Testing Overidentifying Restrictions When the Disturbances Are Small

JOSEPH B. KADANE*

Natural test statistics for the hypothesis that an equation is overidentified have been developed by Anderson and Rubin and by Basmann. If the disturbances are jointly normal, serially uncorrelated, and small, both the above overidentification test statistics have the Fisher variance-ratio distribution asymptotically as the variance of the error terms gets small. This gives an analytic explanation of Monte Carlo results of Basmann. The results given apply to linear models in which predetermined variables are exogenous.

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, \( \lambda \), of the determinantal equation appearing in the theory of the limited information (single equation) maximum likelihood estimator (see [1, Equation 4.14]). They also proposed a conservative test of significance for \( \lambda \), comparing \((T - K)/K_2(\lambda - 1)\) with an \( F \) distribution with \( K_2 \) and \( T - K \) degrees of freedom, large values being significant. (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 article, Anderson and Rubin [2] found that \( T(\lambda - 1) \) has a large-sample asymptotic \( \chi^2 \) distribution with \( L \) degrees of freedom. \( L = K_1 - G_1 \) is the degree of overidentification, where \( G_1 + 1 \) is the number of endogenous variables in the equation.

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 He justified this proposal on heuristic grounds, and referred to an unpublished Monte Carlo study [3] which supports his proposal at one set of parameters and exogenous variable values.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

Additionally he proposed a slightly different test statistic designated \( \lambda_1 \) here, \( (\phi + 1) \) in the notation of [4], based on two-stage least squares estimates of structural parameters, and proposed the same test of significance for \( \lambda_1 \): compare \((T - K)/L(\lambda_1 - 1)\) 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 \( \lambda_1 \).

* Joseph B. Kadane, presently at the Center for Naval Analyses, is on leave from Carnegie-Mellon University where he is associate professor of statistics and social sciences. His previous articles have appeared in Journal of Mathematical Analysis and Applications, American Sociological Review, Science, Sociometry, Econometrica, the Review of Economics and Statistics, and the Annals of Mathematical Statistics. The research for this article was supported by a grant from the National Science Foundation and is an extension of material which appeared in [6].
In a later article Basman [5] derived the exact distribution of \( \lambda \) for a special case. Finally Richardson [7] found the exact moments of \( \lambda \) when \( G_1 = 2 \) for any \( K_2 \) and showed that the moments of \( \lambda \), approach those of the asymptotic distribution conjectured by Basman in [3] as "the concentration parameter" \( \bar{\sigma}^2 \) approaches infinity.

The purpose of this article is to demonstrate that the distribution (for both statistics) originally conjectured by Basman is the first order term of a (random) asymptotic expansion in the variance of the disturbances in the model around zero \(^1\) (\( \sigma \rightarrow 0 \) implies that Richardson's \( \bar{\sigma}^2 \rightarrow \infty \)). Basman's conjecture is established, in this sense, for any linear simultaneous equation model of any size.

Asymptotic series of this type (called small-\( \sigma \) asymptotics) are a useful general method for studying small-sample properties of econometric statistics. Intuitively the idea of small-\( \sigma \) asymptotics is that as \( \sigma \rightarrow 0 \) the model fits with less and less variability. This is a more natural ideal case for many regression problems than the usual large sample case. Elsewhere [6], I have reported an application of small-\( \sigma \) asymptotics to the comparison of alternative econometric estimators.

2. STATEMENT OF THEOREM

Let the complete system

\[
YB + ZT + \sigma U = 0
\]  

have a possibly overidentified (but certainly identified) first equation

\[
y = Y_1\beta + Z_1\gamma + \sigma u
\]  

where \( Y \) is a \( T \times G \) matrix of endogenous variables, partitioned \( Y = (y_1, Y_2, Y_3) \) where \( y_1 \) 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 exogenous variables, assumed to be of rank \( K \), 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 \(-1, \beta', 0', \theta'\)' where \(-1\) is a scalar, \( \beta \) is \( G_1 \times 1 \) and \( 0 \) is a \( G_2 \times 1 \) vector of zeros; \( \Gamma \) is a \( K \times G \) matrix of parameters with first column \( \gamma', 0', \theta' \)' where \( \gamma \) is \( K_1 \times 1 \) and \( 0 \) is a \( K_2 \times 1 \) vector of zeros; \( U \) is a \( T \times G \) matrix of jointly normal residuals with zero means and covariances \( EUu'u = \sigma^2 \delta \delta' \) and with first column \( u \); \( \sigma_u = 1 \), and \( \sigma \) is a (small) positive number. The general \( k \)-class estimate of \( \hat{\xi} \) is

\[
\begin{bmatrix}
\hat{\beta} \\
\hat{\gamma} \\
\hat{\theta}
\end{bmatrix}
= 
\begin{bmatrix}
Y_1'Y_1 - kV^*V^* & Y_1'Z_1 \\
Z_1'Y_1 & Z_1'Z_1
\end{bmatrix}^{-1}
\begin{bmatrix}
(Y_1 - kV^*)' \\
Z_1'
\end{bmatrix} y
\]  

where \( V^* = \bar{\sigma}^2 Y_3 \) and \( \bar{P}_X = I - X(X'X)^{-1}X' \) is the projection onto the space orthogonal to the columns of \( X \), for any matrix \( X \). As is well known, the two-stage least squares estimate corresponds to \( k = 1 \), and limited information (single equation) maximum likelihood corresponds to \( k = \lambda \), where

\(^1\) That is, \( \lambda \) can be written as

\[
\lambda = A_0 + \sigma A_1 + \sigma^2 A_2 + \cdots
\]

where each \( A_i \) is a random variable not involving \( \sigma \). \( A_0 \) is evaluated in this article.
\[ \lambda = \min_{\beta_*} \frac{\beta_*' Y_*' \bar{P}_x Y_* \beta_*}{\beta_*' Y_*' \bar{P}_2 Y_* \beta_*} \]

and \( Y_* = (y, Y_1) \). The minimizing choice, \( \hat{\beta}_* \), in (4), when normalized, can be written as \((-1, \hat{\beta}_1)\) where \( \hat{\beta}_1 \) is the limited information maximum likelihood estimator of \( \beta \). The overidentification test statistic associated with limited information maximum likelihood is \( \lambda \). The overidentification test statistic associated with two-stage least squares, \( \lambda_1 \), is \( \lambda \) above with \( \hat{\beta}_* \) replaced by \((-1, \hat{\beta}_1)\), the two-stage least squares estimate for \((-1, \beta')\). Now we can state:

**Theorem:** Asymptotically as \( \sigma \to 0 \), \((\sigma^2) / (L) \) \((-1, \lambda_1 \to 1 \) and \((\sigma^2) / (L) \) \((-1, \beta') \) have the Fisher variance-ratio distribution \( F_{L,N-K} \).

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

### 3. PROOF OF THEOREM

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

**Proof:**

\[
1 \leq \lambda = \min_{\beta_*} \frac{\beta_*' Y_*' \bar{P}_x Y_* \beta_*}{\beta_*' Y_*' \bar{P}_2 Y_* \beta_*} \\
\leq \frac{(-y' + \beta_1 Y_1') \bar{P}_x (-y + Y_1 \beta_1)}{(-y' + \beta_1 Y_1') \bar{P}_2 (-y + Y_1 \beta_1)}
\]

From (2), \((-y + Y_1 \beta_1) = -Z_1 \gamma - \sigma u \). Also \( \bar{P}_2 Z_1 = \bar{P}_2 Z_1 = 0 \). Hence,

\[
1 \leq \lambda \leq \frac{\sigma^2 u' \bar{P}_x u}{\sigma^2 u' \bar{P}_2 u} = \frac{u' \bar{P}_x u}{u' \bar{P}_2 u}.
\] QED. (5)

Write

\[
[Y_1, Z_1] = X + \sigma V
\]

where \( X \) is constant and in the space spanned by \( Z \), and \( V \) is a random variable not involving \( \sigma \). If the equation is identified, \( \chi \) is of (full) rank \( G_1 + K_1 \), so tr \( \bar{P}_x = T - K + L \).

**Lemma 2:**

\[
\begin{pmatrix} \beta \\ \gamma \end{pmatrix}_k = \begin{pmatrix} \beta \\ \gamma \end{pmatrix} + \sigma (X'X)^{-1}X'u + 0 \sigma^2 \quad \text{if } k = 0 \sigma^2(1).
\]

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

**Proof:**

\[
(V^*, 0) = \bar{P}_x [Y_1, Z_1] = \bar{P}_x [X + \sigma V] = \sigma \bar{P}_x V
\]

Hence using (3),

\[
\begin{pmatrix} \beta \\ \gamma \end{pmatrix}_k = \begin{pmatrix} \beta \\ \gamma \end{pmatrix} + \sigma \{ (X + \sigma V)'(X + \sigma V) - k \sigma^2 V' \bar{P}_x V \}^{-1} (X' + \sigma V'(I - k \bar{P}_x)) u
\]

\[
= \begin{pmatrix} \beta \\ \gamma \end{pmatrix} + \sigma (X'X)^{-1}X'u + 0 \sigma^2 \quad \text{if } k = 0 \sigma^2(1)
\] QED.
Using Lemma 2,

\[ \mathcal{P}_\gamma Y_{\gamma \beta} = \mathcal{P}_{\gamma \gamma} (y, Y_\gamma) \begin{bmatrix} -1 \\ \beta \end{bmatrix} + \sigma \begin{bmatrix} 0 \\ X^* \end{bmatrix} X' u + 0_p(\sigma^2) \]

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

\[ \mathcal{P}_\gamma Y_{\gamma \beta} = \mathcal{P}_{\gamma \gamma} \{-y + Y_\gamma \beta + \sigma Y_1 X^* X' u + 0_p(\sigma^2)\} \]

\[ = \mathcal{P}_{\gamma \gamma} \{-Z_\gamma - \sigma u + \sigma Y_1 X^* X' u + 0_p(\sigma^2)\} \quad \text{(from (2))} \]

\[ = -\sigma \mathcal{P}_{\gamma \gamma} (Z_{\gamma} + X^* X' u) + 0_p(\sigma^2) \]

\[ = -\sigma \mathcal{P}_X u + 0_p(\sigma^2). \]

So

\[ \lambda = \frac{\sigma^2 u^T \mathcal{P}_X \mathcal{P}_X u + 0_p(\sigma^2)}{\sigma^2 u^T \mathcal{P}_X \mathcal{P}_Z u + 0_p(\sigma^2)} \]

\[ = \frac{u^T \mathcal{P}_X u + 0_p(\sigma)}{u^T \mathcal{P}_Z u + 0_p(\sigma)} = 1 + \frac{u^T [\mathcal{P}_X - \mathcal{P}_Z] u}{u^T \mathcal{P}_X u + 0_p(\sigma)}. \]

\( \mathcal{P}_X \) and \(( \mathcal{P}_X - \mathcal{P}_Z)\) are projections and \( \mathcal{P}_Z (\mathcal{P}_X - \mathcal{P}_Z) = 0 \) since \( X \) is in the space spanned by \( Z \). \( u^T [\mathcal{P}_X - \mathcal{P}_Z] u \) and \( u^T \mathcal{P}_Z u \) are therefore independent \( \chi^2 \) distributions with degrees of freedom \( \text{tr}(\mathcal{P}_X - \mathcal{P}_Z) = L \) and \( \text{tr}\mathcal{P}_Z = T - K \), respectively. This completes the proof of the theorem.

REFERENCES


