cemmap
centre for microdata methods and practice

Optimal significance tests in simultaneous equation models

T. W. Anderson

The Institute for Fiscal Studies
Department of Economics, UCL

cemmap working paper CWP18/10

An ESRC Research Centre

Institute for Fiscal Studies, 7 Ridgmount Street, London WC1E 7AE **Tel:** 44 (0)20 7291 4800 **Email:** info@cemmap.ac.uk **www.cemmap.ac.uk**

Optimal Significance Tests in Simultaneous Equation Models *

T. W. Anderson †

June 26, 2010

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.

*TWA10-6-26. The author thanks Naoto Kunitomo (Graduate School of Economics, University of Tokyo) for his generous assistance. An early version of this paper was presented to the James Durbin Seminar October 29, 2009, sponsored by the London School of Economics and University College, London.

†Department of Statistics and Department of Economics, Stanford University

1. Introduction

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

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

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

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

Anderson (1976) pointed out that a structural equation in a simultaneous equation model is the same as a *linear functional relationship* in the statistical literature. Lindley (1953) and Creasy (1956) considered the likelihood ratio test of the slope parameter in this model.

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

2. A simultaneous equation model

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

The second submatrix of (2.7),

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

defines $\boldsymbol{\beta}_1$ except for a multiplicative constant if and only if the rank of $\boldsymbol{\Pi}_{21}$ is $G_1 - 1$. 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. 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

$$\mathbf{Y}\boldsymbol{\beta} = \mathbf{u} \ . \tag{2.9}$$

Later the results will be generalized.

3. Invariance and normalization

The model (2.1), $\mathbf{\Omega}$, (2.8), and $H_0 \ : \ \boldsymbol{\beta} = \boldsymbol{\beta}_0$ is invariant with respect to linear transformations of the exogenous variables

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

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

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

and

$$\mathbf{H}^{+} = \mathbf{Y}^{'}\mathbf{Y} - \mathbf{P}^{+^{'}}\mathbf{A}\mathbf{P}^{+} = \mathbf{H} \ . \tag{3.3}$$

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

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

which determines the constant except for sign. The model (2.1),(2.8) and (3.4) is invariant with respect to transformations :

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

and

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

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

$$\mathbf{P}^{*} = \mathbf{P}\mathbf{\Phi} \ , \ \mathbf{P}^{*^{'}}\mathbf{A}\mathbf{P}^{*} = \mathbf{\Phi}^{'}\mathbf{P}^{'}\mathbf{A}\mathbf{P}\mathbf{\Phi} \ , \ \mathbf{H}^{*^{'}} = \mathbf{\Phi}^{'}\mathbf{H}\mathbf{\Phi} \tag{3.7}$$

and

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

We also consider the model (2.1) and (2.8) when $\mathbf{\Omega}$ (the covariance matrix of a row of $\mathbf{V}$) is known. In this case we can make a transformation (3.5) and (3.6) so $\mathbf{\Omega} = \mathbf{I}$. In that case the first equation in (3.6) is

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

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

$$\mathbf{Y}^* = \mathbf{YO}\ ,\ \mathbf{\Pi}^* = \mathbf{\Pi O}\ ,\ \boldsymbol{\beta}^* = \mathbf{O}'\boldsymbol{\beta}\ ,\ \boldsymbol{\beta}_0^* = \mathbf{O}'\boldsymbol{\beta}_0\ ,\ \boldsymbol{\beta}'\boldsymbol{\beta} = 1\ . \tag{3.10}$$

4. A Canonical form for $G = 2$ and polar coordinates

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

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

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

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

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

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

This is the polar or angular representation of the coefficient.

The $K \times 2$ matrix $\mathbf{\Pi}$ of rank 1 can be parametrized as

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

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

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

Note that

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

is an orthogonal matrix.

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

Now make a transformation (3.1) so $\mathbf{A}^+ = \mathbf{I}_K$, $\mathbf{P}^+ = \mathbf{Q}$,

$$\boldsymbol{\Pi}^+ = \boldsymbol{\pi}\boldsymbol{\alpha}_\theta^{'} \ , \ \mathbf{P}^{'}\mathbf{A}\mathbf{P} = \mathbf{Q}^{'}\mathbf{Q} \ , \ \mathbf{V}^{'}\mathbf{Z}\mathbf{A}^{-1}\mathbf{Z}^{'}\mathbf{V} = \mathbf{W}^{'}\mathbf{W} \ , \tag{4.7}$$

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

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

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

5. The density of Q

The density of $\mathbf{Q}$ is

$$\begin{aligned} \frac{1}{(2\pi)^K} e^{-\frac{1}{2}\mathrm{tr}\mathbf{W}^{'}\mathbf{W}} &= \frac{1}{(2\pi)^K} e^{-\frac{1}{2}\mathrm{tr}\left(\mathbf{Q}^{'}\mathbf{Q}+\boldsymbol{\pi}\boldsymbol{\pi}^{'}-2\boldsymbol{\alpha}_\theta\boldsymbol{\pi}^{'}\mathbf{Q}\right)} \\ &= \frac{1}{(2\pi)^K} e^{-\frac{1}{2}\mathrm{tr}(\mathbf{Q}^{'}\mathbf{Q})-\frac{1}{2}\boldsymbol{\pi}^{'}\boldsymbol{\pi}+\boldsymbol{\pi}^{'}\mathbf{Q}\boldsymbol{\alpha}_\theta} \ . \end{aligned} \tag{5.1}$$

Let

$$\boldsymbol{\pi}^{'}\boldsymbol{\pi} = \lambda^2 \ , \ \boldsymbol{\pi} = \lambda\boldsymbol{\eta} \ , \tag{5.2}$$

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

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

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

$$\boldsymbol{\alpha}_\theta \to \mathbf{O}_a \boldsymbol{\alpha}_\theta \ , \ \boldsymbol{\alpha}_{\theta_0} \to \mathbf{O}_a \boldsymbol{\alpha}_{\theta_0} \ , \ \boldsymbol{\eta}_\theta \to \mathbf{O}_b \boldsymbol{\eta}_\theta \ . \tag{5.4}$$

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

6. Reduction to G

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

Lemma 1 : A function of $\mathbf{Q}$ that is invariant with respect to

$$\boldsymbol{Q} \to \mathbf{O}_a \mathbf{Q} \ , \ \boldsymbol{Q} \to \mathbf{Q}\mathbf{O}_b \ , \tag{6.1}$$

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

Proof : $\mathbf{G}$ is a function of $\mathbf{Q}$ that is invariant. If there are $\mathbf{Q}_1$ and $\mathbf{Q}_2$ such that

$$\mathbf{Q}_1'\mathbf{Q}_1 = \mathbf{Q}_2'\mathbf{Q}_2 \ , \tag{6.2}$$

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

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

7. Density of G.

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

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

See Anderson and Girshick (1944): "Some extensions of the Wishart distribution," *Annals of Mathematical Statistics.*

The density or likelihood of $\mathbf{G}$ is

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

where

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

and $(z/2)^\nu \mathrm{I}^*_\nu(z^2)=\mathrm{I}_\nu(z)$ is the modified Bessel function of order ν (Abramowitz and Stigun).

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

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

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

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

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

Transform $\mathbf{G}$ (2×2) to (r_1,r_2,t), The Jacobian of the transformation is r_2-r_1. See Appendix 1.

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

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

where

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

Let

$$n(r_1,r_2)=\frac{(r_2-r_1)(r_1r_2)^{\frac{1}{2}(K-3)}e^{-(r_1+r_2)/2}}{2^{\frac{1}{2}K+1}\pi^{\frac{1}{2}}\Gamma\left[\frac{1}{2}(K-1)\right]}\ . \tag{7.9}$$

Then the density of r_1, r_2, and t is

$$h(r_1, r_2, t|\theta, \lambda) = n(r_1, r_2) e^{-\frac{1}{2}\lambda^2} \mathbf{\Gamma}^2_{\frac{1}{2}(K-2)}(\lambda^2 c^2) \ . \tag{7.10}$$

Since we have $c^2 = \boldsymbol{\alpha}'_{t-\theta} \mathbf{R} \boldsymbol{\alpha}_{t-\theta}$ is the Lower Right Hand (LRH) corner of $\mathbf{O}'_{t-\theta}\mathbf{R}\mathbf{O}_{t-\theta}$,

$$\begin{aligned}
\mathrm{I}^*_{\frac{1}{2}(K-2)}(\lambda^2 r_2) &= \sum_{j=0}^{\infty} \left(\frac{\lambda^2 c^2}{4}\right)^j \frac{1}{j!\Gamma[j+1+\frac{1}{2}(K-2)]} \\
&= \mathrm{LRH}\left[\sum_{j=0}^{\infty} \left(\frac{\lambda^2}{4}\right)^j \frac{1}{j!\Gamma[j+1+\frac{1}{2}(K-2)]} (\mathbf{O}'_{t-\theta}\mathbf{R}\mathbf{O}_{t-\theta})^j\right] \qquad (7.11) \\
&= \mathrm{LRH}\left[\sum_{j=0}^{\infty} \left(\frac{\lambda^2}{4}\right)^j \frac{1}{j!\Gamma[j+1+\frac{1}{2}(K-2)]} \mathbf{O}'_{t-\theta}\mathbf{R}^j\mathbf{O}_{t-\theta}\right] \\
&= \mathrm{LRH}\left[\mathbf{O}'_{t-\theta} \sum_{j=0}^{\infty} \left(\frac{\lambda^2}{4}\right)^j \frac{1}{j!\Gamma[j+1+\frac{1}{2}(K-2)]} \mathbf{R}^j\mathbf{O}_{t-\theta}\right] \\
&= \mathrm{LRH}\left[\mathbf{O}'_{t-\theta} \begin{bmatrix} \mathrm{I}^*_{\frac{1}{2}(K-2)}(\lambda^2 r_1) & 0 \\ 0 & \mathrm{I}^*_{\frac{1}{2}(K-2)}(\lambda^2 r_2) \end{bmatrix} \mathbf{O}_{t-\theta}\right] \\
&= \mathrm{I}^*_{\frac{1}{2}(K-2)}(\lambda^2 r_1)\sin^2(t-\theta) + \mathrm{I}^*_{\frac{1}{2}(K-2)}(\lambda^2 r_2)\cos^2(t-\theta) \ .
\end{aligned}$$

The density of r_1, r_2, t is

$$\begin{aligned}
(7.12) \qquad & h(r_1, r_2, t|\theta, \lambda) \\
= \ & n(r_1, r_2) e^{-\lambda^2/2} \left[\mathrm{I}^*_{\frac{1}{2}(K-2)}(\lambda^2 r_1)\sin^2(t-\theta) + \mathrm{I}^*_{\frac{1}{2}(K-2)}(\lambda^2 r_2)\cos^2(t-\theta)\right] \ , \\
& 0 \le r_1 \le r_2 < \infty \ , \ -\pi \le t \le \pi \ .
\end{aligned}$$

The (marginal) density of r_1 and r_2 is

$$h(r_1, r_2|\lambda) = \frac{1}{2} n(r_1, r_2) e^{-\lambda^2/2} \left[\mathrm{I}^*_{\frac{1}{2}(K-2)}(\lambda^2 r_1) + \mathrm{I}^*_{\frac{1}{2}(K-2)}(\lambda^2 r_2)\right] \ . \tag{7.13}$$

8. Likelihood ratio criterion

The density (i.e. likelihood) is maximized for

$$(8.1) \qquad H_0 : \theta = \theta_0$$

at $\hat{\theta} = \theta_0$ since $\mathrm{I}^*_\nu(\lambda^2 c^2)$ is an increasing function of c^2. Then

$$\begin{aligned}(8.2) \quad & \max_{H_0} \mathrm{Lhd} \\ = \;& n(r_1, r_2) e^{-\lambda^2/2} \mathrm{I}^*_{\frac{1}{2}(K-2)}(\lambda^2 c_0^2) \\ = \;& n(r_1, r_2) e^{-\lambda^2/2} \left[\mathrm{I}^*_{\frac{1}{2}(K-2)}(\lambda^2 r_1) \sin^2(t-\theta_0) + \mathrm{I}^*_{\frac{1}{2}(K-2)}(\lambda^2 r_2) \cos^2(t-\theta_0) \right] ,\end{aligned}$$

where

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

The likelihood is maximized for

$$(8.4) \qquad H_1 : -\pi \leq \theta \leq \pi$$

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

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

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

$$\begin{aligned}(8.6) \qquad \mathrm{LRC} &= \frac{\max_{H_0} \mathrm{Lhd}}{\max_{H_1} \mathrm{Lhd}} = \frac{\mathrm{I}^*_{\frac{1}{2}(K-2)}(\lambda^2 c_0^2)}{\mathrm{I}^*_{\frac{1}{2}(K-2)}(\lambda^2 r_2)} \\ &= \frac{\mathrm{I}^*_{\frac{1}{2}(K-2)}\{\lambda^2 \left[r_2 - (r_2 - r_1)\sin^2(t-\theta_0) \right]\}}{\mathrm{I}^*_{\frac{1}{2}(K-2)}(\lambda^2 r_2)} \ . \\ &= \frac{\mathrm{I}^*_{\frac{1}{2}(K-2)}(\lambda^2 r_1)\sin^2(t-\theta_0) + \mathrm{I}^*_{\frac{1}{2}(K-2)}(\lambda^2 r_2)\cos^2(t-\theta_0)}{\mathbf{\Gamma}^*_{\frac{1}{2}(K-2)}(\lambda^2 r_2)} \ .\end{aligned}$$

The maximum likelihood estimator of θ is $\hat{\theta} = t$; the maximum likelihood estimator of $\boldsymbol{\beta}$ is $\hat{\boldsymbol{\beta}} = \boldsymbol{\beta}_{\hat{\theta}}$. The LR test is to reject the null hypothesis if the LRC is less than a constant. The null hypothesis is accepted if $\sin^2(t - \theta_0)$ is sufficiently small.

9. Bayes Test

Consider *the prior* probability structure for the parameter θ consisting of a probability of H_0 of $\Pr\{\theta = \theta_0\}$ and a uniform density on H_1

$$\frac{1}{2\pi}\left[1 - \Pr\{\theta = \theta_0\}\right] \quad , -\pi \leq \theta \leq \pi \ . \tag{9.1}$$

Let the loss structure be

parameter \ action	accept H_0	reject H_0
H_0	0	1
H_1	1	0

Then the average risk is

$$\Pr\{\theta = \theta_0\}\Pr\{\text{Reject } H_0|\theta_0\} + \left[1 - \Pr\{\theta = \theta_0\}\right] \int_{-\pi}^{\pi} \frac{1}{2\pi} \Pr\{\text{Accept } H_0|\theta\} d\theta \ . \tag{9.2}$$

The Bayes aqcceptance region is

$$\frac{h(r_1, r_2, t|\theta_0, \lambda)}{\frac{1}{2\pi}\int_{-\pi}^{\pi} h(r_1, r_2, t|\theta_0, \lambda) d\theta} \geq \text{aconstant} \ . \tag{9.3}$$

The ratio (9.3) is

$$\begin{aligned} &\frac{\mathrm{I}^*_{\frac{1}{2}(K-2)}(\lambda^2 r_1)\sin^2(t-\theta_0) + \mathrm{I}^*_{\frac{1}{2}(K-2)}(\lambda^2 r_2)\cos^2(t-\theta_0)}{\frac{1}{2\pi}\int_{-\pi}^{\pi}\left[\mathrm{I}^*_{\frac{1}{2}(K-2)}(\lambda^2 r_1)\sin^2(t-\theta) + \mathrm{I}^*_{\frac{1}{2}(K-2)}(\lambda^2 r_2)\cos^2(t-\theta)\right] d\theta} \\ = &\frac{2\left[\mathrm{I}^*_{\frac{1}{2}(K-2)}(\lambda^2 r_1)\sin^2(t-\theta_0) + \mathrm{I}^*_{\frac{1}{2}(K-2)}(\lambda^2 r_2)\cos^2(t-\theta_0)\right]}{\mathrm{I}^*_{\frac{1}{2}(K-2)}(\lambda^2 r_1) + \mathrm{I}^*_{\frac{1}{2}(K-2)}(\lambda^2 r_2)} \\ = &\frac{2\{\mathrm{I}^*_{\frac{1}{2}(K-2)}(\lambda^2 r_2) - \left[\mathrm{I}^*_{\frac{1}{2}(K-2)}(\lambda^2 r_2) - \mathrm{I}^*_{\frac{1}{2}(K-2)}(\lambda^2 r_1)\right]\sin^2(t-\theta_0)\}}{\mathrm{I}^*_{\frac{1}{2}(K-2)}(\lambda^2 r_1) + \mathrm{I}^*_{\frac{1}{2}(K-2)}(\lambda^2 r_2)} \ . \end{aligned} \tag{9.4}$$

We have used

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} \sin^2\theta d\theta = \frac{1}{2\pi}\int_{-\pi}^{\pi} \cos^2\theta d\theta = \frac{1}{2} \ . \tag{9.5}$$

The left-hand side of (9.4) is a factor times the Likelihood Ratio Criterion.

Theorem 1 : The likelihood ratio test of H_0 is the Bayes solution for a prior alternative of a uniform distribution of θ.

10. Admissibility

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

Definition 1: A is *as good as* B if

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

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

Definition 2: A is *better than* B if the equations above hold with strict inequality for at least one ω.

Definition 3: A is *admissible* if there is no B better than A.

A Bayes test is based on a probability distribution $\mathbf{\Lambda}_0$ on $\mathbf{\Omega}_0$ and $\mathbf{\Lambda}_1$ on $\mathbf{\Omega}_1$. The test with acceptance region

$$\frac{\int f(\mathbf{y}|\theta) d\Lambda_0(\theta)}{\int f(\mathbf{y}|\theta) d\Lambda_1(\theta)} > \text{constant} \tag{10.3}$$

is admissible for testing $\mathbf{\Omega}_0$ vs. $\mathbf{\Omega}_1$. (See Anderson (2003), p.199, for example.)

The Bayes test (9.4) is of the form of (10.3). Thus we have established a main result.

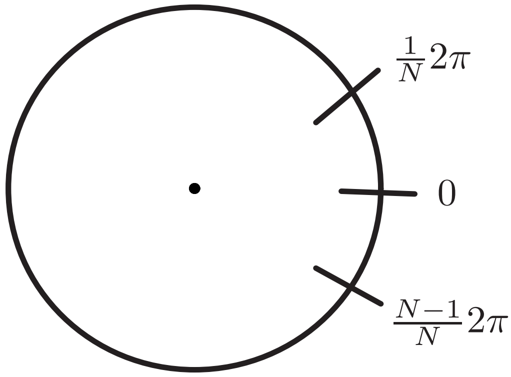

Figure 1: Finite Example

Theorem 2 : The test with acceptance region (9.4) is an admissible invariant test of $H_0 : \theta = \theta_0$.

11. Admissibility over all tests

Now we consider admissibility with respect to all tests. We want to show that the best invariant test of $\theta = \theta_0$ is admissible within the class of all tests. As an example, consider the model in which θ can take on a finite number of values.

Finite example : The possible parameter values are

(11.1) $$\theta = 0, \frac{1}{N}2\pi, \frac{2}{N}2\pi, \cdots, \frac{N-1}{N}2\pi \ .$$

Consider the group of transformations

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

Let these values of θ be labelled as $\theta_0^*, \theta_1^*, \cdots, \theta_{N-1}^*$. Each of them corresponds to a null hpothesis. Define a test of the hypothesis $\theta = \theta_k^*$ by the acceptance region

$A_k^* = A_k^*(t, r_1, r_2)$ in the space of t, r_1, r_2. The set of tests is an *invariant* set if

$$A_k^*(t - \theta_k^*, r_1, r_2) = A_j^*(t - \theta_k^*, r_1, r_2) \tag{11.3}$$

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

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

$$\Pr\{\theta = \theta_j^*\} = \frac{1}{N}\ ,\ j = 0, 1, \cdots, N-1. \tag{11.4}$$

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

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

$$\frac{1}{N}\sum_{k=0}^{N-1} A_k^*(t - \theta_k^* + \theta_i^*, r_1, r_2)\ . \tag{11.5}$$

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

Lemma 2 : If a test with an invariant family of acceptance regions $A_0, A_1, \cdots, A_{N-1}$ is admissible in the set of invariannt tests, it is admissible in the set of all tests.

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

Lemma 2 is a special case of so-called Hunt-Stein theorem to the effect that the best invariant test is admissible in the class of all tests if the group transformations defining invariance is finite or compact. See Zaman (1996), Section 7.9, or Lehmann (1986), Theorem 7 of Chapter 3. The proofs of such theorems are based on the argument that the randomization of the noninvariant tests yields an invariant test that is as good as the noninvariant test.

On the model

$$\mathbf{Q} = \lambda \boldsymbol{\eta} \boldsymbol{\alpha}' + \mathbf{W} \tag{11.6}$$

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

12. Conclusions

12.1 Estimation

Anderson, Stein, and Zaman (1985) considered the estimation of $\boldsymbol{\eta}$ and $\boldsymbol{\alpha}$ in the model $\mathbf{Q} = \lambda \boldsymbol{\eta} \boldsymbol{\alpha}'$, where $\boldsymbol{\eta}'\boldsymbol{\eta} = 1$ and $\boldsymbol{\alpha}'\boldsymbol{\alpha} = 1$. The loss of estimating $\boldsymbol{\alpha}$ by $\hat{\boldsymbol{\alpha}}$ was

$$L(\boldsymbol{\alpha}, \hat{\boldsymbol{\alpha}}) = 1 - (\boldsymbol{\alpha}'\hat{\boldsymbol{\alpha}})^2 = \sin^2(\hat{\theta} - \theta) \tag{12.1}$$

and $\hat{\theta} - \theta$ is the angle between the vector $\boldsymbol{\alpha}$ and an estimator $\hat{\boldsymbol{\alpha}}$. When $G = 2$, this is the model treated in the paper. The estimator t of θ is the LIMLK estimator. Corollary 1 of Anderson, Stein, and Zaman (1985) states that the LIMLK estimator is admissible for the loss function (13.1) and every fixed λ and hence for all λ.

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

12.2 Testing

We consider testing $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0 = \boldsymbol{\beta}_{\theta_0}$ on the basis of $\mathbf{G} = \mathbf{O}_t \mathbf{R} \mathbf{O}_t'$. The risk of a test may depend on θ_0 and θ. Let

$$\gamma(\theta, \theta_0) = \Pr\{\text{Accept} H_0 | \theta, \theta_0\} \ . \tag{12.2}$$

In this notation the significance level of a test is $\gamma(\theta_0, \theta_0)$ and the power of a test is $1-\gamma(\theta, \theta_0)$. The admissibility of the LR test is that for any other test the significance level is greater or the power is not as great or both.

12.3 A more general model.

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

$$\mathbf{Z}_{2.1} = \mathbf{Z}_2 - \mathbf{Z}_1 \mathbf{A}_{11}^{-1} \mathbf{A}_{12} \ , \tag{12.3}$$

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

$$\mathbf{Y}_1 = \mathbf{Z}_1 \left(\boldsymbol{\Pi}_{11} + \mathbf{A}_{11}^{-1} \mathbf{A}_{12} \boldsymbol{\Pi}_{21} \right) + \mathbf{Z}_{2.1} \boldsymbol{\Pi}_{21} + \mathbf{V}_1 \ . \tag{12.4}$$

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

$$\mathbf{A}_{22.1} = \mathbf{Z}_{2.1}' \mathbf{Z}_{2.1} = \mathbf{A}_{22} - \mathbf{A}_{21} \mathbf{A}_{11}^{-1} \mathbf{A}_{12} \ , \tag{12.5}$$

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

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

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

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

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

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

[6] Anderson, T.W. and N. Kunitomo (2008), "Testing a Hypothesis about a Structural Coefficient in a Simultaneous Equation Model for Known Covariance Matrix," (In honor of Raul P. Mentz), Unpublished.

[7] Anderson, T.W. and H. Rubin (1949), "Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations," *Annals of Mathematical Statistics*, Vol. 20, 46-63.

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

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

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

[11] Creasy, M.A. (1956), "Confidence Limits for the Gradiant in the Linear Functional Relationship," *Journal of the Royal Statistical Society*, B, Vol. 18, 65-69.

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

[13] Lindley, D. V. (1953), "Estimation of a Functional Relationship," *Biometrika*, Vol. 40, 47-49.

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

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

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

Appendix A. Jacobian

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

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

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

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

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