Volume Title: Annals of Economic and Social Measurement, Volume 3, number 4

Volume Author/Editor: Sanford V. Berg, editor

Volume Publisher: NBER

Volume URL: http://www.nber.org/books/aesm74-4

Publication Date: October 1974

Chapter Title: Full Information Instrumental Variables Estimation of Simultaneous Equations Systems

Chapter Author: Jerry A. Hausman

Chapter URL: http://www.nber.org/chapters/c10205

Chapter pages in book: (p. 641 - 652)

Annals of Economic and Social Measurement, 3/4, 1974

FULL INFORMATION INSTRUMENTAL VARIABLES ESTIMATION OF SIMULTANEOUS EQUATIONS SYSTEMS

BY J. A. HAUSMAN

Three full information estimators—3SLS, FIML, and full information instrumental variables (FIIV)—are compared, based on an instrumental variable interpretation of FIML. In a test of the estimators on Klein Model I, 3SLS is used to form the instruments for FIIV, and the latter is iterated to compute the FIML estimates. An algorithm is specified for ensuring an increase in the likelihood function at each iteration of the FIML estimator.

1. INTRODUCTION

Full information estimation of simultaneous equation models makes use of all a priori information and thus provides consistent and asymptotically efficient estimates of the parameters of the model. Under suitable regularity conditions the method of maximum likelihood applied to such models attains the Cramer-Rao lower bound as the sample size becomes large. This estimator, the full information maximum likelihood (FIML) estimator, in general requires the iterative solution of a set of nonlinear equations. Therefore other estimators have been proposed, requiring less computational effort but providing equivalent asymptotic properties. Zellner and Theil [11] proposed the method of three-stage least squares (3SLS), and recently Lyttkens [5], Dhrymes [2], and Brundy and Jorgenson [1] have all proposed full information instrumental variables (FIIV) estimators which are also consistent and asymptotically efficient. Extending my previous work [3], in this paper I investigate the relation of all three estimators by examining their properties in the form of instrumental variables estimators.

After specifying the standard linear simultaneous equations model in the next section, in Section 3 I develop an instrumental variable interpretation of FIML (as in [3]). This interpretation of FIML permits easy comparison of other estimators with FIML; it does not require asymptotic expansions that have previously been necessary. Also, 3SLS and the full information estimator are compared with FIML by determining how their instruments differ from the FIML instruments. Lastly, the asymptotic covariance matrix of the FIML estimates is easily determined due to the instrumental variable form of the estimator.

In Section 4 the recently proposed FIIV estimators are shown to be one step of the basic FIML iteration when it is begun with a consistent estimate. The instruments are identical, so that if these estimators are iterated and converge, the resulting estimate will be the FIML estimate. The iterative property does not hold for 3SLS, and I show that the essential difference is that 3SLS ignores overidentifying restrictions in formation of the instruments. Although by the usual first-order definition of efficiency this difference vanishes asymptotically, in finite samples there seems no reason to ignore a priori information.

In Section 5 the properties of the three estimators are tested on Klein Model I, a well-known, three-equation econometric model. The 3SLS estimate is computed first and then used to form the instruments for the FIIV estimator. The FIIV estimator is then iterated, and the FIML estimates computed. As expected, the

FIIV estimates lie "between" the 3SLS estimates and FIML estimates. Even for the small Klein Model I the point estimates of the three models differ substantially.

The FIML estimator as computed here has one severe drawback. it lacks the "uphill property" of ensuring an increase in the likelihood function at each iteration. The uphill property holds only when a certain matrix is positive definite. Therefore in Section 6 I propose an approximation to the matrix which would yield the uphill property when the matrix is not positive definite. The approximation is asymptotically equivalent to the original matrix and is easily computed. Further experiments will be required to ascertain its properties in relation to other commonly used numerical procedures.

In the concluding section I refer to the obvious need for the extension of these techniques to find efficient estimators for nonlinear simultaneous equations models. I have made this extension for the special case of nonlinearity in the parameters, but the case of nonlinearity in the variables remains to be solved. Consistent estimates of the parameters can be found; but as consistency is a weak property, it would be desirable to have asymptotic efficient estimators.

2. Specification and Assumptions for the Linear Case

The standard linear simultaneous equations model is considered where, without restricting generality, all identities are assumed to have been substituted out of the system of equations:

$$YB + Z\Gamma = U. \tag{1}$$

Here Y is the $T \times M$ matrix of jointly dependent variables, Z is the $T \times K$ matrix of predetermined variables, and U is a $T \times M$ matrix of the structural disturbances of the system. The model thus has M equations and T observations. It is assumed that B is nonsingular, $rk(Z) = K$; that all equations satisfy the rank condition for identification; and that the system is stable if lagged endogenous variables are included as predetermined variables. Further, 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 nonsingular probability limits.

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

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

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 (row) vectors of U are thus distributed as multivariate 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 rewrite (1) after choice of a normalization rule:

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

where

$$X_i = [Y_i Z_i]$$

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

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:

$$y = X\delta + u \tag{4}$$

where

$$y = \begin{bmatrix} y_1 \\ \vdots \\ y_M \end{bmatrix}, \quad X = \begin{bmatrix} X_1 & & 0 \\ & \ddots & \\ 0 & & X_M \end{bmatrix}, \quad \delta = \begin{bmatrix} \delta_1 \\ \vdots \\ \delta_M \end{bmatrix}, \quad u = \begin{bmatrix} U_1 \\ \vdots \\ U_M \end{bmatrix}.$$

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]. The instrumental variable interpretation of a maximum likelihood estimator, while known in the case of nonsimultaneous equations models, is here extended to the case of FIML: this extension gives an integrated method in which to interpret the many estimators proposed for econometric models.

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

$$\begin{aligned} \tilde{L}(B, \Gamma, \Sigma) &= (2\Pi)^{-MT/2} \det(\Sigma)^{-T/2} \det(|B|)^T \\ &\cdot \exp[-\tfrac{1}{2}\operatorname{tr}(YB + Z\Gamma)\Sigma^{-1}(YB + Z\Gamma)']. \end{aligned} \tag{5}$$

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

$$\begin{aligned} L(B, \Gamma, \Sigma) &= C + \frac{T}{2}\log\det(\Sigma)^{-1} + T\log\det(|B|) \\ &- \frac{T}{2}\operatorname{tr}\left[\frac{1}{T}\Sigma^{-1}(YB + Z\Gamma)'(YB + Z\Gamma)\right] \end{aligned} \tag{6}$$

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 $\Sigma = T^{-1}(YB + Z\Gamma)'(YB + Z\Gamma)$, thus eliminates Σ from the likelihood function, and leaves 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 equations 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.

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 nonzero elements of B and Γ are set equal to zero:

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

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

$$\text{(9)} \qquad \frac{\partial L}{\partial \Sigma^{-1}}: \; T\Sigma - (YB + Z\Gamma)'(YB + Z\Gamma) = 0$$

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 nonsingular, from equation (2) Σ is positive definite almost surely and so from equation (9)

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

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

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

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

$$\text{(12)} \qquad [(B')^{-1}B'Y' + (B')^{-1}\Gamma'Z'][YB + Z\Gamma]\Sigma^{-1} - Y'(YB + Z\Gamma)\Sigma^{-1} = 0.$$

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

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

Therefore equations (8) and (13) must be solved, and "stacking" them together yields the final form of the necessary conditions where the included variables correspond to the unknown parameters in B and Γ:

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

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:

$$\text{(15)} \qquad \hat{\delta} = (\overline{W}'X)^{-1}\overline{W}_y'$$

where the instruments are

$$\text{(16)} \qquad \overline{W}' = \hat{X}'(S \otimes I_T)^{-1}.$$

The elements of $\overline{W}'$ are then

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

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 nonsingular, 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 $\overline{W}' = Z'(S \otimes I)^{-1}$ while here the included endogenous variables are replaced by consistent estimates which are then used as the instruments.

Equation (15) is nonlinear since both $\hat{X}$ and S depend on $\hat{B}$, $\hat{\Gamma}$, which are elements of $\hat{\delta}$ and would therefore be solved by an iterative process ("Durbin's Method")

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

where subscripts denote iteration number. δ^*, the limit of the iterative process (if it converges), 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 $V = \lim_{n\to\infty} E[(1/T)\partial^2 L(\partial\delta\partial\delta')]$. Thus equation (15) extends the concept of instrumental variables to the maximum likelihood estimation of simultaneous equations models, so that very simple comparisons with other proposed estimators are possible.

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

Three recent papers have proposed new instrumental variable estimators for linear simultaneous equations systems. Here all these estimators are 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 following 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 $i = 1, \ldots, M$. When W_i', the instruments for equation (i), are constructed, to ensure 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, $\Sigma(Z'U) = 0$, 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]

(ii) Construct system instrumental variables $\overline{W}$ using the form of equation (16). $\overline{W}' = \hat{X}'(S \otimes I_T)^{-1}$. Consistent estimates of X are provided from the first step of the procedure. since by definition $\hat{\delta}_i = [\hat{B}_i\hat{\Gamma}_i]'$ and from equation (16) $\hat{X}_i = [Z(\hat{\Gamma}\hat{B}^{-1})_i Z_i]$. Note that this estimate imposes *all* a priori restrictions to estimate the instrumental variables $\overline{W}'$; whereas k-class and 3SLS ignore a priori restrictions in following the instruments $\overline{W}'$, as shown in equation (22).

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

Brundy and Jorgenson stop at this point and have efficient estimates. since their estimates converge in distribution to the FIML estimates. 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 *identical* 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 ensure 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 propagated throughout the entire system rather than being confined to the equation in which it occurs, as in true 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.

The last full information instrumental variable estimator which will be discussed is three stage least squares. In instrumental variable form as first expressed by Madansky [6]. 3SLS has the form

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

where the instruments are

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

[1] Lyttkens' method does not compute S. but rather uses the identity matrix. Thus his estimator is consistent but not generally efficient.

The elements of the instruments matrix $\bar{W}$ are

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

and S is the consistent estimate of Σ derived from the residuals of the structural equations estimated by 2SLS. A comparison of 3SLS and FIML is made in Hausman [3]; the main difference is that 3SLS fails to use all the a priori restrictions in forming the instruments $\bar{W}$. Thus while multicollinearity often makes it extremely difficult to compute the unrestricted instrumental variables, $\bar{W}$, in 3SLS, this problem will no longer exist since FIML and the FIIV estimators use fully restricted estimates of $\bar{W}$. In the finite sample case, the other two estimators might well be preferred to 3SLS since they impose all restrictions in estimation. However, these differences will probably be more serious in medium and large models. Since most empirical studies of full information estimators have concentrated on testing performance of small models, further evidence on larger models should be valuable in evaluating their respective finite sample properties.

5. A Numerical Comparison of 3SLS, FIIV, and FIML

For purposes of comparison the three proposed full information estimation techniques—3SLS, FIIV, and FIML—are applied to the often studied Klein Model I. The model consists of six equations, of which three are identities. The first equation is the consumption function

$$(24) \qquad C_t = \alpha_{11}(W_t + W_t') + \alpha_{12}P_t + \alpha_{13}P_{t-1} + \alpha_{10} + \varepsilon_{1t}$$

where C_t is aggregate consumption, P_t and P_{t-1} are current and lagged total profits, W_t is the private industry wage bill, and W_t' is the government wage bill. The next equation is the investment function

$$(25) \qquad I_t = \alpha_{21}P_t + \alpha_{22}P_{t-1} + \alpha_{23}K_{t-1} + \alpha_{20} + \varepsilon_{2t}$$

where I_t is net investment, P_t and P_{t-1} are again current and lagged profits, and K_t is the capital stock. The last stochastic equation is the wage equation

$$(26) \qquad W_t = \alpha_{31}Q_t + \alpha_{32}Q_{t-1} + \alpha_{33}(t - 1931) + \alpha_{30} + \varepsilon_{3t}$$

where W_t is the wage bill, Q_t and Q_{t-1} are current and lagged private output, and t is the time trend variable. The model is then closed by the three identities

$$(27) \qquad \begin{aligned} K_t &= K_{t-1} + I_t \\ Q_t &= C_t + I_t + G_t \\ P_t &= Q_t - W_t - T_t \end{aligned}$$

where the additional variables G_t and T_t are nonwage government expenditure and business taxes, respectively.

In estimation the three identities are deleted and therefore we are left with three equations having 12 unknown structural parameters. Three of these parameters are the constants corresponding to α_{10}, α_{20}, and α_{30}, and of the nine remaining parameters five correspond to predetermined variables. The time variable t along with the lagged variables P_{t-1}, K_{t-1}, Q_{t-1} are predetermined while

$(W_t + W_t')$, P_t, and Q_t are all endogenous. The estimation is done over Klein's original sample 1921–1941, so there are 21 annual observations.

While 3SLS and FIIV are linear in the sense of only solving a set of linear equations, FIML as shown in equation (15) is nonlinear since the elements of $\bar{W}$ depend on elements of $\hat{\delta}$ and therefore an iterative procedure is needed. Now as equation (19) makes clear, the full information instrumental variables efficient (FIVE) estimator proposed by Brundy and Jorgenson is merely the first step of the iterative process where the initial guesses are derived from a consistent estimation procedure. If equation (19) is iterated, which corresponds to the Dhrymes estimator, and if it converges, then the resulting estimates are the FIML estimates. This iterative procedure is not being advocated as an efficient computational procedure for FIML since it lacks the "uphill property" (to be discussed later), but this experiment is merely to show that Dhrymes' estimator is identical to FIML while the Brundy-Jorgenson estimator corresponds to one iteration of a FIML procedure and when iterated yields FIML.

The 3SLS estimates need no further explanation since the instruments $\bar{W}'$ are formed by using all the predetermined variables while neglecting overidentifying constraints in forming $\bar{W}'$ in equation (22). These initial consistent estimates are then used to form the instruments for the first stage of the FIML iteration. Here in forming $\bar{W}'$ all the overidentifying restrictions are used. The structural parameters corresponding to the first iteration $\hat{\delta}_1$ are then the FIIV estimates. Alternative efficient estimates can be obtained by other initial consistent estimates but all such estimates have identical asymptotic properties. Equation (19) took 41 iterations to reach the convergence criterion of

$$\left|\frac{\hat{\delta}_{k+1} - \hat{\delta}_k}{\hat{\delta}_k}\right| < 0.0005$$

where the norm used is the maximum change in an element of the $\hat{\delta}$ vector.

The final estimates (with asymptotic standard errors) are shown in Table 1. The FIIV estimates in the consumption equation (24) and wage equation (26) are reasonably close to the FIML estimates while the investment equation (25) "still has a long way to go". However, an examination of the covariance matrix presented below shows that the investment equation has by far the largest variance, which, in fact, exceeds the variance of the investment series over the 1921–1941 data period. Therefore it is not suprising that the point estimates of FIML differ markedly from the point estimates of the other estimators. The FIML estimates agree to three significant digits with the FIML estimates of the same model presented by Chernoff and Divinsky in Hood and Koopmans [12], page 284. The maximum of the log likelihood function (6) is -0.54815.

To complete structural estimation for Klein Model I, the asymptotic covariance matrix, S, for each estimator is presented. These estimates all follow from equation (18), which leads to the estimate of the covariance matrix, $S = T^{-1}(YB + Z\Gamma)'(YB + Z\Gamma)$. Along with the coefficient estimates, the covariance estimates provide asymptotic sufficient statistics for the normal distribution of equation (2). If desired, the reduced form coefficients and covariance matrix can then be calculated from $\hat{\Pi} = -\hat{\Gamma}\hat{B}^{-1}$ and $\hat{\Omega} = \hat{B}^{-1\prime}S\hat{B}^{-1}$. The covariance

TABLE 1

3SLS, FIIV, AND FIML ESTIMATES OF KLEIN MODEL 1
(Asymptotic standard errors in parentheses)

Equation	Parameter	Variable	3SLS	FIIV	FIML
Consumption (24)	α_{11}	$W_t + W_t'$	0.79008 (0.03794)	0.80145 (0.03496)	0.80183 (0.03589)
	α_{12}	P_t	0.12489 (0.10813)	−0.17713 (0.21225)	−0.23214 (0.31165)
	α_{13}	P_{t-1}	0.16314 (0.10044)	0.35691 (0.16370)	0.38557 (0.21720)
	α_{10}	1	16.441 (1.305)	17.897 (2.149)	18.341 (2.858)
Investment (25)	α_{21}	P_t	−0.013079 (0.16190)	−0.71470 (0.36873)	−0.80067 (0.49099)
	α_{22}	P_{t-1}	0.75572 (0.15293)	1.0274 (0.28989)	1.0517 (0.35224)
	α_{23}	K_{t-1}	−0.19485 (0.03253)	−0.15044 (0.03299)	−0.14811 (0.02986)
	α_{20}	1	28.178 (6.794)	26.676 (8.026)	27.263 (8.668)
Wage (26)	α_{31}	Q_t	0.40049 (0.03181)	0.24264 (0.04557)	0.23415 (0.04882)
	α_{32}	Q_{t-1}	0.18129 (0.03416)	0.28337 (0.04341)	0.28465 (0.04521)
	α_{33}	t-1931	0.14967 (0.03416)	0.22686 (0.03286)	0.23483 (0.03450)
	α_{30}	1	1.7972 (1.116)	5.3582 (2.034)	5.7939 (2.229)

estimates for the three estimators are:

$$S_{3SLS} = \begin{bmatrix} 0.89176 & & \\ 0.41132 & 2.0930 & \\ -0.39362 & 0.40305 & 0.52003 \end{bmatrix}$$

$$S_{FIIV} = \begin{bmatrix} 1.9859 & & \\ 3.0973 & 10.900 & \\ 0.29126 & 3.3799 & 1.6650 \end{bmatrix}$$

$$S_{FIML} = \begin{bmatrix} 2.1026 & & \\ 3.8754 & 12.764 & \\ 0.48080 & 3.8558 & 1.8007 \end{bmatrix}$$

No unique accepted method exists to evaluate the three estimators. While all three have identical asymptotic properties, the parameter estimates and covariance estimates differ substantially. To evaluate the parameter estimates, the quadratic loss function $R = (\hat{\delta} - \delta)'Q(\hat{\delta} - \delta)$ is used, where Q is the weighting matrix and δ the vector of true (unknown) parameter values. Then the measure of asymptotic expected loss is $\text{tr}(QV^{-1})$, where V^{-1} is the estimated covariance matrix of the structural parameters. Two weighting matrices are used, an identity matrix leading to the trace of V^{-1} and a matrix consisting of ones leading to the sum of the elements of V^{-1}. The results for this experiment on Klein Model I are:

Expected Loss	3SLS	FIIV	FIML
$\text{tr}(V^{-1})$	56.8	73.5	88.8
$\text{tr}(QV^{-1})$	43.4	96.4	122.5

These results are not surprising since 3SLS can be derived as the solution to the programming problem

$$\min_{\delta} [(y - X\delta)'(S^{-1} \otimes Z(Z'Z)^{-1}Z')(y - X\delta)]$$

where S^{-1} is the consistent estimate of covariance matrix from 2SLS. Thus, 3SLS minimizes the generalized distance of the errors when projected onto the subspace spanned by the predetermined variables. FIML, on the other hand, "trades off" the value of the Jacobian, det (B), against a generalized distance term. The FIIV estimator, as expected, falls between the two others. Thus although 3SLS seems to produce a "tighter" estimate, this result may be an illusion due to the different objective functions being maximized. The "true" parameter values would have to be known to make a valid comparison among the three estimators.

6. Computation of FIML Estimates

While it was shown that iterating equation (19) led to the FIML estimate and is equivalent to iterating the FIIV estimator, this procedure is not a very efficient method of computing FIML estimates. An easy way to see the problem is to rewrite equation (19) in the form of a change in the δ vector

$$(28) \qquad \Delta\hat{\delta}_{k+1} = \hat{\delta}_{k+1} - \hat{\delta}_k = (\overline{W}_k'X)^{-1}\overline{W}_k'(y - X\hat{\delta}_k) = (\overline{W}_k'X)^{-1}\overline{W}_k'u_k$$

where u_k is the vector of calculated residuals using $\hat{\delta}_k$. The maximum likelihood estimate is then attained when $\Delta\hat{\delta}_{k+1} = 0$. The problem which arises is that $(\overline{W}_k'X)$ is not in general positive definite and so one cannot hope to prove that the likelihood function will increase at each iteration. This monotonicity property is the "uphill property" referred to in the last section. Since equation (14) shows that $\overline{W}_k'u_k$ is a gradient of the likelihood function, any vector in the same halfspace as the gradient will be in a direction of increase for the likelihood function. Therefore for λ, a scalar, small enough the new estimate

$$(29) \qquad \hat{\delta}_{k+1} = \hat{\delta}_k + \lambda \cdot \Delta\hat{\delta}_{k+1}$$

will give an increased value of the likelihood function provided that $(\overline{W}_k'X)$ is positive definite. This property is very desirable since, in principle, convergence is guaranteed.[2] In practice, it also is an advantage since the procedure does not become stopped at a point where $\Delta\hat{\delta}_{k+1}$ is not zero although the likelihood function is not increasing in the calculated direction.

[2] To actually prove guaranteed convergence, one must show that λ does not become "too small", e.g., that λ is bounded away from zero.

An obvious way around this problem is to use asymptotically equivalent approximation to $(\overline{W}_k'X)$ when it is not positive definite. A possible procedure is to check whether $\overline{W}_k'X$ is positive definite; and if not use

$$\overline{W}'\hat{X} = \hat{X}'(S \otimes I_T)^{-1}\hat{X} \tag{30}$$

where $\hat{X} = \text{diag}(\hat{X}_1, \hat{X}_2, \ldots, \hat{X}_M)$, $\hat{X}_i = [Z(\hat{\Gamma}\hat{B}^{-1})_i Z_i]$, and S is the estimate of the covariance matrix. Since $\hat{\Gamma}$ and $\hat{B}$ are consistent estimates of Γ and B, the matrix $\overline{W}_k'\hat{X}$ is asymptotically equivalent to the matrix $\overline{W}_k'X$. Furthermore, S is positive definite so $\overline{W}_k'\hat{X}$ will be positive definite, and the iterative procedure of equation (30) has the monotonicity property with respect to the likelihood function.

Many other iterative procedures are possible. The Newton-Raphson procedure has often been used. Since the likelihood function is not concave in general, this method, when not in the neighborhood of the optimum, often encounters difficulty choosing directions of increase. A wide class of algorithms based on the Davidon variable metric procedure does guarantee the monotonicity property and has other computational advantages. The choice of a general procedure to calculate FIML estimates will require further experimentation—especially in larger systems, for which almost no results have been reported. The procedure outlined here has desirable asymptotic properties, but its use in actual calculations remains to be evaluated. It does have the computational advantage of not requiring second derivatives, but instead using the vector of instruments used in computing the gradient. This algorithm is the analogous procedure to the Gauss-Newton algorithm for nonlinear least squares. Since the Gauss-Newton algorithm (or minor modifications of it) have proved extremely effective, the algorithm proposed here, with the uphill property modification in equation (30), might also prove effective in computing FIML estimates.

7. Conclusion and Directions for Future Research

An instrumental variable of interpretation of FIML has been developed which permits easy comparison with other proposed instrumental variable estimators such as 3SLS, and the FIIV estimators recently proposed by Lyttkens, Dhrymes, and Brundy and Jorgenson. The exact role of overidentifying constraints becomes clear, and 3SLS is seen to differ from the other full information estimators in failing to use all overidentifying restrictions in forming the instruments. While this difference vanishes asymptotically, it may be of importance in finite samples where the constraints can be expected to hold only approximately.

The instrumental variable interpretation also provides an algorithm (called "Durbin's Method" by Malinvaud) to compute the FIML estimates. This algorithm is tested on Klein Model I and provides acceptable estimates. These estimates are compared to the 3SLS and FIIV estimates, using 3SLS to provide the initial consistent estimates. The algorithm's main shortcoming—lack of the uphill property—is discussed and a technique is proposed to overcome this problem by a positive definite approximation when the $(\overline{W}'X)$ matrix is not definite. This altered "Durbin's Method" may prove computationally valuable since it has the uphill property and does not require computation of second derivatives. Further experience is required before a judgement can be made in its computational efficiency.

For future research the greatest need is the extension of asymptotically *efficient* methods to nonlinear models. In Hausman [3]. I propose two new estimators. nonlinear 3SLS and a nonlinear instrumental variable estimator for the special case of nonlinearity in the parameters. This use covers the common situations of serial correlation and partial adjustment models. Further work needs to be done to find efficient estimators for the case of significant nonlinearity in the variables. A Gauss–Newton procedure like equation (30) seems promising for FIML but will be more complicated because the Jacobian is not constant as in the linear case and second derivates are therefore involved. As FIML will still be time-consuming to compute in the nonlinear problem, less time-consuming nonlinear methods would permit convenient full-information system estimation of non-linear models currently in use. Then the econometric model builder could use efficient estimates to test his various models.

Massachusetts Institute of Technology and NBER Computer Research Center

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, F. J.. "A Simplified Structural Estimator for Large-Scale Econometric Models." *The Australian Journal of Statistics*. 13. 1971. pp. 168–75.
[3] Hausman, J.. "An Instrumental Variable Approach to Full Information Estimators for Linear and Non-Linear Econometric Models." forthcoming in *Econometrica*.
[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.
[11] Zellner. A. and H. Theil. "Three-Stage Least Squares: Simultaneous Estimation of Simultaneous Equations." *Econometrica*. 30. 1962. pp. 54–78.
[12] Hood. W. C. and T. C. Koopmans. *Studies in Econometric Method*. New Haven. 1953.