9-1972

# The Covariance Matrix of the Limited Information Estimator and the Identification Test: Comment

Franklin M. Fisher
*Massachusetts Institute of Technology*

Joseph B. Kadane
*Carnegie Mellon University*, kadane@stat.cmu.edu

*Econometrica*, Vol. 40, No. 5 (September, 1972)

# THE COVARIANCE MATRIX OF THE LIMITED INFORMATION ESTIMATOR AND THE IDENTIFICATION TEST: COMMENT

BY FRANKLIN M. FISHER AND JOSEPH B. KADANE

IN THEIR ARTICLE [**5**], Liu and Breen propose a new estimator of the large-sample asymptotic covariance matrix for the limited information maximum likelihood estimator in simultaneous equations, and express surprise that their estimator is different from the estimator proposed by Chernoff and Divinsky [**1**]. Additionally, they question the interpretation of a statistic used in the past to test overidentifying restrictions.

## 1. ESTIMATORS OF THE LARGE-SAMPLE ASYMPTOTIC COVARIANCE MATRIX

Suppose that $\theta$ is a $1 \times K$ vector of parameters to be estimated, and that $\hat{\theta}_n$ is an estimator of $\theta$ based on $n$ observations. If $\hat{\theta}_n$ is asymptotically unbiased and normal, $\sqrt{n}(\hat{\theta}_n - \theta)c^{-\frac{1}{2}}(\theta)$ converges to a $K$-dimensional unit normal distribution, where $c(\theta)$ depends, in general, on $\theta$. Also $\sqrt{n}(\hat{\theta}_n - \theta)c^{-\frac{1}{2}}(\hat{\theta}_n)$ converges to a $K$-dimensional unit normal distribution. If $\theta_n$ is asymptotically efficient (in the sense that the variance-covariance matrix attains the Cramer-Rao lower bound), one $c(\theta)$ that works can be computed as

$$c(\theta) = \left\{-E_\theta\left[\frac{\partial^2 L}{\partial \theta^2}\right]\right\}^{-1}$$

where $L$ is the log-likelihood function of the parameters $\theta$ given the observations, and the matrix in square brackets is its Hessian.

Liu and Breen correctly compute $c(\theta)$ and $c(\hat{\theta}_n)$. However, they wrongly interpret their results as casting doubt on the results of Chernoff and Divinsky. The Chernoff-Divinsky and Liu-Breen covariance matrices can both be correct, even though they differ.

For example, the sample sum of squares $s^2 = \Sigma(x_i - \bar{x})^2$ is well known to be a sufficient statistic for the variance of a normal distribution. We know $s_1^2 = s^2/(n-1)$ is unbiased; $s_2^2 = s^2/n$ is maximum likelihood; and $s_3^2 = s^2/(n+1)$ is uniformly minimum mean-squared error in the class of estimators of the form $ks^2$. However, $[\sqrt{n}(\bar{x} - \mu)/s_i]$ has an asymptotically unit normal distribution for each estimator $i = 1, 2, 3$.

There is no a priori reason, then, for Liu and Breen to say of their estimator "This result should be identical with, but is in fact different from" the Chernoff-Divinsky estimator. They may differ at each finite sample size, but still both have the requisite large-sample behavior.

We do not find computational simplicity a very appealing argument in this instance, since in either case, the computations would most likely be done by computer. Nor do we find compatibility with Theil's formula for the variance of a $k$-class estimator persuasive. A choice between the Liu-Breen estimator and the Chernoff-Divinsky estimator should be made on the basis of their statistical properties.

## 2. TESTING OVERIDENTIFYING RESTRICTIONS

In the second part of their paper, Liu and Breen criticize the standard likelihood ratio test of overidentifying restrictions. They claim that the test does not, in fact, test the correctness of exclusion restrictions on the jointly dependent variables, because in both the null hypothesis (that the restrictions are correct) and the alternative hypothesis, the same jointly dependent variables are excluded in computing the maximum value of the likelihood function. We shall show that this is not correct, first by reviewing the derivation of the likelihood ratio test, and second by examining what alternative hypotheses the test is consistent against.

For both the null ($i = 1$) and alternative ($i = 2$) hypotheses, consider exclusion restrictions on the first equation of a system in which the other equations are unrestricted. Then, under each model, the concentrated log-likelihood function is well known to be of the form

$$(1) \qquad L^{(i)} = k' + \frac{T}{2} \log l_i,$$

where $k'$ is a constant, $T$ is the number of observations, and $l_i$ is the smallest root of the determinental equation

$$(2) \qquad |W^*_{\Delta\Delta} - lW_{\Delta\Delta}| = 0.$$

Here, $W_{\Delta\Delta}$ is the sample covariance matrix of the residuals of the least squares regressions of the included endogenous variables, $y_\Delta$, on all the predetermined variables, and $W^*_{\Delta\Delta}$ is the similar covariance matrix when the regressions are taken only on included predetermined variables, $z_*$.

It is well known that $l_i$ is equal to one if the equation is just identified or underidentified. Therefore, to test a particular overidentified model as the null hypothesis ($i = 1$) against a particular just identified or underidentified model ($i = 2$) as the alternative hypothesis, $l_1$ is equivalent to the likelihood ratio statistic. Therefore, also, to test a particular overidentified model as the null hypothesis against the alternative of the set of all just identified or underidentified models, the likelihood ratio statistic is again $l_1$. This alternative hypothesis includes: (i) the general equation with no restrictions which is underidentified; and (ii) some restrictions on the endogenous variables but none on the exogenous variables, which is again underidentified. In the particular case stressed by Liu and Breen, the exclusion restrictions for the alternative are the same as those for the null hypothesis.

The consistency of the test based on $l_1$ depends on the probability limit of $l_1$, which turns on the question of whether the equations of the model themselves imply the existence of an equation satisfying the restrictions to be tested, that is, an equation which gives a linear combination of $y_\Delta$ and $z_*$ as equal to a disturbance uncorrelated in the probability limit with any of the predetermined variables. If there is such an equation, $l_1$ will have a probability limit of unity; otherwise, its probability limit will be greater than unity.[1] Another way of putting this is to say that $l_1$ will have a unity probability limit if and only if the model implies the existence of some linear combination of the elements of $y_\Delta$ and $z_*$ which has asymptotically the same residual variance when regressed on the elements of $z_*$ as when regressed on all the predetermined variables. If (and only if) it does, $(W^*_{\Delta\Delta} - W_{\Delta\Delta})$ is asymptotically singular.

Now, if the null hypothesis to be tested is correct, there obviously exists an equation in the form described; it is the equation whose specification is being tested. Under the null hypothesis, therefore, plim $l_1 = 1$.

What of the case in which the null hypothesis is false? Here there are two possibilities. If, even with the restrictions to be tested actually true, the equation in question is just identified or underidentified, then there will always be a way to get from the reduced form of the model to an equation in the crucial form. In this case, $l_1$ will be unity even in the sample and, as is well known, the test fails. As one should expect, one cannot test the truth of identifying restrictions from the data.

On the other hand, if the equation in question is *over*identified when the restrictions are true, then unrestricted reduced forms will *not* be compatible with an equation in the crucial form. If the restrictions are false, therefore, it will not be possible, even in the probability limit, to manipulate the model to obtain a linear expression in $y_\Delta$ and $z_*$ which is equal to a disturbance term only. Any such manipulation will end up with an expression which contains either some of the excluded endogenous or some of the excluded predetermined variables, or both. Hence any linear combination of the elements of $y_\Delta$ and $z_*$ will have a different (smaller) residual variance when regressed on all the predetermined variables than when regressed on the included ones only, and the probability limit of $l_1$ will exceed unity.

Observe that what everything depends on is the truth or falsity of the restrictions to be tested and their identifying power if they are true. The question of whether the equation in question is "really" under-, just, or overidentified by some other set of unknown restrictions is

[1] See Fisher [3, pp. 185–191] for an extended discussion of this.

irrelevant to the behavior of $l$, as it should be, since identification by *unknown* constraints is a contradiction in terms.

Note that all the above analysis is true despite the fact that the matrices in (2) are defined using residuals from regressions in which the excluded endogenous variables do not appear. The exclusion of those variables still plays a crucial role because it determines what residuals are to be used. Perhaps the most illuminating way of looking at the matter in the present context, however, is to observe that if an endogenous variable has been wrongly excluded from the equation to be tested, then, since that variable in the reduced form is a linear function of the disturbances and of *all* the predetermined variables, one can also say, by substitution, that there are predetermined variables wrongly excluded from the equation in a sense directly relevant to the present test.

Again, if one considers the specification error as placing the wrongly excluded variables in the disturbance term, it is easy to see that the test will (in principle) reveal this because the resulting disturbance term will have a smaller residual variance when regressed on all the predetermined variables than when regressed on the included ones only. Whether the excluded predetermined variables affect that disturbance term through being themselves wrongly left out of the equation and thus directly included in the disturbance, or whether they do so by affecting endogenous variables wrongly left out and so included, makes no difference. In either case, the test will reject the null hypothesis that the overidentifying restrictions are true in favor of the alternative that they are false.

However, this is not to deny that there are difficulties with this test: as stressed by Christ [**2**, pp. 539–40],[2] rejection of the null hypothesis does not say which of the restrictions are false or *what* model should be regarded as the new maintained hypothesis. Finally, there is the difficulty, shared by many tests of hypotheses, that the null hypothesis is embedded in, and of lower dimension than the alternative. Thus, if a smooth prior is placed on the parameter space, the null hypothesis would have probability zero. Put another way, the overidentifying restrictions are, at best, approximately true only. Thus, with sufficient data, a consistent test of them will always lead to rejection.

## 3. CONCLUSION

We find the Liu-Breen estimator of the asymptotic covariance matrix of LISE interesting, although their criticism of Chernoff and Divinsky's paper is unfounded. We find their criticism of the specification test of Koopmans and Hood [**4**] unwarranted.

*Massachusetts Institute of Technology*
*and*
*Carnegie-Mellon University*

*Manuscript received May, 1970; revision received October, 1970.*

## REFERENCES

[**1**] CHERNOFF, HERMAN, AND NATHAN DIVINSKY: "The Computation of Maximum Likelihood Estimates of Linear Structural Equations," in W. C. Hood and T. C. Koopmans, eds., *Studies in Econometric Method*, Cowles Commission Monograph No. 14. New York: Wiley, 1953, pp. 236–302.

[**2**] CHRIST, CARL F.: *Econometric Models and Methods*. New York: Wiley, 1966.

[**3**] FISHER, FRANKLIN M.: *The Identification Problem in Econometrics*. New York: McGraw-Hill, 1966.

[**4**] KOOPMANS, T. C., AND W. C. HOOD: "The Estimation of Simultaneous Linear Economic Relationships," in W. C. Hood and T. C. Koopmans, eds., *Studies in Econometric Method*, Cowles Commission Monograph No. 14. New York: Wiley, 1953, pp. 112–199.

[**5**] LIU, T. C., AND W. J. BREEN: "The Covariance Matrix of the Limited Information Estimator and the Identification Test," *Econometrica*, 37 (1969), 222–227.

[2] See also Fisher [**3**, p. 185].