A NOTE ON SOME RECENT DEVELOPMENTS OF ESTIMATING AN OVER-IDENTIFIED SIMULTANEOUS EQUATIONS MODEL BY FIML

by **NICHOLAS C. BALTAS,** (Ph.D.), Agricultural Bank of Greece

1. Introduction

A number of alternative methods can be employed to estimate the parameters in a simultaneous equations model. These methods can be classified into two main catergories: the single-equation or limited information and the full information one. The methods of the first category utilize only the a priori information on restrictions, which apply to any particular structural equation. The full information methods on the other hand are more flexible in the sense that they can be adapted to use information relating to other structural equations as well. The full information maximum likelihood method (FIML) belongs to the second category.

The FIML method has a number of desirable properties like consistency and asymptotic efficiency with a variance equal to the lower bound of the Grammer-Rao inequality provided that the disturbances are normally distributed. The FIML estimates are best asymptotic under conditions including complete identification, stability and regularity (see Sargan (28)). In the case where all equations of a model are just-identified the maximum likelihood estimates can be obtained either directly from the likelihood estimators corresponding to the parameters of a structural model or indirectly from the maximum likelihood of its reduced from. However, for simplicity we can apply the inderect least-squares method to obtain FIML estimators of the parameters with all the desirable properties. If any equation of a model is over-identified, then none of these methods can be used. In this paper the problems of deriving FIML estimators are discussed, when the simultaneous equations model is over-identified. In addition to this, a critical review of literature is presented related to derivation difficulties and with possible ways to overcome them. Finally in a concluding section, the disadvantages of the FIML method are summarized and hints for further research are suggested.

2. The FIML estimator

Consider the structural linear form for all K jointly dependent variables of the model.

$$Y\Gamma + XB + U = 0 \quad . \tag{1}$$

and $$U \sim N(O, \Phi \times I), \; EU'U = \Phi \times I, \; E(X'U)=0, \; r(X)=M \tag{2}$$

where Y is a (T×K) matrix of sample observations of the jointly dependent variables, Γ is a (K×K) non-singular matrix of their coefficient, X is a (T×M) matrix of predetermined variables, B is a (M×K) matrix of their coefficients and U is a (T×K) matrix of disturbances. It is assumed that the variance-covariance matrix Φ (K×K) is non-singular[1], positive definite. The relations (2) mean that, for any time t, the joint distribution of the error terms of the system is normal with mean zero and covariance Φ. The error terms, which are mutually independent and identically distributed as a non-singular K-variate normal distribution, may be contemporaneously correlated but are intertemporally uncorrelated. They also imply that the error terms are uncorrelated with the predeterminated variables of the system, and that no linear dependence exists among the predetermined variables; and the second-order moment matrices of the current predetermined and endogenous variables are assumed to have non-singular probability limits. The non-singularity of Φ is assumed in order to give unique solution of current endogenous variables in terms of the predetermined variables of the system. Also for the computation of FIML estimators it is necessary to suppose that the number of observations T exceed the number of exogenous variables M and the number of endogenous variables K i.e., all equations satisfy the rank condition for identification and that the system is stable if lagged endogenous variables are included as predetermined variables. The derived reduced form of the system is given by:

$$Y = X\Pi + V \text{ where } \Pi = -B\Gamma^{-1}, \; V = -U\Gamma^{-1} \tag{3}$$

The variance-covariance of the reduced form Σ is also assumed to be non-singular

$$\begin{aligned} E(V'V) &= \Sigma \\ &= E(\Gamma'^{-1} U'U\Gamma^{-1}) \\ &= \Gamma'^{-1} E(U'U)\Gamma^{-1} \\ &= \Gamma'^{-1} (\Phi \times I)\Gamma^{-1} \end{aligned} \tag{4}$$

1. The assumption that Φ is nonsingular implies that the system (1) contains no identities. It means that all identities of the system have been solved out.

Now, consider the likelihood function corresponding to the reduced form[2] equations

$$£ = (2\pi)^{-\frac{KT}{2}} |\Sigma|^{-\frac{T}{2}} \exp\left(-\frac{1}{2} \sum_{t=1}^{T} v_t \Sigma^{-1} v_t'\right) \quad (5)$$

Since we are interested, in deriving the estimates of the set of structural parameteres, it is approapriate to express the reduced form parameters matrices V and Σ in terms of B, Γ and Φ.

The likelihood of the typical elements of endogenous variables given the x's is provided by

$$P(y_t \mid x_t) = P(v_t \mid x_t) \left| \frac{\partial v_t}{\partial y_t} \right| \quad (6)$$

where $|\partial v_t \mid \partial y_t|$ is the absolute value of the Jacobian of the transformation of the following matrix

$$\begin{bmatrix} \frac{\partial v_{1t}}{\partial y_{1t}} & \frac{\partial v_{1t}}{\partial y_{2t}} & \cdots\cdots & \frac{\partial v_{1t}}{\partial y_{kt}} \\ \frac{\partial v_{2t}}{\partial y_{1t}} & \frac{\partial v_{2t}}{\partial y_{2t}} & \cdots\cdots & \frac{\partial v_{2t}}{\partial y_{kt}} \\ \frac{\partial v_{kt}}{\partial y_{1t}} & \frac{\partial v_{kt}}{\partial y_{2t}} & \cdots\cdots & \frac{\partial v_{kt}}{\partial y_{kt}} \end{bmatrix} \quad (7)$$

the determinant of which is equal to $|\Gamma|^T$. So equation (6) can be rewritten as,

$$P(y_t \mid x_t) = |\Gamma|^T P(v_1) P(v_2) \ldots\ldots P(v_T) \quad (8)$$

Substituting (1) into (5) and taking (8) into consideration, the joint density of Y_t is:

$$£ = (2\pi)^{-\frac{KT}{2}} |\Gamma|^T \Phi^{-\frac{T}{2}}$$

$$\exp.\left\{-\frac{1}{2} \sum_{t=1}^{T} (y_t \Gamma + x_t B) \Phi^{-1} (y_t \Gamma + x_t B)'\right\} \quad (9)$$

2. In the estimation of the reduced form it might be better to avoid total commitment to exactly restricted reduced form models. Instead we propose to estimate a model, which is only stochastically dependent on the structural restrictions, a procedure that represents the degree of uncertainty concerning the structural restrictions.

since $|\Sigma| = |\Gamma|^{-2} |\Phi|$, $|\Sigma|^{-\frac{T}{2}} = |\Gamma|^{T} |\Phi|^{-\frac{T}{2}}$ and

$$-\frac{1}{2} \sum_{t=1}^{T} v_t' \Sigma^{-1} v_t' = -\frac{1}{2} (Y\Gamma+XB) (\Phi^{-1} \times I) (Y\Gamma+XB)' =$$

$$= -\frac{1}{2} \sum_{t=1}^{T} (y_t \Gamma + x_t B) \Phi^{-1} (y_t\Gamma + x_t B)'$$

To simplify the notation we will adopt the following equivalences:

$$A = \binom{\Gamma}{B} \qquad Z = (Y \quad X)$$

Using this notation we may write the logarithm of the likelihood function

$$L(B,\Gamma,\Phi;Z) = \log \pounds = -\frac{KT}{2} \log(2\pi) + T\log|\Gamma| - \frac{T}{2} \log|\Phi| - \frac{1}{2} \sum_{t=1}^{T} z_t A\Phi^{-1} A'z_t'$$

or $$L(B,\Gamma,\Phi;Z) = -\frac{KT}{2} \log(2\pi) + T\log|\Gamma| - \frac{T}{2} \log|\Phi| - -\frac{T}{2} \mathrm{tr}\ \Phi^{-1} (A'\Lambda A) \quad (10)$$

since $$\sum_{t=1}^{T} z_t A\Phi^{-1} A'z_t' = \mathrm{tr}\ \Phi^{-1} A'Z'ZA = T\mathrm{tr}\ \Phi^{-1} A'\Lambda A$$

where $$\Lambda = \frac{1}{T} Z'Z = \begin{bmatrix} m_{yy} & m_{yx} \\ m_{xy} & m_{xx} \end{bmatrix}$$

is the sample variance matrix. Both the determinant of Γ and the last term in (10) are nonlinear functions of the unknown parameters and consequently the corresponding set of first-order conditions are nonlinear too.

Since no restrictions have been placed on the elements of Φ, we can maximize[3] L with respect to the unknown elements of B, Γ and Φ. First, we obtain an estimator for Φ in terms of the unknown parameters A. Substituting this maximizing value of

---

3. Hausman (17, 18) follows a different procedure: he concentrates on the presence of the Jacobian $|\Gamma|$ in the likelihood function, which differentiates the simultaneous equations problem from the Zellner (29) multivariate least squares problem. The reason is that if the Jacobian of the transformation from V to Y, $\partial V/\partial Y$, were an identity matrix, the FIML estimator would be the generalized least squares estimators. Then he maximizes the likelihood with respect to A, Γ and Σ and sets the partial derivatives equal to zero.

Φ in (10) we obtain a likelihood function in terms of A and Γ. Finally we maximize this function with respect to A and Γ. Proceeding the estimation we take:

$$\frac{\partial L}{\partial \Phi} = -\frac{T}{2}\ \Phi^{-1} + \frac{T}{2}\ \Phi^{-1}\ A'\Lambda A\ \Phi^{-1} = 0 \qquad (11)$$

$$\text{and } \hat{\Phi} = A'\Lambda A$$

$$\text{because } \frac{\log|\Phi|}{\partial \Phi} = \Phi^{-1}, \quad \frac{\partial \mathrm{tr}\Phi^{-1}\ A'\Lambda A}{\partial \Phi} = -\Phi^{-1}\ A'\Lambda A\ \Phi^{-1}$$

inserting $\hat{\Phi}$ in (10) we obtain the reduced likelihood as a function of B and Γ

$$L(B,\Gamma;Z) = -\frac{KT}{2}\ \log(2\pi) + T\log|\Gamma| - \frac{T}{2}\log|A'\Lambda A| - \frac{KT}{2} \qquad (12)$$

and $$L(B,\Gamma;Z) = -\frac{KT}{2}\ [1+\log(2\pi)] + T\log|\Gamma| - \frac{T}{2}\log\ |A'\Lambda A| \qquad (13)$$

Now equation (13) must be maximized with respect to the unknown parameters B and Γ. It is noted that by differentiating, the estimating equations obtained are highly non-linear in the unknown parameters. For instance, differentiating with respect to γ_{ij} and putting the derivatives equal to zero, we get,

$$\frac{\partial L}{\partial \gamma_{ij}} = \frac{T}{|\Gamma|} \cdot \frac{\partial|\Gamma|}{\partial \gamma_{ij}} - \frac{T}{2\,|A'\Lambda A|} \cdot \frac{\partial\,|A'\Lambda A|}{\partial \gamma_{ij}} = 0 \qquad (14)$$

Since $|\Gamma|$ is a function of the coefficient of endogenous variables in all equations, equalizing the derivative to zero, we obtain a system of non-linear equations in the unknown coefficients.[4] For solving such non-linear equations, the use of an iterative method becomes indispensable.

3. Iterative methods

In the literature there are several iterative methods currently available. H. Chernoff and N. Divinsky (6) and H. Eisenpress and J. Greenstadt (11) have proposed suitable algorithms for computing the FIML estimators. Chow (7) presented two iterative methods of estimation of parameters in a system of simultaneous linear stochastic equations.[5] The first iterative approach (direct method) is almost identical

---

4. For example, for the coefficient γ^{ij}, $T/|\Gamma| \cdot \partial|\Gamma| / \partial\gamma_{ij} = \gamma^{ij}$ is the (i, j=1,2,..., T) element of Γ^{-1}, which involves all elements of Γ in a non-linear fashion.

5. A somewhat different interepretation of interdependent equations of a simultaneous equations model and an iterative method of estimation known as "fix point" is described in E.J. Mosback and H. Wold (23).

with the gradient method proposed by Chernoff and Divinsky. The alternative technique is based on Newton's method in which the values of coefficients converge to the correct values very rapidly, aw compared to the direct method. The drawback of this method lies not in its effectiveness as an oprimizer, but rather in the excessive cost required in the calculation of gradient and Hessian matrix. Other methods such as the Davidon (9), Fletcher-Powell (14) e.t.c. have been developed to reduce these costs. This can be obtained by providing to the latter algorithms good initial approximations to the Hessian matrix. An alternative method described as the linearized maximum likelihood method, which has the same optimal large-sample properties, as long as the a priori information is correct, has been suggested by T.J. Rothenberg and C.T. Leenders (27). Although this method is much simpler than the FIML, it is computationally more troublesome than the three-stage squares method, which has the same asymptotic properties.

Lyttkens (22), Dhrymes (10), and Brundy and Jorgenson (5) proposed full information instrumental variables (FIIV) estimators, which are also consistent and asymptotically efficient. Hausman (17) showed that the FIIV estimators, are particular cases of the basic FIML iteration, when it starts with consistent estimate. The instruments are identical, so that if these estimators are iterated and converge, the resulting estimate will be the FIML estimate. Hausman proposed an algorithm, which follows an analogous procedure to the Gauss-Newton algorithm for non-linear least squares. Since the Gauss-Newton algorithm has proved effective, the algorithm proposed by Hausman might also be effective in computing the FIML estimates. This approach yields the "uphill property", which ensures an increase in the likelihood function at each iteration, not only when a certain matrix is positive definite, but also when the matrix is not positive definite. Although this procedure has desirable asymptotic properties, its use in actual calculations remains to be evaluated.

Dagenais (8) developed a unified iterative generalized least squares approach for computing FIML estimates for both linear and non-linear equations models of medium size. The proposed algorithm is essentially a gradient method which, at each iteration, multiplies the gradient by a suitably chosen positive definite matrix. When applied to linear models, the basic procedure[6] resembles to that of Hausman (17).

An alternative estimation procedure has been proposed by Powell (26). This method[7] does not require any derivative in determining the unknown parameters of

6. In the nonlinear case, it is similar to both, that proposed by Bernd et.al. (3) and Amemiya's method (1), which is the nonlinear analogue of Hausman's procedure. In contrast with the linear case, it does not produce an asymptotically efficient second round estimator even if the initial estimator is consistent, but as in the linear case, it illustrates the similarity and the difference between the maximum likelihood and non-linear three-stage squares estimator.

7. It has also been extensively evaluated as a numerical method technique and appears to be among the most efficient computational methods presently available (see M. Box, D. Davies and W. Swann (4)).

an over-identified simultaneous equations model. The approach of estimating parameters, by the FIML method considers the likelihood function to be a general non-linear function of the unrestricted structural parameters, and maximizes it directly using conjugate directions of search method. This yields a ready approximation to the variance-covariance matrix. The Powell algorithm provides an explicit expression for the estimates of B and Γ. From this, we can derive the maximum likelihood estimates of the restricted reduced form, $P = -B\Gamma^{-1}$. The FIML estimators are equivalent to the so called least generalized residual variance estimators, which are obtained by minimizing the determinant of the variance-covariance matrix of the reduced form residuals.[8]

Since FIML is an iterative technique, some initial values must be given to start the iteration. Apparently, if those values are reasonably close to the real values, it might be expected that the iteration procedure can be made to converge faster and the computation time will be minimized. As such initial values will be used, those will be obtained from a single-equation methods like 2SLS, LIML etc.

In a very recent paper Fair and Parke (12) obtained full information estimates of a fairly large nonlinear model. This was made feasible due to a development of an algorithm by one of the authors (Parke (24)). This algorithm is fundamentally different from the algorithm used by others i.e., Dagenais (8). Like Powell's algorithm, it does not require any derivative of the likelihood function. It, instead, takes advantage of certain features of the model's structural equation. There are, however, some practical points that arise when trying to use the Hausman test in the present context, and the current application of the test has only been partly successful. This is likely due to small sample problems. It thus appears that more observations are needed before the Hausman tests can be applied with much confidence.

4. Concluding remarks

The FIML method of estimation presupposes and utilizes a priori information concerning the linearity of the complete system of equations. Estimators are obtained by maximizing this logarithmic likelihood function subject to all the a priori restrictions built into the model. In fact, more efficient estimates can be obtained as more valid restrictions are utilized.

Although the FIML estimators have many desirable statistical properties, they also possess a number of less desirable qualities. The large amount of computation required is one of the disadvantages. Generally the more computations are needed, the greater the chance of introducing, unwittingly, rounding-off errors into the answer. Another drawback is that so many assumptions have to be made to enable

8. See A. Goldberger, (15), pp. 352-356.

us to obtain the estimators sought. The normal distribution of the error term is a strict assumption.[9] There may be occasions on which this assumption appears too strong. What is somewhat more serious is the implicit assumption that the model is well specified, which means that it portrays reasonably the interrelation existing among the variables considered. With this notion we do not mean that the model is not linear, rather it has been mis-specified. That is, if we have wrongly specified even one parameter, then the estimates of all other parameters in the model are bound to be affected in some way.

Moreover, though many interesting alternative optimization algorithms have been mentioned in this paper, particularly the one proposed by Powell (mainly when the initial values are near to the optimum), the FIML method is generally time-consuming to compute the non-linear parameters of the model. Hence, the future research will have to be concentrated on less-consuming iterative methods, which would permit convenient full-information estimation of a simultaneous equations model. Moreover, the choice of a general procedure to calculate the FIML estimates will require further experimentation, especially in larger systems, for which almost no results have been reported.

REFERENCES

1. T. AMEMIYA, "The Maximum Likelihood and the Nonlinear Three-Stage Least Squares Estimator in the General Nonlinear Simultaneous Equation Model", *Econometrica*, Vol. 45, 1977, pp. 955-968.
2. N.C. BALTAS, *An Econometric Investigation of the Interrelationship between Capital Formation and Economic Growth of Greece*, An Unpublished Ph. D Thesis, University of Birmingham, 1974.
3. E.K. BERNDT, B.H. HALL, R.E. HALL and J.A. HAUSMAN, "Estimation and Inference in Nonlinear Structural Models", *Annals of Economic and Social Measurement*, Vol. 3/4, 1974, pp. 653-665.
4. M. BOX, D. DAVIES and S. SWANN, *Non-linear Optimization Techniques*, I.C.I., Monograph, No 5, Oliver and Boyd, London 1969.
5. J. BRUNDY and D.W. JORGENSON, "Efficient Estimation of Simultaneous Equation Systems by Instrumental Variables", *Rev. of Econ. and Stat.*, Vol. 53, 1971, pp. 207-224.
6. H. CHERNOFF and N. DIVINSKY, "The Computations of Maximum Likelihood Estimates of Linear Structural Equations", in *Studies in Econometric Methods*, W.C. Hood and T.C. Koopmans (Eds), New York, John Wiley, 1953.
7. G.C. CHOW, "Two Methods of Computing Full Information Maximum Likelihood Estimates in Simultaneous Stochastic Equations", *Inter. Econ. Rev.*, Vol. 9, No 1, Feb. 1968, pp. 100-112.

---

9. With another procedure, which does not require the normality assumption, we can obtain the "Gaussian estimators" of the unknown parameters. These are defined as the estimators derived by maximizing a function like (9), which would have been the likelihood for the observed sample if the errors had been normal varieties. Koppmans and others have named these quasi maximum likelihood estimators (see Fisk (13), p. 62).

8. M.G. DAGENAIS, "The Computation of FIML Estimates as Iterative Generalized Least Squares Estimates in Linear and Nonlinear Simultaneous Equations Models", *Econometrica,* Vol. 46, 1978, pp. 1351-1362.
9. W.C. DAVIDON, "Variable Metric Method for Minimization", A.E.C., Research and Development Report ANL5990, 1959.
10. P.J. DRYMES, "A Simplified Structural Estimator for Large-Scale Econometric Models", *The Australian Journal of Statistics,* 13, 1971, pp. 168-175.
11. H. EISENPRESS and J. GREENSTADT, "The Estimation of Nonlinear Econometric Systems", *Econometrica,* Vol. 34, 1966, pp. 851-861.
12. R.C. FAIR and W.R. PARKE, "Full-Information Estimates of a Nonlinear Macroeconometric Model", *Journal of Econometrics,* Vol. 13, 1980, pp. 269-291.
13. P.R. FISK, *Stochasticaly Dependent Equations: An introductory Text for Econometricians,* Griffin's Statistical Monographs and Courses, No 21, 1967.
14. R. FLETCHER and M.J.D. POWELL, "A Rapidly Convergent Descent Method for Minimization", *The Computer Journal,* VI July, 1963, pp. 163-168.
15. A.S. COLDBERGER, *Econometric Theory,* John Wiley and Sons, Inc. New York, 1964.
16. S.M. GOLDFELD and R.E. QUANDT, *Non-linear Methods in Econometrics,* North-Holland Publishin Co., Amsterdam 1972.
17. J.A. HAUSMAN, "Full Information Instrumental Variables Estimation of Simultaneous Equations Systems", *Annals of Economic and Social Measurement,* Vol. 3/4, 1974, pp. 641-652.
18. J.A. HAUSMAN "An Instrumental Variable Approach to Full-Information Estimators for Linear and Certain Non-Linear Econometric Models", *Econometrica,* 43, 1975, pp. 727-738.
19. D.F. HENDRY, "The Structure of Simultaneous Equations Estimators", *Journal of Econometrics,* No 4, 1976, pp. 51-88.
20. W.C. HOOD and T.C. KOOPMANS, *Studies in Econometrics Methods,* J. Wiley, New York, 1953.
21. J. JOHNSTON, *Econometric Methods,* McGraw-Hill, 1972 (Second Edition).
22. E. LYTTKENS, "Symmetric and Asymmetric Estimation Methods" in E. Mosback and H. Wold, *Interdependent Systems,* Amsterdam, 1970, pp. 434-459.
23. E. J. MOSBACK and H. WOLD, *Interdependent Systems, Structure and Estimation,* Amsterdam: North-Holland Publishing Company, 1970.
24. W.R. PARKE, "An algorithm for FIML and 3SLS estimation of large nonlinear models", (forthcoming).
25. M.J.D. POWELL, "An Efficient Method for Finding the Minimum of a Function of a Several Variables without Calculating Derivatives", *The Computer Journal,* Vol. 7, 1964, p. 155.
26. M.J.D. POWELL, "A Method for Minimizing a Sum of Squares of Non-linear Functions without Calculating Derivatives", *The Computer Journal,* Vol. 7, 1964, pp. 303-307.
27. T.J. ROTHENBERG and C.T. LEENDERS, "Efficient Estimation of Simultaneous Equations Systems", *Econometrica,* Vol. 32, 1964, pp. 57-76.
28. J.D. SARGAN, "Three Stage Least Squares and Full Maximum Likelihood Estimates", *Econometrica,* Vol. 32, No 1-2, 1964, pp. 77-81.
29. A. ZELLNER, "An Efficient Method of Estimating Semmingly Unrelated Regressions and Tests for Aggregation Bias", *Jour. of Amer. Stat. Assoc.,* Vol. 57, 1962, pp. 348-368.