Instrumental variables interpretations of FIML and nonlinear FIML

Calzolari, Giorgio and Sampoli, Letizia

IBM Scientific Center, Pisa, Italy., Prometeia Calcolo, Bologna, Italy

4 September 1989

Instrumental Variables Interpretations of FIML and Nonlinear FIML

by **Giorgio CALZOLARI** *and* **Letizia SAMPOLI**

Centro Scientifico IBM
Pisa

Prometeia Calcolo
Bologna

1. INTRODUCTION

For a linear system of simultaneous equations, an *instrumental variables* interpretation of the *full information maximum likelihood* estimator was proposed by Durbin (1963, unpublished paper), and by Hausman (1974, 1975). When the covariance matrix of the structural system is completely unrestricted, FIML estimates can be obtained from iterating to convergence a full information instrumental variables procedure like Brundy and Jorgenson's (1971) FIVE method. Instruments for the included endogenous variables are obtained as linear combinations of all the predetermined variables of the system, through the matrix of restricted reduced form coefficients. In other words, in each iteration instruments for the included endogenous variables are obtained from the simultaneous solution of the structural form equations (without error terms and with coefficients set at the previous iteration values).

Extending Hausman's interpretation to the general class of nonlinear systems of simultaneous equations with normal errors and unrestricted covariance matrix, Amemiya (1977) shows that FIML estimates also in this case can be obtained by iterating to convergence an appropriate instrumental variables method. Instruments are obtained from the variables and functions of variables (linear or nonlinear) that appear in the structural equations, properly *purged* of their stochastic component. In the particular case of a linear system, instruments for the included endogenous variables turn out to be obtained by *purging* the variables of their residuals in the restricted reduced form. Therefore the resulting instruments are exactly the same as for Hausman's method.

When zero restrictions are imposed in the covariance matrix, instrumental variables interpretations of FIML are available in the literature for linear models.

If the covariance matrix is diagonal (a case analyzed by Koopmans, Rubin and Leipnik, 1950, and by Rothenberg, 1973), the instrumental variables whose iterated application converges to FIML are illustrated, for example, in Maddala (1981, p.199), or Hausman (1983, p.427). Instruments for the endogenous variables included in one equation are obtained as linear combinations of all the predetermined variables of the system (as before), but also of the estimated residuals of all the other structural equations. The coefficients of the linear combination are such that the same values for the instruments would be obtained by *purging* each endogenous variable included in equation i only of that *part* of its restricted reduced form residual that derives from the i th structural residual.

This interpretation can be extended to the case of a block-diagonal covariance matrix, as shown by Friedmann (1985). We can linearly combine in addition to the predetermined variables of the system, also the estimated residuals of all the structural equations uncorrelated with the given one. Again, this is the same as *purging* the included endogenous variables of the structural residuals of all the equations that are correlated with the given one.

Somehow more complicated is the case of zero restrictions in the covariance matrix, which do not imply a diagonal or block-diagonal structure. [1] This is the case analyzed by Hausman, Newey and Taylor (1987). Again the result is

(1) An elegant analysis of the different implications of zero restrictions, when they imply or not a diagonal or block-diagonal structure of the covariance matrix, is made in Friedmann (1985) by means of the concept of *connected versus correlated* error terms.

that the efficient instruments for the endogenous variables included in one equation should be obtained by linearly combining not only the predetermined variables of the system, but also the estimated residuals of all the structural equations that are uncorrelated with the given one. The complication with respect to the previous cases is in the coefficients (or weights) of the linear combination of residuals: these coefficients do not have the simple expression they have in the previous cases.

All the methods summarized above show that we can get FIML estimates by iterating to convergence an instrumental variables formula that is perfectly consistent with the intuitive *textbook-type* interpretation of efficient instruments: instruments for an equation must be uncorrelated with the error term of the equation, but at the same time must have the highest correlation with the explanatory variables (e.g. Fomby, Hill and Johnson, 1984, p.258).

However, if our purpose is to obtain FIML from iterating to convergence some full information instrumental variables, the intuitive *textbook-type* interpretation of the efficient instruments is not necessarily helpful, and can be too restrictive.

The purpose of this paper is to show that, in the *full information* framework, there is a much wider flexibility in the choice of the instruments. Instruments can be efficient [2] even if, against intuition, they are not *purged enough* of correlation with the error term: for example, the instruments for the endogenous variables or functions of endogenous variables included in one equation do not need to be *purged* of the residuals of equations that are correlated with the given one. Viceversa, instruments can be efficient even if they are *purged too much*: for example, if there are zero covariance restrictions, instruments may be *purged* also of the estimated residuals of equations uncorrelated with the given one.

(2) With efficient instruments we simply mean that their iterated application, if converging, leads to FIML; we do not mean that their application, starting from any consistent estimate, attains asymptotic efficiency after the first iteration. As explained in Amemiya (1977), in the general nonlinear system the instrumental variables do not attain efficiency after the first iteration, unless the initial estimate is already asymptotically efficient. The same happens for the linear system, with diagonal or block-diagonal covariance matrix, as shown in Friedmann (1985).

We show these results for general nonlinear systems of simultaneous equations, with additive random error terms which are independently and identically distributed like multivariate normal. Then the linear system follows as a particular case. We first show why there is a wide flexibility in the choice of instruments, then exemplify some extreme consequences with a list of examples.

2. NOTATIONS AND ESTIMATION

We follow the notations and implicitly accept the assumptions in Amemiya (1977, 1982). Let the simultaneous equation model be represented as

$$f_{i,t} = f_i(y_t, x_t, a_i) = u_{i,t} \qquad \begin{cases} i = 1,2,\dots,n \\ t = 1,2,\dots,T \end{cases} \tag{1}$$

where y_t is the $n \times 1$ vector of endogenous variables at time t, x_t is the vector of predetermined variables at time t and a_i is the $k_i \times 1$ vector of unknown structural coefficients in the i th equation. The $n \times 1$ vector of random error terms at time t, $u_t = (u_{1,t}, u_{2,t}, \dots, u_{n,t})'$, is assumed to be independently and identically distributed as $N(0, \Sigma)$; the $n \times n$ matrix Σ is symmetric and positive definite, and can be either unrestricted or subject to linear restrictions (in particular, zero restrictions like $\sigma_{i,j} = \sigma_{j,i} = 0$). The complete $\sum^{n} k_i \times 1$ vector of unknown structural coefficients of the system will be indicated as $a = (a_1', a_2', \dots, a_n')'$. Identities, if any, have been previously substituted out of the system.

The log-likelihood of the whole sample can be expressed as

$$L_T = -\frac{T}{2} \log |\Sigma| + \sum_{t=1}^{T} \log \left| \left| \frac{\partial f_t}{\partial y_t'} \right| \right| - \frac{1}{2} \sum_{t=1}^{T} f_t' \Sigma^{-1} f_t \tag{2}$$

where $f_t = (f_{1,t}, f_{2,t}, \dots, f_{n,t})' = u_t$ and the Jacobian determinant $|\partial f_t / \partial y_t'|$ is taken in absolute value. We define, for the i th equation, $g_{i,t} = \partial f_{i,t} / \partial a_i$, which is a column vector with the same length as a_i, and is a function of y_t, x_t and a.

If, for any u_t, the system has a unique solution for y_t, we can regard $g_{i,t}$ as a function of u_t, x_t and a. We need, in particular, its derivative with respect to $u_{i,t}$, which is

$$\frac{\partial g_{i,t}}{\partial u_{i,t}} = \left(\frac{\partial g_{i,t}}{\partial y_t'} \right) \left(\frac{\partial f_t}{\partial y_t'} \right)^{\cdot,i} \tag{3}$$

where $(\partial f_t / \partial y_t')^{\cdot,i}$ is the i th column of the inverse of the $n \times n$ Jacobian matrix $(\partial f_t / \partial y_t')$.

Differentiating the log-likelihood (2) with respect to the coefficients of the i th equation we get

$$\frac{\partial L_T}{\partial a_i} = \sum_{t=1}^{T} \frac{\partial g_{i,t}}{\partial u_{i,t}} - \sum_{t=1}^{T} g_{i,t} f_t' \sigma^{\cdot,i} \tag{4}$$

where $\sigma^{\cdot,i}$ is the i th column of Σ^{-1}.

We now introduce the $T \times n$ matrix of the structural form functions over time, F, whose t,i th element is $f_i(y_t, x_t, a_i)$ (its t th row is f_t'), and the $T \times k_i$ matrix G_i, whose t th row is $g_{i,t}'$, so we can write

$$\sum_{t=1}^{T} g_{i,t} f_t' = G_i' F \tag{5}$$

We also indicate with $S = F'F/T$ the sample covariance matrix of the structural residuals $(n \times n)$. If $i_{\cdot,i}$ is the i th column of the $n \times n$ unit matrix I_n, we build for the i th equation an $n \times n$ matrix Q_i, which has $i_{\cdot,i}$ as i th column, while the remaining $n - 1$ columns are completely arbitrary

$$Q_i = [\times, \times, \dots, i_{\cdot,i}, \dots, \times] = \begin{bmatrix} \times & \times & . & . & 0 & . & . & \times \\ \times & \times & . & . & 0 & . & . & \times \\ . & . & . & . & . & . & . & . \\ . & . & . & . & 1 & . & . & . \\ . & . & . & . & . & . & . & . \\ \times & \times & . & . & 0 & . & . & \times \end{bmatrix} \tag{6}$$

Given Q_i, we build the $n \times n$ matrix P_i as

$$P_i = Q_i \Sigma S^{-1} = Q_i \Sigma \left(\frac{F'F}{T} \right)^{-1} \tag{7}$$

Whatever choice we make for the arbitrary elements of Q_i (whose number is $n^2 - n$ for any i), we obviously get

$$P_i \, S \, \sigma^{\cdot,i} = P_i \, \frac{F'F}{T} \, \sigma^{\cdot,i} = i_{\cdot,i} \tag{8}$$

and therefore

$$\frac{\partial g_{i,t}}{\partial u_{i,t}} = \frac{\partial g_{i,t}}{\partial u'_t} \, i_{\cdot,i} = \frac{\partial g_{i,t}}{\partial u'_t} \, P_i \, \frac{F'F}{T} \, \sigma^{\cdot,i} \tag{9}$$

Substituting equations (5) and (9) into (4) we get

$$\begin{aligned} \frac{\partial L_T}{\partial a_i} &= \sum_{t=1}^{T} \frac{\partial g_{i,t}}{\partial u'_t} \, P_i \, \frac{F'F}{T} \, \sigma^{\cdot,i} - G'_i F \sigma^{\cdot,i} \\ &= - \left[G'_i - \left(\frac{1}{T} \sum_{t=1}^{T} \frac{\partial g_{i,t}}{\partial u'_t} \right) P_i \, F' \right] F \sigma^{\cdot,i} \end{aligned} \tag{10}$$

Defining the matrix of instruments for the i th equation as

$$\hat{G}'_i = G'_i - \left(\frac{1}{T} \sum_{t=1}^{T} \frac{\partial g_{i,t}}{\partial u'_t} \right) P_i \, F' \tag{11}$$

then building as in Amemiya (1977, p.962) the $nT \times \sum_{i=1}^{n} k_i$ block-diagonal matrices $G = diag\,(G_1, ..., G_n)$ and $\hat{G} = diag\,(\hat{G}_1, ..., \hat{G}_n)$ and stacking the equations (10) for $i = 1, 2, ..., n$, we get for the gradient the expression

$$\frac{\partial L_T}{\partial a} = - \hat{G}' \, (\Sigma^{-1} \otimes I_T) \, vec\, F \tag{12}$$

Setting (12) to zero gives a nonlinear equation that can be solved, as in Amemiya (1977, eq.4.8), iterating to convergence [3] the instrumental variables formula

$$\hat{a}^{(m+1)} = \hat{a}^{(m)} - \left[\hat{G}' \, (\Sigma^{-1} \otimes I_T) \, G \right]^{-1} \hat{G}' \, (\Sigma^{-1} \otimes I_T) \, vec\, F \tag{13}$$

[3] Iterations stop when $\hat{G}' (\Sigma^{-1} \otimes I_T)\, vec\, F = 0$ but, of course, convergence is not guaranteed. The joint use of the instrumental variables formula and of the *line search* maximization of the likelihood, analogous to Dagenais (1978) or Calzolari, Panattoni and Weihs (1987), is helpful since it improves considerably the probability and the speed of convergence.

To make iteration (13) feasible in practice, we must replace G, $\hat{G}$, F and Σ with values computed at the coefficients estimate of the m th iteration, $\hat{a}^{(m)}$. For G and $\hat{G}$ this means that we must compute derivatives at $\hat{a}^{(m)}$; for F we must take the matrix of the m th iteration structural residuals; finally, for Σ, we must take the appropriate estimate, whose expression changes considerably if we consider the unrestricted case or the case where restrictions are imposed.

Equation (11) defines the matrices of efficient instruments with a wide degree of arbitrariness given the presence of the matrices P_i, whose only requirement is that they must satisfy the condition given by equation (8).

3. THE INSTRUMENTAL VARIABLES

Interpreting the matrix G_i is particularly simple for an equation that is linear in the coefficients, even if nonlinear in the variables. The t th column of G_i' is the $k_i \times 1$ vector $g_{i,t} = \partial f_{i,t}/\partial a_i$, and $-g_{i,t}$ is nothing but the vector of values at time t of the explanatory variables and functions of variables included in the i th equation. Therefore, $-G_i$ is simply the $T \times k_i$ matrix of *regressors* of the i th equation that can be normalized as

$$y_i = - G_i \, a_i + u_i \tag{14}$$

The $k_i \times 1$ vector of instruments at time t, $\hat{g}_{i,t}$ (the t th column of $\hat{G}_i'$) is obtained by *purging* $g_{i,t}$ of $(T^{-1} \sum_{t=1}^{T} \partial g_{i,t}/\partial u_t')P_i f_t$

$$\hat{g}_{i,t} = g_{i,t} - \left(\frac{1}{T} \sum_{t=1}^{T} \frac{\partial g_{i,t}}{\partial u_t'} \right) P_i f_t \tag{15}$$

For an element of $g_{i,t}$ that is not function of current endogenous variables, the first derivatives with respect to the error terms are zero, thus implying a row of zeroes in the $k_i \times n$ matrix $\partial g_{i,t}/\partial u_t'$. This element of $g_{i,t}$ remains therefore unchanged in $\hat{g}_{i,t}$, as expected, whatever the choice of the arbitrary elements of Q_i (and therefore P_i). For an element of $g_{i,t}$ that is function of current endogenous variables, the choice of the arbitrary elements of Q_i can considerably affect the way in which it will be *purged* of the correlation with the error terms. The analysis of some extreme cases will better illustrate this point.

3.1. If Σ is unrestricted, then solving the first order conditions $\partial L_T/\partial\Sigma^{-1} = 0$ leads to the well known estimate of the covariance matrix $\hat{\Sigma} = F'F/T = S$ that must be substituted into equation (13) at each iteration. If we choose $Q_i = I_n$ (the $n \times n$ unit matrix, the same for all equations), then from equation (7) we get $P_i = I_n$ and, therefore

$$\hat{g}_{i,t} = g_{i,t} - \left(\frac{1}{T} \sum_{t=1}^{T} \frac{\partial g_{i,t}}{\partial u_t'} \right) f_t \tag{16}$$

that is Amemiya's (1977) instruments.

3.2. If the model is linear, we can write

$$f(y_t, x_t, a) = f_t = B y_t + \Gamma x_t = u_t \tag{17}$$

If Σ is unrestricted (so that $\hat{\Sigma} = S = F'F/T$) and y_j is an explanatory variable included in equation i (so that $y_{j,t}$ is an element of the vector $-g_{i,t}$), then we have $(T^{-1} \sum^{T} \partial y_{j,t}/\partial u_t') = \partial y_{j,t}/\partial u_t' = b^{j,\cdot}$, the j th row of B^{-1}. If we choose $Q_i = I_n$ (and therefore $P_i = I_n$), then we replace $y_{j,t}$ with the instrument $\hat{y}_{j,t} = y_{j,t} - b^{j,\cdot} u_t$, that is we *purge* $y_{j,t}$ of its restricted reduced form residual (which is $b^{j,\cdot} u_t$, and therefore is a linear combination of all the structural residuals, coefficients being elements of B^{-1}). This is obviously equal to the value of $y_{j,t}$ calculated from the simultaneous solution of the model (17, that is considering all the overidentifying coefficients restrictions), after setting to zero the error terms: $\hat{y}_{j,t} = - b^{j,\cdot} \Gamma x_t$, where coefficients are set to the previous iteration values, as in Hausman (1974, 1975).

3.3. Still for the example 3.2 (and 3.1, with simple modifications), we may set to zero all the $n - 1$ arbitrary columns of Q_i ($n^2 - n$ elements), choosing

$$Q_i = [0, 0, \ldots, 0, i_{\cdot,i}, 0, \ldots, 0] \tag{18}$$

and therefore we get from equation (7), being Σ unrestricted, $P_i = Q_i$ and

$$P_i f_t = [0, 0, \ldots, 0, u_{i,t}, 0, \ldots, 0]' \tag{19}$$

For the instrument replacing in equation i the explanatory endogenous variable $y_{j,t}$ we have, therefore

$$\hat{y}_{j,t} = y_{j,t} - b^{j,i} u_{i,t} = - b^{j,\cdot} \Gamma x_t + \sum_{k \neq i} b^{j,k} u_{k,t} \tag{20}$$

Surprisingly we get, even if Σ is unrestricted, the instrument that would *traditionally* be used if Σ is diagonal, as in Maddala (1981, p.199), or Hausman (1983, p.427). The endogenous regressor in equation i is *purged* only of the i th equation residual. Viewing this from the other point, the structural residuals of all equations but the i th (even if correlated with the i th residual) enter the linear combination that produces the efficient instrument for equation i.

3.4. Still for examples 3.1 and 3.2, we may set to zero only some rows and columns of Q_i, and take the others from the unit matrix. For example, if $i > 2$, we may take

$$Q_i = [0, 0, i_{\cdot,3}, i_{\cdot,4}, \ldots, i_{\cdot,i}, \ldots, i_{\cdot,n}] \tag{21}$$

then it will be $P_i = Q_i$ and

$$\hat{g}_{i,t} = g_{i,t} - \sum_{k>2} \left(\frac{1}{T} \sum_{t=1}^{T} \frac{\partial g_{i,t}}{\partial u_{k,t}} \right) f_{k,t} \tag{22}$$

for the nonlinear model. If the model is linear

$$\hat{y}_{j,t} = y_{j,t} - \sum_{k>2} b^{j,k} u_{k,t} = - b^{j,\cdot} \Gamma x_t + b^{j,1} u_{1,t} + b^{j,2} u_{2,t} \tag{23}$$

The instruments for the endogenous regressors of equation i ($i > 2$) are not *purged* of the residuals of equations 1 and 2. We get, even if Σ is unrestricted, the instruments that would *traditionally* be used if Σ is block-diagonal, the first block being given by the first two equations (e.g. Friedmann, 1985). For the linear case we can also say that equation (23) uses $u_{1,t}$ and $u_{2,t}$ as extra instrumental variables for estimation, even in absence of covariance restrictions.

3.5. The model is linear, as in example 3.2, but Σ is diagonal (and so is Σ^{-1}). Differentiating the log-likelihood (2) with respect to the diagonal elements of Σ^{-1} and equating derivatives to zero we get

$$\hat{\Sigma} = diag\,[s_{1,1}, \dots, s_{n,n}] = diag\left[\frac{1}{T}\sum_{t=1}^{T} f_{1,t}^2, \dots, \frac{1}{T}\sum_{t=1}^{T} f_{n,t}^2\right] \tag{24}$$

If we choose for Q_i the unit matrix, then P_i will generally be a full matrix with fairly complicated elements (only asymptotically it would be equal to the unit matrix I_n , if the coefficients estimate is consistent). This implies that endogenous regressors will be partially *purged* of a linear combination of the residuals of all the equations (only partially, because coefficients of the linear combination are not equal to the elements of B^{-1}). Viewing this from the other point, in the linear case all residuals enter, with complicated coefficients, into the linear combination that forms the efficient instruments.

3.6. In the same case of example 3.5 we choose an $n \times n$ matrix Q_i whose rows are all zeroes except the i th row, whose elements are all $\neq 0$: the i,j th element is $s_{i,j} / s_{j,j} = \sum^{T} f_{i,t} f_{j,t} / \sum^{T} f_{j,t}^2$ (the i,i th element is 1, so the condition that the i th column of Q_i^{-1} must be equal to $i_{.,i}$ is satisfied). With this choice of Q_i we get the same $P_i f_t$ as in equation (19): any endogenous regressor in equation i is *purged* only of the ith equation residual, which now is the expected result, Σ being diagonal. However the instruments are exactly the same as in example 3.3, where Σ had no restrictions.

3.7. The model is linear and Σ is two-blocks block-diagonal, the first two equations giving the first block. From the first order conditions we get that $\hat{\Sigma}$ is obtained from $F'F/T$ by simply cancelling the two off-diagonal blocks. If we choose for Q_i the unit matrix, then P_i will generally be a full matrix with fairly complicated elements. This implies that an endogenous regressor of equation i will be replaced by a linear combination of the predetermined variables of the system and of all residuals, including those of the first two equations.

3.8. For the model of example 3.7, we partition the $n \times n$ matrix S

$$\frac{F'F}{T} = S = \begin{bmatrix} S_{1,1} & S_{1,2} \\ S_{2,1} & S_{2,2} \end{bmatrix} \tag{25}$$

where $S_{1,1}$ has dimensions 2×2 . We indicate with $\Sigma_{1,1}$ the first 2×2 block of the block-diagonal Σ and with $\Sigma^{1,1} = (\Sigma_{1,1})^{-1}$ its inverse. For the i th equation ($i > 2$) we may choose the following matrix Q_i

$$Q_i = \left[\begin{array}{cc|c} 0 & 0 & 0 \; . \; . \; . \; 0 \\ 0 & 0 & 0 \; . \; . \; . \; 0 \\ \hline \multicolumn{2}{c|}{S_{2,1}\Sigma^{1,1}} & I_{n-2} \end{array}\right] , \tag{26}$$

so that

$$P_i = Q_i \Sigma S^{-1} = [0, 0, i_{.,3}, i_{.,4}, \dots, i_{.,i}, \dots, i_{.,n}] \tag{27}$$

as for example 3.4. Instruments for the endogenous regressors of equation i ($i > 2$) are given by a linear combination of the predetermined variables of the system and of the residuals of equations 1 and 2 (which are uncorrelated with equation i). This now is the expected result, as in Friedmann (1985), Σ being block-diagonal. The instruments, however, are exactly the same as in equation (23), where Σ was unrestricted.

3.9. The model is linear and some linear restrictions are imposed on the Σ matrix. These restrictions can also be of the zero-type, but without implying a diagonal or block-diagonal structure. At each iteration, the estimate of Σ can be obtained from the first order conditions related only to the unrestricted elements of Σ, as in Hausman, Newey and Taylor (1987). Alternatively, $\hat{\Sigma}$ can be obtained from the constrained maximization of the log-likelihood (2), using the Lagrange multipliers, as in Friedmann (1985). In any case we get an estimate $\hat{\Sigma}$ generally different from $S = F'F/T$ (only asymptotically they will be equal, if coefficients are consistently estimated), so that $\hat{\Sigma}S^{-1}$ is generally different from the unit matrix. Therefore, if we choose $Q_i = I_n$, we get a fairly complicated linear combination of the structural residuals of all equations, coefficients being generally different from the elements of B^{-1}. This implies that instruments are only partially *purged* of the correlation with all the error terms or, which is the same, instruments involve a linear combination of all the structural residuals. Different linear combinations, but still involving structural residuals of all equations, will be obtained by choosing Q_i as in equations (18) or (21).

3.10. Let us suppose that, in example 3.9, the endogenous variable $y_{j,t}$ is a regressor of equation i. Suppose that we can choose a matrix P_i which has the following form

$$P_i = [w, 0, 0, \ldots, 0] + I_n \tag{28}$$

where the $n \times 1$ vector $w = [w_1, \ldots, w_n]'$ is not identically zero. Then

$$P_i f_t = \begin{bmatrix} w_1 f_{1,t} + f_{1,t} \\ w_2 f_{1,t} + f_{2,t} \\ \cdot \quad \cdot \quad \cdot \\ w_i f_{1,t} + f_{i,t} \\ \cdot \quad \cdot \quad \cdot \\ w_n f_{1,t} + f_{n,t} \end{bmatrix} \tag{29}$$

and the instrument replacing $y_{j,t}$ in equation i is, therefore

$$\hat{y}_{j,t} = y_{j,t} - \sum_{k=1}^{n} b^{j,k}(w_k f_{1,t} + f_{k,t}) = -b^{j,\cdot}\,\Gamma x_t - b^{j,\cdot}\, w f_{1,t} \tag{30}$$

This is a linear combination of all the predetermined variables of the system (with the usual reduced form coefficients) and of the residual of the first structural equation (with a complicated coefficient given by $-b^{j,\cdot}w$). If $i > 1$ and the linear restriction is $\sigma_{i,1} = \sigma_{1,i} = 0$, this is the result in Hausman, Newey and Taylor (1987, eq.3.11). It remains to prove that equation (28) is an acceptable choice for the matrix P_i. If $s_{1,\cdot}\sigma^{\cdot,i} \neq 0$ (which, for $i > 1$, is made possible by the restrictions on Σ), and we choose $w = (i_{\cdot,i} - S\sigma^{\cdot,i}) / (s_{1,\cdot}\sigma^{\cdot,i})$, the resulting P_i satisfies equation (8) and therefore produces efficient instruments.

3.11. The only condition needed in the last example, in order to ensure that an instrument with the form of equation (30) is efficient for equation i, is $s_{1,\cdot}\sigma^{\cdot,i} \neq 0$. If $i = 1$, such a condition is obviously satisfied when Σ is either restricted or not. The case of Σ unrestricted is not surprising, because it gives $w = 0$, thus producing the traditional instruments of example 3.2. On the contrary, the risult is quite surprising if Σ is diagonal. In such a case, in fact, $s_{1,\cdot}\sigma^{\cdot,1} = 1$, and the vector w is

$$w = i_{\cdot,1} - S\sigma^{\cdot,1} = i_{\cdot,1} - \frac{s_{\cdot,1}}{s_{1,1}} \tag{31}$$

While $w_1 = 0$, the other elements of w are generally $\neq 0$. Therefore, if $y_{j,t}$ is a regressor of the first equation, the efficient instrument given by equation (30) involves the usual linear combination of all the predetermined variables of the system (with the restricted reduced form coefficients), plus the structural residual of the first equation (with a complicate coefficient given by $-b^{j,\cdot}w = \left[\sum_{k=2}^{n} b^{j,k}s_{k,1}\right] \Big/ s_{1,1}$). The traditional case (example 3.6) used exactly the reverse: the structural residuals of all the equations except the first.

4. CONCLUSION

The usual interpretations of efficient instruments should be carefully revised for the *full information* case. There are sentences often recurring in the literature that should be reconsidered. One is, for example, that efficient instruments for an equation must be uncorrelated with the error term of the equation, but at the same time must have the highest correlation with the explanatory variables. Another one is that, when there are covariance restrictions, FIML uses the uncorrelated structural residuals to perform better instruments. In the *full information* framework these sentences are too restrictive. We have shown in this paper that efficient instruments can be chosen with a much greater flexibility.

REFERENCES

Amemiya, T. (1977): "The Maximum Likelihood and the Nonlinear Three-Stage Least Squares Estimator in the General Nonlinear Simultaneous Equation Model", *Econometrica* 45, 955-968.

Amemiya, T. (1982): "Correction to a Lemma", *Econometrica* 50, 1325-1328.

Brundy, J. M., and D. W. Jorgenson (1971): "Efficient Estimation of Simultaneous Equations by Instrumental Variables", *The Review of Economics and Statistics* 53, 207-224.

Calzolari, G., L. Panattoni, and C. Weihs (1987): "Computational Efficiency of FIML Estimation", *Journal of Econometrics* 36, 299-310.

Dagenais, M. G. (1978): "The Computation of FIML Estimates as Iterative Generalized Least Squares Estimates in Linear and Nonlinear Simultaneous Equations Models", *Econometrica* 46, 1351-1362.

Durbin, J. (1963): "Maximum Likelihood Estimation of the Parameters of a System of Simultaneous Regression Equations". London School of Economics: discussion paper presented at *The European Meeting of the Econometric Society*, Copenhagen.

Fomby, T. B., R. C. Hill, and S. R. Johnson (1984): *Advanced Econometric Methods*. New York: Springer-Verlag.

Friedmann, R. (1985): "Multivariate Predictions from Structural Econometric Models with Covariance Restrictions". University of Bielefeld: discussion paper presented at *The Fifth International Symposium on Forecasting*, Montreal.

Hausman, J. A. (1974): "Full Information Instrumental Variables Estimation of Simultaneous Equations Systems", *Annals of Economic and Social Measurement* 3, 641-652.

Hausman, J. A. (1975): "An Instrumental Variable Approach to Full Information Estimators for Linear and Certain Nonlinear Econometric Models", *Econometrica* 43, 727-738.

Hausman, J. A. (1983): "Specification and Estimation of Simultaneous Equation Models", in *Handbook of Econometrics*, ed. by Z. Griliches and M. D. Intriligator. Amsterdam: North-Holland Publishing Company, Vol.I, 391-448.

Hausman, J. A., W. K. Newey, and W. E. Taylor (1987): "Efficient Estimation and Identification of Simulaneous Equation Models with Covariance Restrictions", *Econometrica* 55, 849-874.

Koopmans, T. C., H. Rubin, and R. B. Leipnik (1950): "Measuring the Equation Systems of Dynamic Economics", in *Statistical Inference in Dynamic Economic Models*, ed. by T. C. Koopmans. New York: John Wiley & Sons, Cowles Commission Monograph No.10, 53-237.

Maddala, G. S. (1981): "Statistical Inference in Relation to the Size of the Model", in *Large-Scale Macro-Econometric Models*, ed. by J. Kmenta and J. B. Ramsey. Amsterdam: North-Holland Publishing Company, 191-218.

Rothenberg, T. J. (1973): *Efficient Estimation with A Priori Information*. New Haven: Yale University Press, Cowles Foundation Monograph No.23.