6-1-1968

The Distribution, When the Residuals Are Small, of Statistics Testing Overidentifying Restrictions

Joseph B. Kadane

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS

AT YALE UNIVERSITY

Box 2125, Yale Station
New Haven, Connecticut

COWLES FOUNDATION DISCUSSION PAPER NO. 251

THE DISTRIBUTION, WHEN THE RESIDUALS ARE SMALL, OF STATISTICS TESTING OVERIDENTIFYING RESTRICTIONS

Joseph B. Kadane

June 19, 1968

THE DISTRIBUTION, WHEN THE RESIDUALS ARE SMALL,

OF STATISTICS TESTING OVERIDENTIFYING RESTRICTIONS*

by

Joseph B. Kadane

Abstract

In the estimation of simultaneous equation econometric models, overidentifying restrictions improve estimates of the remaining parameters. Natural test statistics for the hypothesis that an equation is overidentified have been developed by Anderson and Rubin and by Basmann. If the residuals are jointly normal, serially uncorrelated, and small, both the above overidentification test statistic have the Snedecor F distribution asymptotically as the variance of the residuals gets small. This gives analytic confirmation of Monte Carlo results of Basmann. The results given apply to linear models in which predetermined variables can be exogenous or lagged endogenous.

1. Introduction

Anderson and Rubin [1] found that the likelihood ratio statistic for testing the overidentifying restrictions on a single equation in a system of simultaneous equations is equivalent to the smallest root, λ,

---

*The research described in this paper was carried out under a grant from the National Science Foundation.

of the determinental equation appearing in the theory of the limited information (single equation) maximum likelihood estimator. They also proposed a conservative test of significance for λ, comparing $\frac{T-K}{K_2}(\lambda-1)$ with an F distribution with K_2 and $T-K$ degrees of freedom and rejecting for large values. (Here T is the sample size, K the number of predetermined variables in the system and K_2 the number of predetermined variables excluded from the equation in question.)

In a later paper, Anderson and Rubin [2] found that $T(\lambda-1)$ has a large-sample asymptotic χ^2 distribution with L degrees of freedom where L is the degree of overidentification.

Basmann [4] pointed out the difficulty that the proposed conservative test for λ does not coincide with the large-sample test as the sample size increases and proposed that $\frac{T-K}{L}(\lambda-1)$ be compared to an F distribution with L and $T-K$ degrees of freedom. He justified this proposal on heuristic grounds, and referred to an unpublished Monte-Carlo study [3] which supports his proposal at one set of parameter and exogenous variable values. His criticism does not imply that Anderson and Rubin erred, but rather that their proposed test of significance is <u>very</u> conservative.

Additionally he proposed a slightly different test statistic $\hat{\lambda}$ ($\hat{\phi}+1$ in the notation of [4]) based on two-stage least squares (GCL) estimates of structural parameters, and proposed the same test of significance for $\hat{\lambda}$: compare $\frac{T-K}{L}(\hat{\lambda}-1)$ to an F distribution with L and T-K degrees of freedom. The Monte Carlo study reported in [4] strongly supported this approximation to the distribution of $\hat{\lambda}$.

In a later article Basmann [5] derived the exact distribution of $\hat{\lambda}$ for a special case. Finally Richardson [7] derived the excat distribution of $\hat{\lambda}$ in two special cases (of which one was the same as that of Basmann above) and showed that $\hat{\lambda}$ has the asymptotic distribution conjectured by Basmann in [3] as "the concentration parameter" $\bar{\mu}^2$ approaches infinity.

The purpose of this paper is to demonstrate that the distribution (for both statistics) originally conjectured by Basmann is the first order term of a (random) Taylor series expansion as the variance of the residuals in the model approaches zero (which implies that Richardson's $\bar{\mu}^2 \to \infty$.) Basmann's conjecture is established, in this sense, for any linear simultaneous equation model of any size. The results here apply to models with or without lagged endogenous variables appearing as predetermined variables.

Taylor series of this type (called small-σ asymptotics) are a useful general method for studying small-sample properties of econometric statistics. Intuitively the idea of small-σ asymptotics is that the model is getting increasingly good as $\sigma \to 0$. This is a more natural ideal case for many regression problems than the usual large sample case. Other work [6] applies small-σ asymptotics to the comparison of alternative econometric estimators.

2. Statement of Theorem

Let the complete system

(1) $$YB + Z\Gamma + \sigma U = 0$$

have a possibly overidentified (but certainly identified) first equation

(2) $$y = Y_1\beta + Z_1\gamma + \sigma u$$

where Y is a $T \times G$ matrix of endogenous variables, partitioned $Y = (y, Y_1, Y_2)$ where y is $T \times 1$, Y_1 is $T \times G_1$ and Y_2 is $T \times G_2$ $(G = G_1 + G_2 + 1)$; Z is a $T \times K$ matrix of predetermined variables, partitioned $Z = (Z_1, Z_2)$ where Z_1 is $T \times K_1$ and Z_2 is $T \times K_2$ $(K = K_1 + K_2)$; B is a non-singular $G \times G$ matrix of parameters with first column $(-1, \beta', 0')$ where -1 is a number, β is $1 \times G_1$ and 0 is a $1 \times G_2$ vector of zeros; Γ is a $K \times G$ matrix of parameters with first column $(\gamma', 0')$, where γ is $1 \times K_1$ and 0 is a $1 \times K_2$ vector of zeros; U is a $T \times G$ matrix of jointly normal residuals with zero means and covariances $Eu_{ti}u_{t'j} = \sigma_{ij}\delta_{tt'u}$; and σ is a (small) positive number. The general k-class estimate of $\begin{pmatrix}\beta\\ \gamma\end{pmatrix}$ is

$$\begin{pmatrix}\beta\\ \gamma\end{pmatrix}_k = \begin{bmatrix} Y_1' \quad Y_1 - kV'V & Y_1' \quad Z_1 \\ Z_1' \quad Y_1 & Z_1' \quad Z_1 \end{bmatrix}^{-1} \begin{bmatrix} (Y_1 - kV)' \\ Z_1' \end{bmatrix} y$$

where $V = P_Z Y_1$ and $P_X = I - X(X'X)^{-1}X'$ is the projection onto the space orthogenal to the columns of X , for any matrix X . As is well-known, the two-stage least squares (GCL) estimate, corresponds to $k = 1$, and limited information (single equation) maximum likelihood corresponds to $k = \lambda$, where

$$\lambda = \operatorname*{Min}_{\beta_*} \frac{\beta_*' Y_*' P_{Z_1} Y_* \beta_*}{\beta_*' Y_*' P_Z Y_* \beta_*} = \frac{\hat{\beta}_*' Y_*' P_{Z_1} Y_* \hat{\beta}_*}{\hat{\beta}_*' Y_*' P_Z Y_* \hat{\beta}_*} \tag{4}$$

and $Y_* = (y, Y_1)$.

$\hat{\beta}_*'$ in (4), when normalized, can be written as $(-1, \hat{\beta}_\lambda')$ where $\hat{\beta}_\lambda$ is limited information maximum likelihood estimator of β . The identifiability test statistic associated with limited information maximum likelihood is λ . The identifiability test statistic associated with two-stage least squares, λ_1 , is λ above with $(-1, \hat{\beta}_1')$, the two stage least squares estimate for $(-1, \beta')$, substituted for $\hat{\beta}$.

Now we can state

<u>Theorem:</u> Asymptotically as $\sigma \to 0$, λ and λ_1 each have the same distribution as $1 + X_1/X_2$ where X_1 has a χ^2 distri-

bution with L degrees of freedom, independent of X_2 which has a χ^2 distribution with T-K degrees of freedom.

Actually the proof applies to any k-class estimator where $k = O_p(1)$ as $\sigma \to 0$. In particular, the analogous statistics with ordinary least squares or Nagar's unbiased k-class estimators have the above distribution.

3. Proof of Theorem

Lemma 1: $\lambda = O_p(1)$ as $\sigma \to 0$

Proof

$$1 \leq \lambda = \min_{\hat{\beta}} \frac{\hat{\beta}' Y_*' P_{Z_1} Y_* \hat{\beta}}{\hat{\beta}' Y_*' P_Z Y_* \hat{\beta}}$$

$$\leq \frac{(-y' + \beta' Y_1^*) P_{Z_1} (-y + Y_1 \beta)}{(-y' + \beta' Y_1') P_Z (-y + Y_1 \beta)}$$

From (2), $(-y + Y_1\beta) = -Z_1\gamma - \sigma u$. Also

$$P_{Z_1} Z_1 = P_Z Z_1 = 0 .$$

Hence

(5) $$1 \leq \lambda \leq \frac{\sigma^2 u' P_{Z_1} u}{\sigma^2 u' P_Z u} = \frac{u' P_{Z_1} u}{u' P_Z u} .$$

In the case when the predetermined variables Z are in fact exogenous (and hence considered constant), the proof of lemma 1 is complete, as the expression on the right hand side of (5) is a random variable not involving σ. In the case in which lagged endogenous variables are permitted as predetermined variables, however, more explanation is required.

Write $Z = R + \sigma S$, in general, where R is constant and S is a random variable, depending on the lag structure, not involving σ. Then since Z is partitioned $Z = (Z_1, Z_2)$, R and S can be partitioned conformably as $R = (R_1, R_2)$ and $S = (S_1, S_2)$.

Now

$$P_{Z_1} = P_{R_1} + O_p(\sigma) \quad \text{and}$$

$$P_Z = P_R + O_p(\sigma) .$$

Hence (5) can be written

$$1 \leq \lambda \leq \frac{u'P_{R_1} u + O_p(\sigma)}{u'P_R\ u + O_p(\sigma)} = O_p(1) . \qquad \text{QED.}$$

Similarly, the reduced form for a lag structure can be written

$$[Y_1, Z_1] = X + \sigma V^*$$

where X is constant and is the space spanned by R, and V^* is a random variable not involving σ.

Lemma 2

$$\begin{pmatrix}\beta\\ \gamma\end{pmatrix}_k = \begin{pmatrix}\beta\\ \gamma\end{pmatrix} + \sigma(X'X)^{-1}X'u + O_p(\sigma^2) \quad \text{if} \quad k = O_p(1)$$

[In particular, Lemma 2 applies if $k = 1$ and if $k = \lambda$ (using Lemma 1).]

Proof

$$(V, 0) = P_Z[Y_1, Z_1] = P_Z[X + \sigma V^*] = \sigma P_Z V^* \ .$$

Hence using (3),

$$\begin{pmatrix}\beta\\ \gamma\end{pmatrix}_k = \begin{pmatrix}\beta\\ \gamma\end{pmatrix} + \sigma\{(X + \sigma V^*)'(X + \sigma V^*) - k\sigma^2 V^{*\prime} P_Z V^*\}^{-1}\{X' - \sigma(I-kP_Z)V^{*\prime}\}u$$

$$= \begin{pmatrix}\beta\\ \gamma\end{pmatrix} + \sigma(X'X)^{-1}X'u + O_p(\sigma^2) \quad \text{if} \quad k = O_p(1) \qquad \text{QED.}$$

Using Lemma 2,

$$P_{Z_1}Y_*\hat{\beta}_k = P_{Z_1}(y, Y_1)\left\{\begin{pmatrix}-1\\ \beta\end{pmatrix} + \sigma\begin{bmatrix}0\\ \cdots\\ X^*\end{bmatrix}X'u + O_p(\sigma^2)\right\}$$

$$\text{where} \quad (X'X)^{-1} = \begin{bmatrix}X^*\\ \cdots\\ X^{**}\end{bmatrix}$$

$$P_{Z_1}Y_*\hat{\beta}_k = P_{Z_1}\{-y + Y_1\beta + \sigma Y_1 X^* X'u + O_p(\sigma^2)\}$$

$$= P_{Z_1}\{-Z_1\gamma - \sigma u + \sigma Y_1 X^* X'u + O_p(\sigma^2)\} \quad \text{(from (2))}$$

$$= P_{Z_1}\{-\sigma u + \sigma Y_1 X^* X'u + \sigma Z_1 X^{**} X'u + O_p(\sigma^2)\}$$

$$= -\sigma P_{Z_1} P_X u + O_p(\sigma^2) \ .$$

So

$$\lambda = \frac{\sigma^2 u' P_X P_{Z_1} P_X u + 0_p(\sigma^3)}{\sigma^2 u' P_X P_Z P_X u + 0_p(\sigma^3)}$$

$$= \frac{u' P_X u}{u' P_Z u} + 0_p(\sigma) = 1 + \frac{u' [P_X - P_R] u}{u' P_R u} + 0_p(\sigma)$$

P_R and $(P_X - P_R)$ are projections and $P_R(P_X - P_R) = 0$ since X is in the space spanned by R. $u'[P_X - P_R]u$ and $u'P_R u$ are therefore independent χ^2 distributions with degrees of freedom $tr(P_X - P_R) = L$ and $trP_R = T\text{-}K$ respectively. This completes the proof of the theorem.

REFERENCES

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

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

[3] Basmann, R.L. (1959, unpublished mineo) "An Experimental Investigation of Some Approximate Finite Sample Tests of Linear Restrictions on Matrices of Regression Coefficients."

[4] Basmann, R.L. (1960) "On Finite Sample Distributions of Generalized Classical Linear Identifiability Test Statistics," Journal of the American Statistical Association, 55, pp. 650-659.

[5] Basmann, R.L. (1966) "On the Application of the Identifiability Test Statistic in Predictive Testing of Explanatory Economic Models," Econometric Annual of the Indian Economic Journal, pp. 387-423.

[6] Kadane, J.B. (1966, unpublished dissertation, Stanford University) "Comparison of Estimators in Simultaneous Equation Econometric Models when the Residuals are Small."

[7] Richardson, D.H. (1967) "On the Distribution of the Identifiability Test Statistic," mineo Paper No. 8, Department of Economics, University of Kansas.