COWLES COMMISSION DISCUSSION PAPER: STATISTICS: No. 341

November 23, 1949

A Note on Limited Information Estimates and Methods for their Appraisal[1]

by Carl Christ

As indicated in a previous note on models and identification (Cowles Commission Discussion Paper, Statistics, No. 340), consistent estimates can be obtained for the parameters of an identified model if the model is a parametric family of structures, and if the true structure belongs to the model. These estimates are obtained by maximizing a likelihood function of parameters and variables, subject to certain restrictions implied in the model.

The efficiency of the estimates is an increasing function of the amount of a priori information about the model which is used in obtaining the estimates. If so little information is used that the structures within the model are not identified, then the efficiency of the estimates becomes zero.

When estimates of all the parameters of the model are obtained simultaneously by maximizing their joint likelihood function subject to all the a priori restrictions which are implied in the model, then under certain assumptions[2] the estimates are consistent and efficient. Estimates so obtained are called full maximum likelihood estimates, here abbreviated to full max llh. They are the best estimates we know how to get, precisely because they use all of the a priori information which we assume we have concerning the structure. They are relatively expensive to compute, however, and so sometimes it is desired to sacrifice efficiency in favor of economy.

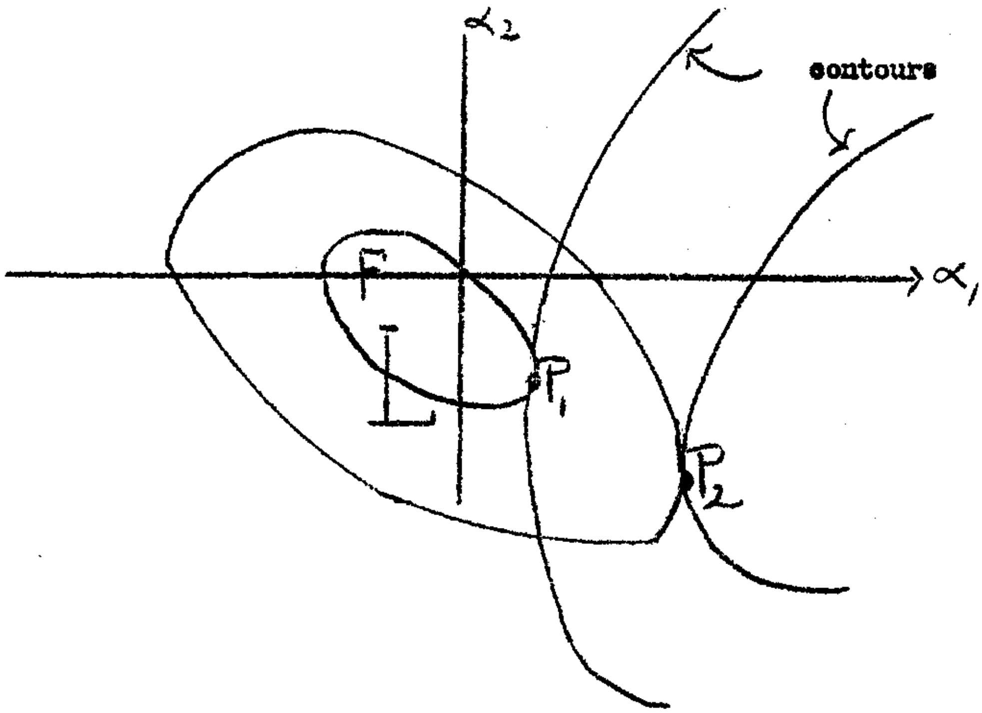
---

1. This note draws heavily on discussions with Leo Hurwicz.

2. The disturbances must be normally distributed with means zero and variances constant, and serially uncorrelated. The equations must be linear in the parameters.

The limited information single equation maximum likelihood method, here abbreviated to LISE, yields consistent estimates under certain conditions[2] because it is based on the maximization of a likelihood function. It is more economical than full max llh, but it yields less efficient estimates because it does not use all the a priori restrictions on the likelihood function. It estimates the parameters of one equation at a time, and in estimating the parameters of a given equation it neglects all a priori information about the remainder of the model except for knowledge of the existence and the values of at least a certain number, M, of the predetermined variables which appear in the rest of the model but do not appear in the equation to be estimated (M + 1 is equal to the number of jointly dependent variables appearing in the equation to be estimated; this condition is necessary, and usually sufficient, for identification).

The full max llh and the LISE methods can be compared graphically by considering the two regions in the parameter space which correspond to the restrictions used by the two methods. In two dimensions, the situation is essentially as shown, where the large region L contains the points in the parameter space which are permissible according to the a priori restrictions used by LISE, and the smaller region F contains the points which are still permissible if <u>all</u> the a priori restrictions implied in the model are used. Suppose the likelihood function has contours as shown. Then the maximization of the likelihood function leads to the full max llh estimates given by the coordinates of P_1 and the LISE estimates given by P_2. If the true point is within the region F, i.e., if the true structure belongs to the model, then it is clear that the variances of the full max llh estimates P_1 will usually be smaller than those of the LISE estimates P_2, and can never be larger.



The remainder of this note is concerned with methods of appraising a set of LISE estimates, i.e., a particular structure, obtained for a model which is a parametric family of structures and which is supposed to represent time series in an economic system, for example the U. S. economy. There are three stages at which the evaluation may take place: (1) after the model has been formulated (i.e., after the a priori restrictions have been set up) but before any estimation has been done; (2) after estimates have been obtained but before predictions based on the estimates have been checked against observations; (3) after predictions have been checked. There are also three aspects of the structure upon which evaluation may be based: (a) the character of the a priori restrictions; (b) the estimated values of the parameters; (c) the values of the disturbances calculated from the observations and the estimated parameters. Thus there is a nine-chotomy of methods of evaluation, which can be represented in a 3 x 3 table thus:

	(1) before estimation	(2) after estimation	(3) after checking predictions
(a) restrictions		*	*
(b) estimates	*		
(c) disturbances	*		

The spaces containing asterisks can be neglected, because everything that can be discovered by merely examining the a priori restrictions can be discovered before estimation, and nothing can be discovered from estimates and calculated disturbances until after estimation. Therefore there are five situations to discuss: (1a), (2b), (2c), (3b), and (3c). They will be taken up in that order.

(1a) Character of restrictions. This type of evaluation consists chiefly of examining the model to see whether it makes good economic sense or not: of taking it apart, so to speak, to see what it implies about economic behavior, what inconsistencies can be found in it, what kinds of things might invalidate certain of its restrictions, etc.

(2) After estimation, (b) estimates. Several possibilities exist here. First, the estimates can be examined to see whether they have the approximate magnitudes and particularly the algebraic signs which are expected on the basis of theoretical and other information about elasticities, marginal propensities, etc. Second, the estimated sampling variance of each estimate can be examined to see how much confidence can be placed in its sign or in its approximate size. For instance, an estimate having an a priori wrong sign is a red flag according as its own estimated variance is small in comparison.

Third, for any equation of the model a test can be applied to the largest characteristic root λ_1 of the following equation, which is used to obtain the

LISE estimates of the given equation of the model:

$$\det \left(W(W^{*} - W)^{-1} - \lambda I \right) = 0$$

Here W is the covariance matrix for the disturbances to the regressions of the jointly dependent variables appearing in the given equation of the model on the predetermined variables assumed to be known to appear in the entire model; W* is the covariance matrix for the disturbances to the regressions of the jointly dependent variables appearing in the given equation of the model on the predetermined variables appearing in the given equation of the model; λ is a scalar, and I is the identity matrix. Anderson and Rubin have shown that under the assumptions of the LISE method the quantity $T \log \left(1 + \frac{1}{\lambda_1} \right)$ has the χ^2 distribution asymptotically as the sample size T increases, with the number of degrees of freedom equal to the number of overidentifying restrictions (in the form of predetermined variables excluded) which exist on the given equation of the model. $1 + \frac{1}{\lambda_1}$ can never be less than 1, and if it is close to 1 in an overidentified model it means that the effect of excluding the excluded predetermined variables is only slightly detrimental to the variances, i.e., increases then only slightly, which is what we want. This χ^2 test of the largest root λ_1 is a sort of overall test of the totality of restrictions and assumptions applied to the model; if λ_1 takes a value which is very improbable under the hypothesis that all these assumptions are true, then we have only a very generalized alarm signal which cannot point to a specific remedy. The test is limited in usefulness by two additional things: its power function is unknown so that a "favorable" verdict has an uncertain meaning, and the distribution of $T \log \left(1 + \frac{1}{\lambda_1}\right)$ is not known for small samples of the size ordinarily used.

Fourth, the estimates can be compared with least squares estimates of the same equations, although it is not clear just what can be learned about the LISE estimates by this process - it is possible to construct examples in which the LISE and least squares estimates are close together, and examples in which they are far apart. The LISE estimates can also be compared with the full max llh estimates.

(2) After estimation, (c) disturbances. The calculated disturbances for the sample period can be tested to see whether they are random, not autocorrelated. If they show significant autocorrelation, the structure is suspect because it is obtained on the assumption of non-autocorrelated disturbances and because there is systematic variation which it does not explain.

The disturbances can be examined to see whether they are very large according to some intuitive standard of how large they are expected to be (for example, we may believe that the disturbance to a good production function in a specific industry will usually be a small percentage of output in that industry, and that the disturbance to the corresponding investment equation will often be a larger percentage of investment). But this method is of doubtful usefulness, because it is not always possible to tell whether disturbances are due to the existence of several systematic factors which have been neglected, or to a real randomness in the phenomenon studied, especially if the disturbances appear to be random.

The disturbances can be broken into two groups, and the hypothesis can be tested that the variance of the disturbances in one group is the same as that in the other. This affords a test of one of the assumptions underlying both the LISE and the least squares estimates. The groups can be high-income years and low-income years, or increasing-income years and decreasing-income years, or years before a certain date and years after, or cetera.

(3) After prediction, (b) estimates. In this category the only thing that might be done is to recompute the estimates for a new sample including the period

for which predictions are made, and see what kind of a change occurs in the estimates as a result.

(3) After prediction, (c) disturbances. First, the calculated disturbances to the reduced-form equations for the prediction period can be found; these are the errors of prediction, measuring the differences between actual and predicted values of the jointly dependent variables. Tolerance intervals can be prepared for these calculated disturbances, using estimates of their variances based on the assumption that the estimated structure will be as good an explanation of the future as it was of the sample period. If a calculated disturbance falls outside its tolerance interval, we believe that this assumption is not true, but again as in the case of the characteristic root test we have only a generalized symptom of trouble. If all calculated disturbances fall inside their tolerance intervals, this may mean either (1) that we have a good structure or (2) that we have a terrible structure which however is not worse in the prediction period than it was in the sample period. (This remark arose in discussions with H. Markowitz.)

(It should further be stated that since the limited information method, for a model which is overidentified and linear in parameters but not in variables, provides as many reduced-form predictions of a given jointly dependent variable as there are equations containing that variable in the model, there will probably be ambiguities in any LISE procedure involving the reduced form of such a model. They might be avoided, however, by solving the estimated structure directly to obtain the reduced form.)

Second, the same procedure can be applied to the calculated disturbances of the equations of the model (structural equations). This is what I have called the ks* test in Discussion Paper 269. It avoids two objections inherent in the application to the reduced form case, because it localizes trouble to the structural equation(s) at fault, and it gives exactly one answer for each equation. But it is still subject to the difficulty that it cannot tell a good equation from a bad one which has merely not gotten worse.

Third, the calculated disturbances to the reduced form for the prediction period can be compared with the errors of prediction made by other methods, such as full max llh, or the least squares method, or the naive models which simply extrapolate y_{t-1} or $y_{t-1} + \Delta y_{t-1}$, or the reduced form obtained by solving the structure as estimated by least squares.