Econometrica, Vol. 52, No. 4 (July, 1984)

NOTES AND COMMENTS

THE COVARIANCE MATRIX OF THE INFORMATION MATRIX TEST

BY TONY LANCASTER

IN THIS NOTE we point out how the covariance matrix of the information matrix test, due to White [2], can be estimated without the computation of analytic third derivatives of the density function.

Let F be the logarithm of a density function depending on p parameters $\boldsymbol{\theta}$ and let $F_1(p \times 1)$ and $F_2(p \times p)$ contain $\{\partial F/\partial \theta_i\}$ and $\{\partial^2 F/\partial \theta_i \, \partial \theta_j\}$ respectively.

The basis of the information matrix test is the well-known equality $E(F_2) = -E(F_1 F_1')$, or in an obvious scalar notation,

$$\int F_2^{ij} f \, du = -\int F_1^i F_1^j f \, du \qquad (i, j = 1, 2, \ldots, p), \tag{1}$$

where $f = \exp\{F\}$, the variate is u and F_1 and F_2 are calculated at the true value of θ. If we differentiate both sides of (1) with respect to θ_k[1] we readily find, writing F_3^{ijk} for $\partial^3 F/\partial \theta_i \, \partial \theta_j \, \partial \theta_k$,

$$E(F_3^{ijk}) = -\left\{ E(F_2^{ij} F_1^k) + E(F_2^{ik} F_1^j) + E(F_2^{jk} F_1^i) + E(F_1^i F_1^j F_1^k) \right\} \qquad (i, j, k = 1, 2, \ldots, p). \tag{2}$$

If we let the superscript c denote the operation of column stacking a matrix and let $F_3(p^2 \times p)$, have ith row which contains the θ derivatives of the ith element of F_2^c, then the matrix form of (2) is

$$E(F_3) = -E(F_2 \otimes F_1 + F_1 \otimes F_2) - E(F_2 + F_1 F_1')^c F_1'. \tag{3}$$

The information matrix test based on n independent realizations of u compares the sample average values of F_2 and $-F_1 F_1'$, whose expectations are equal according to (1). Specifically if, following White's notation,[2] we let

$$d = (F_2 + F_1 F_1')^c, \qquad D_n = n^{-1} \textstyle\sum d \qquad (q \times 1), \quad q = p(p+1)/2,^3$$

$$\nabla D_n = n^{-1} \textstyle\sum \partial d_j/\partial \theta_k, \qquad \nabla D = E(\partial d_j/\partial \theta_k) \qquad (q \times p),$$

$$A_n = n^{-1} \textstyle\sum F_2, \qquad A = E(F_2) \qquad (p \times p),$$

$$B_n = n^{-1} \textstyle\sum F_1 F_1', \qquad B = E(F_1 F_1') \qquad (p \times p),$$

then the test statistic takes the form

$$\mathscr{I}_n = n \hat{D}_n' \hat{V}^{-1} \hat{D}_n \tag{4}$$

[1] Mild regularity conditions additional to those given by White are needed to justify differentiating under the integral sign.

[2] Except that we omit the observation subscript, i, for notational clarity. Summation is over $i = 1, 2, \ldots, n$.

[3] q is the number of distinct elements of F_2 and in what follows we assume row repetitions have been deleted from d and D_n.

where the $\hat{}$ indicates evaluation of θ at the maximum likelihood estimate, $\hat{\theta}$, and $\hat{V}$ is a consistent estimator of the covariance matrix of $\sqrt{n}\,\hat{D}_n$ which is asymptotically $N(0, V)$ when F is correctly specified. The matrix V is

$$V = E(d - \nabla D \cdot A^{-1} \cdot F_1)(d - \nabla D \cdot A^{-1} \cdot F_1)'$$

and White's $\hat{V}$ replaces this expectation by its sample analogue with θ set equal to $\hat{\theta}$, i.e. it replaces E by $n^{-1}\sum$ and θ by $\hat{\theta}$ throughout, an estimator we shall call $\hat{V}_1$. However, differentiating d we find

$$\begin{aligned}\nabla D &= E[F_3 + F_1 \otimes F_2 + F_2 \otimes F_1] \\ &= -E\big[(F_2 + F_1F_1')^c F_1'\big] \\ &= -E(dF_1')\end{aligned}$$

using (3). Thus, when F is correctly specified

$$\begin{aligned}V &= E(dd') + \nabla D \cdot A^{-1}E(F_1F_1')A^{-1}\nabla D' - \nabla D \cdot A^{-1}E(F_1d') \\ &\quad - E(dF_1')A^{-1}\nabla D' \\ &= E(dd') - E(dF_1')[E(F_1F_1')]^{-1}E(F_1d')\end{aligned}$$

which is consistently estimated under the assumptions of [2] by replacing E by $n^{-1}\sum$ and θ by $\hat{\theta}$. This estimator, $\hat{V}_2$, can be recognized as the inverse of the upper left submatrix of the inverse of the nonnegative definite matrix

$$n^{-1}(\hat{Y}'\hat{Y})$$

where $\hat{Y} = (\hat{Y}_1\ \hat{Y}_2)$, of order $n \times (q + p)$, and $\hat{Y}_1$ has rows of the form $\hat{d}'$, $\hat{Y}_2$ rows of the form $\hat{F}_1'$, and the hat indicates calculation at $\theta = \hat{\theta}$. The estimator is thus nonsingular if $\hat{Y}'\hat{Y}$ is, and does not require analytic third derivatives of the log likelihood function. With

$$\hat{V}_2 = n^{-1}\big(\hat{Y}_1'\hat{Y}_1 - \hat{Y}_1'\hat{Y}_2(\hat{Y}_2'\hat{Y}_2)^{-1}\hat{Y}_2'\hat{Y}_1\big)$$

and since $\hat{D}_n' = n^{-1}\iota'\hat{Y}_1$ where ι is a column of n ones the information matrix statistic, (4), with $\hat{V}_2$ as covariance matrix estimator is[4]

$$\mathscr{I}_n = \iota'\hat{Y}_1\big(\hat{Y}_1'\hat{Y}_1 - \hat{Y}_1'\hat{Y}_2(\hat{Y}_2'\hat{Y}_2)^{-1}\hat{Y}_2'\hat{Y}_1\big)^{-1}\hat{Y}_1'\iota. \tag{5}$$

An alternative representation of $\mathscr{I}_n$ exploits the fact that, from the likelihood equations, $\iota'\hat{Y}_2 = 0$ which implies that $\mathscr{I}_n$ of (5) is also given by the more concise expression

$$\mathscr{I}_n = \iota'\hat{Y}(\hat{Y}'\hat{Y})^{-1}\hat{Y}'\iota. \tag{6}$$

Since $\iota'\iota = n$ this expression can be recognized as n times an R^2 statistic in the regression of ι on $\hat{Y}$, i.e of unity on $\hat{d}'$ and $\hat{F}_1'$, but in which the total sum of squares is taken around the origin rather than the mean.

[4]For a test based on fewer than all $q = p(p + 1)/2$ elements of d one deletes columns of $\hat{Y}_1$ and corresponding rows and columns of $\hat{V}_2$.

This note was prompted by Chesher's [1] elegant demonstration[5] that the information matrix test is a score test of model specification against the alternative of local random parameter heterogeneity.

University of Hull

Manuscript received April, 1983; revision received July, 1983.

[5] A revised version under the same title is contained in this issue of *Econometrica*, 52(1984), 865–872.

REFERENCES

[1] CHESHER, A.: "Testing for Neglected Heterogeneity," University of Birmingham Discussion Paper, A282, 1982.
[2] WHITE, H.: "Maximum Likelihood Estimation of Misspecified Models," *Econometrica*, 50(1982), 1–25.