1)\sin^2(t-\theta_0) \leq \frac{1}{\lambda}\log\frac{\pi_0}{1-\pi_0}. \tag{7.14}$$

This is a likelihood ratio test and is a Bayes test.

Theorem 7.1. *The Bayes test with acceptance region of the LR Test is admissible in the set of invariant tests.*

Proof. Let the expected loss of the LR Test, which is the Bayes test, be $\mathcal{E}L(\theta, a)$ where the expectation is over the distribution of $\boldsymbol{Q}$ and the parameters $\boldsymbol{\eta}$ and $\boldsymbol{\alpha}$. If this test is not admissible, there is another test that is better in the sense that the expected loss is less than the expected loss for the LRT for some π_0, $\boldsymbol{\eta}$, $\boldsymbol{\alpha}$ and less than or equal for all π_0, $\boldsymbol{\eta}$, $\boldsymbol{\alpha}$. This conclusion contradicts the definition of the Bayes test. □

8 Admissibility over all tests

8.1 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 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

$$\boldsymbol{Q} = \lambda \boldsymbol{\eta} \boldsymbol{\alpha}' + \boldsymbol{W} \tag{8.1}$$

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

8.2 An example

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

$$\theta = 0, \frac{1}{N}2\pi, \frac{2}{N}2\pi, \ldots, \frac{N-1}{N}2\pi. \tag{8.2}$$

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

$$\theta \longrightarrow \theta + \frac{j}{N}2\pi, \quad t \longrightarrow t + \frac{j}{N}2\pi, \qquad j = 0, 1, \ldots, N-1. \tag{8.3}$$

Let these values of θ be labeled as $\theta_0^*, \theta_1^*, \ldots, \theta_{N-1}^*$. 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

$$A_k^*(t - \theta_k^*, r_1, r_2) = A_j^*(t - \theta_j^*, r_1, r_2) \tag{8.4}$$

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

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

$$\Pr\{\theta = \theta_j^*\} = \frac{1}{N}, \qquad j = 0, 1, \ldots, N-1. \tag{8.5}$$

Suppose the set of tests are not necessarily invariant; that is, (8.4) 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

$$\frac{1}{N} \sum_{k=0}^{N-1} A_k^*(t - \theta_k^* + \theta_i^*, r_1, r_2). \tag{8.6}$$

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

Lemma 8.1. *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. □

9 Comments

9.1 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 $\boldsymbol{G} = \boldsymbol{P}'\boldsymbol{AP}$.

9.2 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.

9.3 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,

$$\boldsymbol{\beta} = \begin{bmatrix} 1 \\ -\boldsymbol{\beta}_2 \end{bmatrix}. \tag{9.1}$$

With the conventional normalization the model is not invariant with respect to orthogonal transformations; for the orthogonal matrix $\boldsymbol{O}$ defined by (4.6),

$$\boldsymbol{O}'\boldsymbol{\beta} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} 1 \\ -\beta_2 \end{bmatrix} = \begin{bmatrix} \cos\theta - \beta_2 \sin\theta \\ -\sin\theta - \beta_2 \cos\theta \end{bmatrix} \tag{9.2}$$

which is of the form $(1, -\beta_2^*)$ only if $\theta = 0$, that is, if $\boldsymbol{O} = \boldsymbol{I}$. 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.

10 Admissibility of estimators

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

$$L(\boldsymbol{\alpha}, \hat{\boldsymbol{\alpha}}) = 1 - \left(\boldsymbol{\alpha}'\hat{\boldsymbol{\alpha}}\right)^2 = \sin^2\left(\hat{\theta} - \theta\right). \tag{10.1}$$

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 (10.1) 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 λ, $\boldsymbol{\eta}$, 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 λ, $\boldsymbol{\eta}$, and $\boldsymbol{\alpha}$.

Invariant estimation can be explored in terms of the model $\boldsymbol{Q} = \lambda\boldsymbol{\eta}\boldsymbol{\alpha}' + \boldsymbol{W}$, where $\boldsymbol{\eta}'\boldsymbol{\eta} = 1$ and $\boldsymbol{\alpha}'\boldsymbol{\alpha} = 1$, $\boldsymbol{\eta}$ is $K \times 1$ and $\boldsymbol{\alpha}$ is 2×1. We can show that the LIMLK estimator of $\boldsymbol{\beta}$ is the *best invariant estimator* of $\boldsymbol{\beta}$. The LIMLK estimator of $\boldsymbol{\beta}$ is defined by

$$\boldsymbol{G}\hat{\boldsymbol{\beta}} = r_1\hat{\boldsymbol{\beta}}, \qquad \hat{\boldsymbol{\beta}}'\hat{\boldsymbol{\beta}} = 1; \tag{10.2}$$

the LIMLK estimator is the eigenvector of $\boldsymbol{G}$ corresponding to the smaller eigenvalue of $\boldsymbol{G}$ (Section 6). Similarly the maximum likelihood estimator of $\boldsymbol{\alpha}$ is the eigenvector corresponding to the larger eigenvalue. The maximum likelihood estimator of θ is t, where $\tilde{\boldsymbol{\beta}} = (\cos t, \sin t)'$.

We use the loss function (10.1) where

$$\boldsymbol{\alpha} = \begin{bmatrix} -\sin\theta \\ \cos\theta \end{bmatrix}, \qquad \hat{\boldsymbol{\alpha}} = \begin{bmatrix} -\sin\hat{\theta} \\ \cos\hat{\theta} \end{bmatrix}. \tag{10.3}$$

This loss function is invariant with respect to transformations

$$\boldsymbol{\alpha} \longrightarrow \boldsymbol{O}_a\boldsymbol{\alpha}, \qquad \hat{\boldsymbol{\alpha}} \longrightarrow \boldsymbol{O}_a\hat{\boldsymbol{\alpha}} \tag{10.4}$$

or equivalently $\theta \to \theta + a$ and $\hat{\theta} \to \hat{\theta} + a$ or equivalently $\boldsymbol{\beta} \to \boldsymbol{O}_a'\boldsymbol{\beta}$ and $\hat{\boldsymbol{\beta}} \to \boldsymbol{O}_a'\hat{\boldsymbol{\beta}}$.

We can show that the LIMLK estimator is Bayes for the loss function $L(\boldsymbol{\alpha}, \hat{\boldsymbol{\alpha}})$ and the prior uniform distributions on $\boldsymbol{\eta}'\boldsymbol{\eta} = 1$ and $\boldsymbol{\alpha}'\boldsymbol{\alpha} = 1$. The Bayes loss function is

$$\begin{aligned} &\frac{1}{(2\pi)^K}e^{-\frac{1}{2}\lambda^2}\iiint\left[1 - \left(\hat{\boldsymbol{\alpha}}'\boldsymbol{\alpha}\right)^2\right]e^{-\frac{1}{2}\operatorname{tr}\boldsymbol{Q}'\boldsymbol{Q}+\lambda\boldsymbol{\eta}'\boldsymbol{Q}\boldsymbol{\alpha}}\,d\boldsymbol{Q}\,dU_2(\boldsymbol{\alpha})\,dU_K(\boldsymbol{\eta}) \\ &\qquad = \frac{1}{(2\pi)^K}e^{-\frac{1}{2}\lambda^2}\iiint\sin^2\left(\hat{\theta} - \theta\right)e^{-\frac{1}{2}\operatorname{tr}\boldsymbol{Q}'\boldsymbol{Q}+\lambda\boldsymbol{\eta}'\boldsymbol{Q}\boldsymbol{\alpha}}\,d\boldsymbol{Q}\,dU_2(\boldsymbol{\alpha})\,dU_K(\boldsymbol{\eta}). \end{aligned} \tag{10.5}$$

The Bayes estimator $\hat{\boldsymbol{\alpha}}$ is the function of $\boldsymbol{Q}$ that is invariant and minimizes the Bayes loss. Application of the lemma to the Bayes loss gives the Bayes loss as the minimum with respect to $\hat{\theta}$ of

$$\frac{1}{(2\pi)^K} e^{-\frac{1}{2}\lambda^2} \iint \left[1 - \left(\hat{\boldsymbol{\alpha}}'\boldsymbol{\alpha}\right)^2\right] e^{-\frac{1}{2}\operatorname{tr}\boldsymbol{Q}'\boldsymbol{Q}+\lambda\sqrt{\boldsymbol{\alpha}'\boldsymbol{Q}'\boldsymbol{Q}\boldsymbol{\alpha}}}\, d\boldsymbol{Q}\, dU_2(\boldsymbol{\alpha})$$
$$= \frac{1}{(2\pi)^{K+1}} e^{-\frac{1}{2}\lambda^2} \iint \sin^2\left(\hat{\theta} - \theta\right) e^{-\frac{1}{2}\operatorname{tr}\boldsymbol{Q}'\boldsymbol{Q}+\lambda\sqrt{\boldsymbol{\alpha}'_{t-\theta}\boldsymbol{R}\boldsymbol{\alpha}_{t-\theta}}}\, d\boldsymbol{Q}. \quad (10.6)$$

Since $\boldsymbol{\alpha}'_{t-\theta}\boldsymbol{R}\boldsymbol{\alpha}_{t-\theta} = r_2 - (r_2 - r_1)\sin^2(t-\theta)$, (10.6) is minimized for $\hat{\theta} = \theta$ and $\theta = t$. Hence the LIMLK estimator is the Bayes estimator and hence is the best invariant estimator for every λ. In turn this implies that the LIMLK estimator is admissible. This approach is somewhat different from that of Anderson, Stein, and Zaman.

11 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

$$\boldsymbol{Z}_{2.1} = \boldsymbol{Z}_2 - \boldsymbol{Z}_1\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12}, \quad (11.1)$$

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

$$\boldsymbol{Y}_1 = \boldsymbol{Z}_1\left(\boldsymbol{\Pi}_{11} + \boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12}\boldsymbol{\Pi}_{21}\right) + \boldsymbol{Z}_{2.1}\boldsymbol{\Pi}_{21} + \boldsymbol{V}_1. \quad (11.2)$$

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

$$\boldsymbol{A}_{22.1} = \boldsymbol{Z}_{2.1}'\boldsymbol{Z}_{2.1} = \boldsymbol{A}_{22} - \boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12}, \quad (11.3)$$

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

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 http://dx.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 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)." *Estadistica* 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 http://dx.doi.org/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 http://dx.doi.org/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.