working paper
department
of economics

AN INSTRUMENTAL VARIABLE APPROACH
TO FULL-INFORMATION ESTIMATORS
FOR LINEAR AND NON-LINEAR ECONOMETRIC MODELS

Jerry A. Hausman

Number 127 March 1974

massachusetts
institute of
technology
50 memorial drive
cambridge, mass. 02139

AN INSTRUMENTAL VARIABLE APPROACH TO FULL-INFORMATION ESTIMATORS FOR LINEAR AND NON-LINEAR ECONOMETRIC MODELS

Jerry A. Hausman

Number 127 March 1974

This is a revised version; the original was printed in January 1973.

1. INTRODUCTION

In this paper an investigation of asymptotically efficient estimators for linear and nonlinear simultaneous equation econometric models is undertaken. By using an instrumental variable approach the equivalence of previously proposed linear estimators to full information maximum likelihood (FIML) follows in a straightforward manner, and a class of new estimators which includes a nonlinear three stage least squares estimator (NL3SLS) and nonlinear full-information instrumental variables estimator are proposed and shown to be asymptotically equivalent to FIML.

First, an instrumental variable interpretation of FIML is developed by investigating the first order conditions for the maximum of the likelihood function without first concentrating the likelihood function. The essential difference between 3SLS and FIML then becomes evident. The difference between the two estimators is that FIML uses all over-identifying restrictions in forming the instruments while 3SLS ignores some of these restrictions. While this difference in forming the instruments is of no importance asymptotically as is known by the earlier results of Sargan [9] and Rothenberg and Leenders [8], in finite samples there seems no reason not to use all known prior information. The a priori restrictions give a more useful criterion than Dhrymes' [3] recent interpretation of a difference in 'purging' the endogenous variables since all other proposed estimators can be shown to be equivalent by simply proving asymptotic equivalence of the instruments used to those instruments used by FIML estimator.

The next result is to derive the necessary conditions on the number of observations to permit computation of the FIML estimate. The FIML estimate can be computed in a class of cases where all efficient limited information estimators such as two stage least squares (2SLS) and limited information maximum likelihood (LIML) are infeasible. 3SLS is also infeasible in this class of cases. The reason that the FIML estimate exists is again that all a priori restrictions are used while the other estimators neglect some over-identifying restrictions in forming the instruments. Thus a partial solution to the much studied problem of simultaneous equation estimation with undersized samples is given. Previous authors in their almost exclusive attempts to extend limited information methods failed to realize that an appropriate full information method, by using all prior information, could make estimation possible. Also, I point out an error of Klein [4] on degrees of freedom restrictions for FIML estimation. I establish that FIML has less stringent degrees of freedom requirements than other estimators rather than more stringent requirements as he asserted.

Then using the instrumental variable interpretation, a relation between FIML and the class estimators recently proposed by Dhrymes [2], Lyttkens [5], and Brundy and Jorgenson [1] is established. The full information instrumental variable estimators are shown to be special cases of the basic FIML iteration. Furthermore, if they are iterated and converge, the resulting estimates are the FIML estimates.

Lastly, FIML is considered in the nonlinear case; and it is shown that in the special case of nonlinearity in the parameters the instrumental variable interpretation can be extended to provide an asymptotically efficient estimator with less computation needed than the FIML estimator. In a similar way a nonlinear three state least squares estimator is proposed and demonstrated to be asymptotically equivalent to FIML. NL3SLS again neglects some over-identifying restrictions in forming the instruments so that in finite sample the instrumental variable estimator which uses all the restrictions seems preferable.

2. Specification and Assumptions for the Linear Case

The standard linear simultaneous equations model is considered first, where all identities are assumed to have been substituted out of the system of equations:

$$(1)\ YB + Z\Gamma = U.$$

where Y is the T x M matrix of jointly dependent variables, Z is the T x K matrix of predetermined variables, and U is a T x M matrix of the structural disturbances of the system. The model thus has M equations and T observations. It is assumed that B is nonsingular, $rk(Z) = K$, and that all equations satisfy the rank condition for identification. Also if lagged endogenous variables are included as predetermined variables, the system is assumed to be stable. Lastly, an orthogonality assumption, $E(Z'U) = 0$, between the predetermined variables and structural errors is required; and the second order moment matrices of the current predetermined and endogenous variables are assumed to have non-singular probability limits.

The structural errors are assumed to be mutually independent and identically distributed (iid) as a nonsingular M-variate normal (Guassian) distribution:

$$(2)\ U \sim N(0,\ \Sigma \otimes I_T)$$

where Σ is positive definite almost surely, and no restrictions are placed on Σ. Thus for the present we assume the presence of contemporaneous correlation but no intertemporal correlation. The (column) vectors of U are thus distributed as univariate normal, $U_i \sim N(0, \Sigma)$.

Now the identification assumptions will exclude some variables from each equation so let r_i and s_i denote the number of included jointly dependent and predetermined variables, respectively, in the i^{th} equation. Then rewriting (1) after choice of a normalization rule:

$$(3)\ y_i = X_i\ \delta_i + U_i \qquad\qquad (i = 1, 2, \ldots, M)$$

where

$$X_i = [Y_i\ Z_i]$$

$$\delta_i = \begin{bmatrix} \beta_i \\ \gamma_i \end{bmatrix}$$

so that X_i contains the $t_i = r_i + s_i - 1$ variables whose coefficients are not known a priori to be zero. It will prove convenient to stack these M equations into a system:

$$(4)\ y = X\delta + u$$

where $y = \begin{bmatrix} y_i \\ \vdots \\ y_M \end{bmatrix}$, $X = \begin{bmatrix} X_1 & & 0 \\ & \ddots & \\ 0 & & X_M \end{bmatrix}$, $\delta = \begin{bmatrix} \delta_1 \\ \vdots \\ \delta_M \end{bmatrix}$, $u = \begin{bmatrix} u_1 \\ \vdots \\ u_M \end{bmatrix}$.

3. An Instrumental Variable Interpretation of FIML

The technique used to derive an instrumental variable interpretation of FIML is similar to, but not identical with, a proposal by Durbin in an unpublished paper. While not deriving Durbin's result from the likelihood function, Malinvaud states the estimator which he calls 'Durbin's Method' [7, pp. 686-7]. However, the resulting estimator differs from the estimator proposed here by not making full use of the identifying restrictions and being identical only in the case of a just-identified system. The instrumental variable interpretation of a maximum likelihood estimator while known in the case of non-simultaneous equation models is here extended to the case of FIML thus giving an integrated method in which to interpret the many estimators proposed for econometric models.

Given assumption (2) the likelihood function of the sample is

$$(5)\ \tilde{L}(B,\ \Gamma,\ \Sigma) = (2\Pi)^{-MT/2}\ \det(\Sigma)^{-T/2}\ \det(B)^{T}$$

$$\cdot\ \exp\ [-\ \frac{1}{2}\ \mathrm{tr}(YB + Z\Gamma)'\ \Sigma^{-1}\ (YB + Z\Gamma)]$$

Taking logs and rearranging, we derive the function to be maximized

$$(6)\ L(B,\ \Gamma,\ \Sigma) = C + \frac{T}{2}\ \log\ \det(\Sigma)^{-1} + T\ \log\ \det(B)$$

$$-\ \frac{T}{2}\ \mathrm{tr}\ [\frac{1}{T}\ \Sigma^{-1}\ (YB + Z\Gamma)'\ (YB + Z\Gamma)].$$

where the constant, C may be disregarded in maximizing the likelihood function. Since no restrictions have been placed on the elements of Σ, the usual procedure is to 'concentrate' the likelihood function by partially maximizing the function with respect to Σ. This procedure sets $\Sigma = T^{-1}(YB + Z\Gamma)' (YB + Z\Gamma)$ and thus eliminates Σ from the likelihood function, leaving a function $L^*(B, \Gamma)$ to be maximized. Our procedure instead concentrates on the presence of the Jacobian det(B) in the likelihood function which differentiates the simultaneous equation problem from the Zellner [10] multivariate least squares problem. For if the Jacobian of the transformation from U to Y, ${}^{\partial U}/_{\partial Y}$, were an identity matrix, the maximum likelihood estimator would be the generalized least squares estimator. Also it will be seen in a later section that the Jacobian is crucial in the development of a non-linear FIML estimator.

To maximize the log likelihood function $L(B, \Gamma, \Sigma)$, the necessary conditions for a maximum are the first order conditions obtained by differentiating (6) using the relation ${}^{\partial \log \det(A)}/_{\partial A} = (A')^{-1}$. Note that the a priori restrictions have been imposed so that only elements corresponding to non-zero elements of B and Γ are set equal to zero:

(7) $\frac{\partial L}{\partial B}$: $T(B')^{-1} - Y \quad (YB + Z\Gamma)\ \Sigma^{-1} = 0$

(8) $\frac{\partial L}{\partial \Gamma}$: $- Z' \ (YB + Z\Gamma)\ \Sigma^{-1} = 0$

(9) $\frac{\partial L}{\partial \Sigma}$: $T\Sigma - (YB + Z\Gamma)' \ (YB + Z\Gamma) \ = 0$

Concentration of the likelihood function follows from solving for Σ in equation (9); here we solve for T using equation (9). Since the M-variate distribution has been assumed non-singular, from equation (2) Σ is positive definite almost surely so from equation (9),

(10) $T \cdot I = (YB + Z\Gamma)' \ (YB + Z\Gamma)\ \Sigma^{-1}$.

(7)

Substituting this result for the first term in equation (7) yields

$$(11)\ (B')^{-1}\ (YB + Z\Gamma)'\ (YB + Z\Gamma)\ \Sigma^{-1} - Y'\ (YB + Z\Gamma)\ \Sigma^{-1} = 0.$$

The first term in (11) represents the presence of the non-identity Jacobian, but this term can be simplified by rearranging to get

$$(12)\ [(B')^{-1}\ B'Y' + (B')^{-1}\ \Gamma'Z'][YB + Z\Gamma]\ \Sigma^{-1}$$

$$- Y'\ (YB + Z\Gamma)\ \Sigma^{-1} = 0.$$

Noting that in equation (12) the first and last terms are identical with opposite sign, we have the desired first order condition

$$(13)\ (B')^{-1}\ \Gamma'Z'\ (YB + Z\Gamma)\ \Sigma^{-1} = 0$$

Therefore equations (8) and (13) must be solved and 'stacking' them together yields the final form of the necessary conditions

$$(14)\ \begin{pmatrix} -Z' \\ (B')^{-1}\ \Gamma'Z' \end{pmatrix} (YB + Z\Gamma)\ \Sigma^{-1} = 0.$$

Rewriting equation (14) in the form of equation (4), the FIML estimator $\hat{\delta}$ of the unknown elements of δ in instrumental variable form is:

$$(15)\ \hat{\delta} = (\bar{W}'X)^{-1}\ \bar{W}'_{y}$$

where the instruments are

(16) $\bar{W} = \hat{X}\ (S \otimes I_T)^{-1}.$

The elements of $\bar{W}$ are then

(17) $\hat{X} = \text{diag}(\hat{X}_1,\ \hat{X}_2,\ \ldots.,\ \hat{X}_M),\ \hat{X}_i = [Z(\hat{\Gamma}\hat{B}^{-1})_i\ Z_i]$

and from equation (9)

(18) $S = T^{-1}\ (Y\hat{B} + Z\hat{\Gamma})'\ (Y\hat{B} + Z\hat{\Gamma}).$

The instrumental variable interpretation of equations (15) and (16) is immediate since the second order moment matrices exist and are non-singular, and by the orthogonality assumption $E(Z'U) = 0$. In the instrumental variable interpretation of generalized least squares where only predetermined variables appear in X, the instruments are all the predetermined variables $\bar{W} = Z\ (S \times\ I)$ while here the included endogenous variables are replaced by consistent estimates which are then used as the instruments.

Equation (15) is non-linear since both $\hat{X}$ and S depends on $\hat{B}$, $\hat{\Gamma}$ which are elements of $\hat{\delta}$ and would therefore be solved by an iterative process ('Durbin's Method') where subscripts here denote iteration number:

(19) $\hat{\delta}_{k+1} = (\bar{W}_k'\ X)^{-1}\ \bar{W}_k'y.$

The limit of the iterative process, if it converges, δ^*, is the FIML estimate with asymptotic covariance matrix $(\hat{X}^{*\prime}\ (S^* \otimes\ I_T)^{-1}\ X)^{-1}$ since asymptotically

(20) $\sqrt{T}\ (\hat{\delta} - \delta) \overset{A}{\sim} N\ (0,\ V^{-1})$

where $V = -\operatorname{plim} E[\frac{1}{T} \partial^2 L / \partial\delta\partial\delta']$. Thus equation (15) extends the concept of instrumental variables to the maximum likelihood estimation of simultaneous equation models so that very simple comparisons with other proposed estimators are possible.

4. Equivalence of FIML and 3SLS

An instrumental variable interpretation of 3SLS was first advanced by Madansky [6]. In this interpretation the 3SLS estimator has the form

$$(21)\ \hat{\delta}_{3SLS} = (\tilde{W}' X)^{-1} \tilde{W}' y$$

where here the instruments are

$$(22)\ \tilde{W} = Z\ (\tilde{S} \otimes Z' Z)^{-1} Z' X.$$

The elements of $\tilde{W}$ are

$$(23)\ X = \text{diag}\ (X_1, \ldots, X_M),\ X_i = [Y_i\ Z_i].$$

and $\tilde{S}$ is the consistent estimate of Σ derived from the residuals of the structural equations estimated by 2SLS. The essential differences between FIML and 3SLS may be discovered by an examination of the difference in instruments between equation (16) and equation (22). The first difference is the consistent estimation of the variance covariance matrix Σ. They are asymptotically equivalent in probability limit since by consistency

$$(24)\ \text{plim}\ S = \text{plim}\ \tilde{S} = \Sigma.$$

The second difference is that the FIML estimator uses <u>all</u> a priori restrictions in computing the instruments, while as seen from equation (22), 3SLS uses an unrestricted estimate in computing the instruments. Again asymptotic equivalence follows since

(25) $\text{plim}\ \hat{\Gamma}\ \text{plim}\ \hat{B}^{-1} = \text{plim}\ (Z'Z)^{-1}\ Z'\ Y$

since $\text{plim}\ Z'U = 0$ and $\hat{\Gamma}$, $\hat{B}^{-1}$ have finite probability limits by assumption. Therefore the two differences lies in not making complete use of the identifying restrictions in estimating the instruments $\bar{W}$ and $\tilde{W}$ and in different estimates of the covariance matrix Σ. In finite samples the two methods can be equivalent only in the just identified case since the instruments would then be identical. Lastly the equivalence results of Sargan [9] and Rothenberg and Leenders [8] are obtained without the necessity of an asymptotic expansion by the result that $\bar{W}$ and $\tilde{W}$ are equivalent in the probability limit. Furthermore, any other asymptotically efficient instrumental variable estimator may be proved equivalent to FIML by the same technique.

5. The Incorrectness of Klein's Degrees of Freedom Restrictions

In the finite sample case it is known that important restrictions on the use of 3SLS and efficient limited information estimators such as LIML and 2SLS are degrees of freedom restrictions. The binding restriction is usually that the number of observations T must be no less than the total number of predetermined variables in the system K. Evaluation of the inner terms of the 3SLS instruments

(26) $(\tilde{S} \otimes Z'Z)^{-1} = (\tilde{S}^{-1} \otimes (Z'Z)^{-1})$

and use of the elementary result

(27) Lemma 1: $\text{rk}\ (AB) \leq \min\ (\text{rk}\ (A), \text{rk}\ (B))$

implies that $\text{rk}\ (Z'Z) \leq \min\ (K, T)$ and so for $Z'Z$ to be of full rank it is necessary that $K \leq T$. We now show that this degrees of freedom restriction is not binding for FIML and develop exact degrees of freedom restrictions. Consideration of the instruments $\bar{W}$ in equation (16) implies that S^{-1} must exist. From equation (18) $S = T^{-1}\ \hat{U}'\ \hat{U}$ such that application of lemma 1 results in the condition that $\text{rk}\ (S) \leq \min\ (M, T)$. Therefore a necessary condition for estimability of FIML is $M \leq T$. That this condition is almost surely sufficient follows from

(28) Lemma 2: Let $X_1, \ldots, X_T$ be a random sample from an absolutely continuous M-variate distribution with non-singular covariance matrix Σ. Then the sample covariance matrix S is positive definite with probability one iff $T \geq M$.

Proof: Necessity follows from straightforward application of lemma 1 (27) and sufficiency comes from the following argument. The moment matrix is positive semi-definite and the determinant vanishes only if the observations are linearly dependent. But if the joint distribution of the observations is absolutely continuous given Σ non-singular, the probability of an exact linear relationship is zero.

After insuring the nonsingularity of S, the only remaining task is to derive conditions for the nonsingularity of $(\bar{W}' X)$. The necessary condition here after again using Lemma 1 (27) is that

(29) $T \geq r_i + s_i - 1$ for all $i = 1, \ldots, M$.

This condition is just the usual least squares condition that the number of 'right hand side' variables must not exceed the number of observations. It is not often a binding restriction in FIML estimation. As 3SLS must also satisfy the condition of Lemma 2 since $\tilde{S}$ is a moment matrix, it is seen by application of the order condition for identification that the conditions in finite samples for the FIML estimate are weaker than those for 3SLS estimation. The order condition of identification states that for each equation the number of predetermined variables excluded from the equation must be at least as great as one less than the number of endogenous variables included in the equation

(30) $K - s_i \geq r_i - 1$ for all $i = 1, \ldots, M$.

Rearranging gives $K \geq s_i + r_i - 1$ and since 3SLS requires $T \geq K$ it is seen that also for 3SLS, $T \geq r_i + s_i - 1$ for all i. Thus collecting results we have the following:

Theorem 1: Under the assumptions for the linear simultaneous equations model of Section 2, the following conditions are necessary and almost surely sufficient for the estimability of the FIML estimate: (i) the number of observations must be at least as great as the number of endogenous variables, $M \leq T$ (ii). For each equation, the number of observations must be at least as great as the number of included "right hand side" variables after normalization, $T \geq r_i + s_i - 1$ for all $i = 1, \ldots, M$. the estimability of 3SLS requires a strengthening of condition (ii) so that the number of observations must be at least as great as the total number of exogenous variables, $T \geq K$. Lastly, the FIML and 3SLS estimates are identical in the just-identified case.

These necessary conditions are in conflict with those of Professor Klein [4, p. 175-6] who places the following "degrees of freedom" restrictions on FIML estimation:

$$(31)\quad (i)'\ K < T \qquad (ii)'\ M < T \qquad (iii)'\ K + M < T \qquad (iv)'\ \sum_{i=1}^{M} (r_i + s_i - 1) < MT$$

Except for using a strong rather than weak inequality, Klein's condition (ii)′ is identical to our condition (i). Condition (iv)′ corresponds to (ii) when both sides are summed over all equations, but it is too weak; the condition must hold for each equation. Conditions (i)′ and (iii)′ would place greater restrictions on FIML estimation than 3SLS if correct; but Professor Klein is incorrect in claiming that the moment matrix including all endogenous and predetermined variables must be nonsingular (his $W'W$ matrix on p. 176).

He has neglected to impose the a priori restrictions which enter in equations (15) and (16) and this make the requirements for estimation of FIML <u>weaker</u>, not stronger, than 3SLS as he implicitly asserts. Intuitively, this result again follows from the difference in instruments for FIML and 3SLS, $\bar{W}$ and $\tilde{W}$, respectively. FIML imposes all a priori restrictions in computing the instruments while 3SLS neglects these restrictions. That FIML imposes all a priori restrictions is also the reason for its estimability when efficient limited information methods, 2SLS and LIML, cannot be used. Since they too treat all equations but the one being estimated as just identified, they will have two of the restrictions of 3SLS: $T \geq K$ and $T \geq r_i + s_i - 1$ for all equations to be estimated. Thus in many actual cases where limited information estimation or 3SLS estimation is impossible, the full information maximum likelihood estimate can be computed.

6. The Relationship of FIML to Recently Proposed Instrumental Variable Estimators

Three recent papers have proposed instrumental variable estimators for linear simultaneous equation systems. Here these estimators are all shown to be particular cases of the basic FIML iteration developed in equation (19). Lyttkens [5], and Dhrymes [2], and Brundy and Jorgenson's [1] estimators all have the form:

(i) Construct a consistent estimate of the structural parameters (δ, Σ). These initial consistent estimates may be obtained by the use of consistent, but possibly inefficient, instrumental variable estimators using the format of equation (3). This procedure is always possible so long as $T \geq r_i + s_i - 1$ for all $i = 1, \ldots, M$ which is condition (ii) of Theorem 1. In constructing the instruments for equation i, W_i, to insure consistency it is necessary to include all s_i predetermined variables from equation i as instruments. The remaining $r_i - 1$ instruments can be constructed by regressing the $r_i - 1$ jointly dependent variables in equation i on a subset of all the excluded predetermined variables. By the orthogonality assumption, $E(Z'U) = 0$, the estimates $\hat{\delta}_i$ will be consistent but, in general, not efficient estimates. This procedure is followed for all M equations; and S, a consistent estimate of Σ, is derived from the residuals of the structural equations in the usual manner.(1)

---

1. Lyttken's method does not compute S, but rather uses the identity matrix. Thus his estimator is consistent but not generally efficient.

(ii) Construct system instrumental variables $\bar{W}$ using the form of equation (16), $\bar{W} = \hat{X}\ (S \otimes I_T)^{-1}$. Consistent estimates of X are provided from the first step of the procedure since by definition $\hat{\delta}_i = [\hat{\beta}_i\ \hat{\gamma}_i]'$ and from equation (16) $\hat{X}_i = [Z\ (\hat{\Gamma}\ \hat{B}^{-1})_i\ Z_i]$. Note that <u>all</u> a priori restrictions are being imposed to estimate the instrumental variables $\bar{W}$ rather than unrestricted estimates as in k-class and 3SLS instruments $\tilde{W}$ as shown in equation (22).

(iii) Estimate the structural parameters as in equation (19), $\hat{\delta} = (\bar{W}'\ X)^{-1}\ \bar{W}'\ y$. If desired, compute efficient estimates of Σ and the reduced form parameters.

Brundy and Jorgenson stop at this point and have efficient estimates since their estimates converge in distribution to the FIML estimates by an identical argument as that of equations (24) and (25). Lyttkens and Dhrymes propose an iterative process between steps (ii) and (iii) while unaware of the properties of the final estimates. But since this procedure is in every way identical to equation (19), by the earlier derivation if the iteration converges the estimates $(\hat{\delta}^*, S^*)$ are the FIML estimates! Thus these iterated instrumental procedures will be numerically <u>identical</u> to FIML if both use identical initial consistent estimates. Thus Dhrymes' [2] question of the effect of the initial estimates used in step (i) is answered for small samples; and for large samples even without identical initial estimates, under the usual regularity conditions the Cramer-Rao theorem can be invoked to insure a unique maximum likelihood estimate almost surely.

Also, note that the so-called limited information procedure proposed by Brundy and Jorgenson is misnamed. The procedure is identical to Lyttkens in using the identity matrix as an estimate of the contemporaneous correlation matrix Σ. This procedure is not limited information since it utilizes all the a priori restrictions on the δ_i in estimating the instrumental variables of step (ii). Thus any error of misspecification will be propogated throughout the entire system rather than being confined to the equation in which it occurs as in time limited information methods. Since the a priori restrictions are being imposed, FIML or its one iteration special case might as well be used to provide fully efficient estimates rather than only consistent estimates which the Brundy-Jorgenson 'limited information' procedure gives.

Lastly, while multicollinearity often makes computation of the unrestricted instrumental variables, $\tilde{W}$, used in 3SLS as in equation (22) extremely difficult, since FIML and the single iteration procedures use fully restricted estimates $\bar{W}$ an in equation (17) this problem will no longer exist. Thus in the full information context, procedures using principal components need not be used for the multicollinearity problem. Also, in the finite sample case since all a priori restrictions are being imposed these instrumental variable procedures might well be preferred to 3SLS which imposes the restrictions only in the final stage. In 3SLS the estimates of the included right hand side endogenous variables often differs little from the actual and presumably non-orthogonal variables due to lack of degrees of freedom, but FIML and the instrumental variable procedures by imposing all the a priori restrictions will often have many more degrees of freedom in estimation.

7. FIML for Nonlinear Systems

Consider the general nonlinear simultaneous equation system

$$(32)\quad F(Y, Z; \alpha) = U.$$

Here F is an MT vector of functions $[f_1, f_2, \ldots., f_M]$ which in a neighborhood in R dimensional space of the true parameter values, α^*, are assumed uniformly bounded and three times differentiable with uniformly bounded derivatives. Also, the f_i are assumed continuous with respect to Y and Z. The system is assumed to be identified and α to belong to a compact subset of R dimensional space. As before the structural errors are assumed i.i.d. and distributed as a nonsingular M-variate normal distribution.

The log of the likelihood function is

$$(33)\quad L(\alpha, \Sigma) = C + {}^{T}/_{2} \log \det (\Sigma)^{-1} + \sum_{t=1}^{T} \log | J_t |$$

$$- \frac{T}{2} \operatorname{tr} [\frac{1}{T} \Sigma^{-1} F (Y, Z; \alpha)' F (Y, Z; \alpha)]$$

where $|J_t|$ is the Jacobian of the transformation from U to Y. Note the important complication introduced by the nonlinear structural system is that in equation (33) the Jacobian is no longer constant as in equation (6) but instead varies with each observation. Therefore, the first order conditions cannot be simplified as in equations (11) and (12) to provide a convenient iterative procedure. The log likelihood in principle can be maximized by straightforward 'hill-climbing' algorithems but this procedure may prove impractical unless special assumptions are made about the structural system.

One very special case of the general model, which nevertheless is quite common to econometric models, is that of non-linearity only in the parameters. Here the structure is linear in the variables and nonlinear in the parameters A which are analytic non-linear functions of an R dimensional vector of parameters α so that

$$(34)\quad U = XA(\alpha) = YB(\alpha) + Z\Gamma(\alpha)$$

where $X = [Y\ Z]$. Two important examples of such a structure are linear simultaneous equation models with autoregressive errors and partial adjustment or distributed lag models containing a 'desired' stock which is a function of structural parameters. Writing out the log of the likelihood function where $B(\alpha)$ are the parameters of the endogenous variables gives

$$(35)\quad L(\alpha, \Sigma) = C + \frac{T}{2} \log \det (\Sigma)^{-1} + T \log \det (B(\alpha))$$

$$- \frac{T}{2} \operatorname{tr} [\frac{1}{T} \Sigma^{-1} (XA(\alpha))' (XA(\alpha))].$$

The Jacobian is once again constant and the first order conditions are

$$(36)\quad \frac{\partial L}{\partial \alpha} : \quad T(B'(\alpha))^{-1} \frac{\partial B}{\partial \alpha} - (\frac{\partial A}{\partial \alpha})' X' (XA(\alpha)) \Sigma^{-1} = 0$$

$$(37)\quad \frac{\partial L}{\partial \Sigma} : \quad T\Sigma - (XA(\alpha))' (XA(\alpha)) = 0$$

Noting that ${}^{\partial B}/_{\partial \alpha}$ is a submatrix of ${}^{\partial A}/_{\partial \alpha}$ and using the same substitution technique of equations (11) and (12) yields the non-linear iterative equation

$$(38)\quad \hat{\alpha}_{k+1} = (\bar{\bar{W}}_k' \; X \cdot \frac{\partial A}{\partial \alpha})^{-1} \; \bar{\bar{W}}_k' \; y$$

where the instruments are

$$(39)\quad \bar{\bar{W}} = \frac{\partial A}{\partial \alpha} \cdot \hat{X} \; (S \otimes I_T)^{-1}.$$

The elements of $\bar{\bar{W}}$ are

$$(40)\quad \hat{X} = \mathrm{diag}\; (\hat{X}_1, \hat{X}_2, \ldots., \hat{X}_M),\; \hat{X}_i = [Z \; (\hat{\Gamma}(\alpha) \; \hat{B}(\alpha)^{-1})_i \; Z_i]$$

and from equation (37):

$$(41)\quad S = T^{-1} \; (XA(\alpha))' \; (XA(\alpha))$$

The limit of the iterative process α^* is the FIML estimate with asymptotic covariance matrix $(^{\partial A}/_{\partial \alpha} \cdot \hat{X}^{*\prime} \; (S^* \otimes I_T)^{-1} \cdot X \cdot {}^{\partial A}/_{\partial \alpha})^{-1}$

In Section 4 the instrumental variable interpretation of 3SLS was given and its asymptotic equivalence to FIML shown. In a similar manner for the non-linear in parameters system of equation (34) a non-linear 3SLS estimator (NL3SLS) may be defined

$$(42)\quad \hat{\alpha}_{NL3SLS} = (\tilde{\tilde{W}}' \cdot X \frac{\partial A}{\partial \alpha})^{-1} \; \tilde{\tilde{W}}' \; Y$$

with the instruments

$$(43)\quad \tilde{\tilde{W}} = Z \; (\tilde{\tilde{S}} \otimes Z'Z)^{-1} \; Z' \; X \; \frac{\partial A}{\partial \alpha}$$

where $\tilde{\tilde{S}}$ is a consistent estimate of Σ. Note that the NL3SLS estimator is nonlinear due to the presence of the $\frac{\partial A}{\partial \alpha}$ matrix and will therefore require an iterative procedure. However, it will require less computation than the FIML estimator since the large block $Z\ (\tilde{\tilde{S}} \otimes Z'Z)^{-1}\ Z'\ X$ remains constant while FIML revises $\hat{X}$ on each iteration. The asymptotic equivalence of NL3SLS and FIML follows from application of equation (25) with the asymptotic covariance matrix of NL3SLS being $(\frac{\partial A}{\partial \alpha}\ \ X'\ Z\ (\tilde{\tilde{S}} \otimes Z'Z)^{-1}\ Z'\ X\ \frac{\partial A}{\partial \alpha})$.

For asymptotic efficiency there is no need to define a non-linear 2SLS estimator for the computation of $\tilde{\tilde{S}}$. Any consistent estimate will do; and in particular, the estimate derived by not imposing the across-parameter constraints is easily shown to be consistent. For example in the autoregressive case, a parameter from the autoregressive specification will usually multiply more than one of the other parameters. An unconstrained estimate which treats each of the terms as different will yield consistent estimates of the disturbances from which a consistent estimate of Σ follows in the usual way.

Another estimator which is asymptotically efficient and uses more restrictions in the estimation of the instruments than does 3SLS is the nonlinear analogue of the Lyttkens, Dhrymes, and Brundy and Jorgenson procedures. It corresponds to one step of the FIML iteration:

(i) Construct a consistent estimate of the structural parameters $(A(\alpha), \Sigma)$. A consistent, but inefficient, instrumental variables procedure on each of the M equations in which the nonlinear constraints are not imposed is the most simple procedure. A consistent estimate S follows in the usual manner.

(ii) Use these consistent estimates to form instruments with equation

(40) defining $\hat{X}$ and the iterative procedure

(44) $\hat{\hat{\alpha}} = (\bar{\bar{W}}' X \frac{\partial A}{\partial \alpha})^{-1} \bar{\bar{W}}' y.$

Here the instruments $\bar{\bar{W}}$ remain constant with respect to the $\hat{X}$ term while the $\frac{\partial A}{\partial \alpha}$ matrix changes on each iteration until convergence is achieved. Thus the iterative procedure differs from FIML where $\hat{X}$ is also changing with each iteration.

Thus three instrumental variable estimators have been proposed to treat the non-linear in parameters case: FIML, NL3SLS, non-linear instrumental variables. Each provides asymptotically efficient estimates with FIML requiring the most computation since both $\hat{X}$ and $\frac{\partial A}{\partial \alpha}$ are changing across iterations. The other two procedures keep $\hat{X}$ constant and iterate only over $\frac{\partial A}{\partial \alpha}$. As before, on a consideration of degrees of freedom the non-linear instrumental variables procedure might be preferred to NL3SLS.

8. Conclusions

The conceptual framework of instrumental variables has been used to demonstrate the close relation of FIML, 3SLS, and recently proposed instrumental variable procedures. While Madansky had established an instrumental variable interpretation of 3SLS, the instrumental variable interpretation of FIML is new and leads to an extremely simple asymptotic equivalence by showing convergence in distribution of the two estimators. The other instrumental variable procedures are shown to be one step of the FIML iteration, and therefore if they are iterated, will yield the FIML estimate.

Exact degrees of freedom requirements for estimability of FIML are calculated and are weaker, not stronger than those for 3SLS as Professor Klein implied. Thus FIML is a possible estimator when 3SLS, 2SLS, LIML, and other k-class estimators cannot be used. This result follows since FIML imposes all a priori restrictions in forming the instruments while the other estimators use unrestricted estimates as instruments. Thus the multicollinearity problem present in computing 3SLS will be lessened by using the a priori restrictions.

Lastly, FIML and two new estimators, NL3SLS and non-linear instrumental variables, are developed for the important case of a structural system which is nonlinear in the parameters. All three procedures require an iterative method and are asymptotically efficient. Again, FIML requires the most computation while NL3SLS does not impose a priori restrictions in forming the instruments.

References

[1] Brundy, J. and D.W. Jorgenson, "Efficient Estimation of Simultaneous Equation Systems by Instrumental Variables", Review of Economics and Statistics, 53, 1971, pp. 207-224.

[2] Dhrymes, P.J., "A Simplified Structural Estimator for Large-Scale Econometric Models", The Australian Journal of Statistics, 13, 1971, pp. 168-75.

[3] Dhrymes, P.J., "Small Sample and Asymptotic Relations Between Maximum Likelihood and Three Stage Least Squares Estimators", Econometrica, 41, 1973, pp. 357-64.

[4] Klein, L.R., "Estimation of Interdependent Systems in Macroeconometrics", Econometrica, 37, 1969, pp. 171-92.

[5] Lyttkens, E., "Symmetric and Asymmetric Estimation Methods", in E. Mosback and H. Wold, Interdependent Systems, Amsterdam 1970, pp. 434-59.

[6] Madansky, A., "On the Efficiency of Three-Stage Least Squares Estimation", Econometrica, 32, 1964, pp. 51-6.

[7] Malinvaud, E., Statistical Methods of Econometrics, Amsterdam 1970.

[8] Rothenberg, T. and C. Leenders, "Efficient Estimation of Simultaneous Equation Systems", Econometrica, 32, 1964, pp. 57-76.

[9] Sargan, J.D., "Three-Stage Least Squares and Full Maximum Likelihood Estimates", Econometrica, 32, 1964, pp. 77-81.

[10] Zellner, A., "An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias", _Journal of the American Statistical Association_, 57, 1962, pp. 348-68.