STATISTICA, anno LXVIII, nn. 3-4, 2008

LEAST ORTHOGONAL DISTANCE ESTIMATOR OF STRUCTURAL PARAMETERS IN SIMULTANEOUS EQUATION MODELS

L. Pieraccini, A. Naccarato

1. INTRODUCTION

The starting idea whose development has given rise to Least Orthogonal Distance Estimator (LODE) can be traced back to the work of the senior coauthor (Pieraccini, 1969), in which 2SLS were obtained as generalized least squares estimator applied to the system of the so called identifying restrictions; the result was afterwards extended to 3SLS (Pieraccini, 1978). With this in mind and making reference to the work of K. Pearson (1901), the LODE method of estimation has been derived under the consideration that the over-identifying restrictions are nothing else but linear relations between variables affected by error.

In its first formalization, LODE structural parameters' estimation of endogenous and exogenous variables were not treated symmetrically: parameters of endogenous variables were derived minimizing a quadratic form obtained by the over identifying restrictions while those of exogenous ones were obtained in the same way as in LIML (Pieraccini, 1983, 1988, 1992). Taking into account the results of many simulation experiments (Cau, 1990; Sbrana, 2001; Zurlo, 2006) a modified version of Limited Information LODE is here presented in which structural parameters' estimates both for endogenous and exogenous variables are directly derived from the whole system of identifying restrctions.

Furthermore, two recent contributions have increased the interest about the method: its extension to the case of Full Information (Naccarato, 2007), which has shown the versatility of the method to cope with simultaneous estimation of the whole system's structural parameters, and a very extensive simulation experiment (Naccarato and Zurlo, 2007) which has confirmed the Limited Information (LI) method's good performances and shown a quite substiantial improvement for its Full Information (FI) version.

In this paper we give a complete illustration of the new version of LODE method, the principles upon which the method is based and the formal derivation of its properties.

The structure of this paper is as follows: after briefly reviewing the simultaneous equation models to establish notation (section 2), conditions for identification are presented in the new context (section 3) and Limited Information LODE is derived (section 4). The next three paragraphs are devoted to FI LODE derivation (section 5), to the estimation of the matrix of variances and covariances for error components (section 6) and to the consistency of FI LODE (section 7). Finally (section 8), few words of conclusion, with a short summary of the more important results of the recent simulation experiment, close the contribution.

2. SIMULTANEOUS EQUATIONS MODELS

Making use of standard notations, the structural form of a simultaneous equations system can be written as follows:

$$\underset{n,m}{Y}\underset{m,m}{\Gamma}+\underset{n,k}{X}\underset{k,m}{B}+\underset{n,m}{U}=\underset{n,m}{0}, \tag{1}$$

where $\underset{n,m}{Y}$ and $\underset{n,k}{X}$ are the matrices of the observation on the m endogenous variables y and the k exogenous variables x; $\underset{n,m}{U}$ is the matrix of disturbances. Γ and B are the $m\times m$ and $k\times m$ matrices of structural parameters of endogenous variables and exogenous variables.

For the matrix U standard hypotheses are supposed to hold:

$$\begin{aligned} &E(vecU)=0\\ &E(vecU(vecU)^T)=\Omega\otimes I, \end{aligned} \tag{2}$$

where $\underset{m,m}{\Omega}$ is the constant over observations variance-covariance matrix. Furthermore, the following assumptions

$$\begin{aligned} &\underset{n\to\infty}{p\lim}\frac{1}{n}U^TU=\Omega\\ &\underset{n\to\infty}{p\lim}\frac{1}{n}X^TU=\underset{k,m}{0}\\ &\underset{n\to\infty}{p\lim}\frac{1}{n}X^TX=\underset{k,k}{\Sigma_x}, \end{aligned} \tag{3}$$

are generally made.

Under non singularity condition for Γ, the reduced form of system (1) is derived as

$$\underset{n,m}{Y}=\underset{n,k}{X}\underset{k,m}{\Pi}+\underset{n,m}{V}, \tag{4}$$

in which

$$\underset{k,m}{\Pi} = -\underset{k,m}{B}\,\underset{m,m}{\Gamma^{-1}}$$
$$\underset{n,m}{V} = -\underset{n,m}{U}\,\underset{m,m}{\Gamma^{-1}}, \tag{5}$$

where in the second equation it is

$$E(V) = 0$$
$$E(V^T V) = n(\Gamma^{-1})^T \Omega \Gamma^{-1}, \tag{6}$$

Post-multiplying by Γ the first of equations (5) we obtain

$$\underset{k,m}{\Pi}\ \underset{m,n}{\Gamma} = -\underset{k,m}{B}, \tag{7}$$

that represent the link between reduced and structural form parameters.

Since (7) is a system of equations with $m+k$ unknowns and k equations, usual exclusion constrains have to be introduced to find solutions with respect to Γ and B in terms of Π.

If endogenous and exogenous variables are not all included in *i-th* equation, it is possibile to consider the following partition of the overall matrix of endogenous variables with respect to the included and excluded variables:

$$\underset{n,m}{Y} = \begin{bmatrix} \underset{n,m_{1i}}{Y_{1i}} & \vdots & \underset{n,m_{2i}}{Y_{2i}} \end{bmatrix},$$

where the first m_{1i} colums refer – as usual – to endogenous variables included in the *i-th* equation while the last m_{2i} colums refer to endogenous variables not included in it. In the same way the vector of Γ's in *i-th* equation can be reordered as

$$\underset{m,1}{\Gamma_i} = \begin{bmatrix} \underset{m_{1i},1}{\Gamma_{1i}} \\ \cdots \\ \underset{m_{2i},1}{0} \end{bmatrix},$$

where the first m_{1i} elements of Γ_i are the coefficients of endogenous variables included in *i-th* equation while the remaining m_{2i} elements equal to 0 are related to the excluded ones. Notice that when defining the vector Γ_i no normalization rule has yet been introduced.

Similarly the following partition of the exogenous variables' matrix is considered

$$\underset{n,k}{X} = \left[\underset{n,k_{1i}}{X_{1i}} \;\vdots\; \underset{n,k_{2i}}{X_{2i}} \right],$$

where $\underset{n,k_{1i}}{X_{1i}}$ and $\underset{n,k_{2i}}{X_{2i}}$ are the sub-matrices corresponding to the exogenous variables included and excluded from the *i-th* equation respectively. Accordingly the vector of parameters is defined as

$$\underset{k,1}{B_i} = \begin{bmatrix} \underset{k_{1i},1}{B_{1i}} \\ \cdots \\ \underset{k_{2i},1}{0} \end{bmatrix},$$

where the first k_{1i} parameters are related to the exogenous variables included in the *i-th* equation while the last k_{2i} are zeros and they are related to the excluded ones.

With these partitions the *i-th* structural equation is written in the following form

$$Y_{1i}\Gamma_{1i} + X_{1i}B_{1i} + U_i = 0 .$$

The reordering of endogenous and exogenous variables in the structural form induces a new ordering in the reduced form as follows

$$\begin{aligned} Y_{1i} &= X_{1i}\Pi_{11}^{i} + X_{2i}\Pi_{12}^{i} + V_{1i} \\ Y_{2i} &= X_{1i}\Pi_{21}^{i} + X_{2i}\Pi_{22}^{i} + V_{2i} , \end{aligned}$$

or

$$[Y_{1i} \;\vdots\; Y_{2i}] = [X_{1i} \;\vdots\; X_{2i}] \begin{bmatrix} \Pi_{11}^{i} & \Pi_{21}^{i} \\ \Pi_{12}^{i} & \Pi_{22}^{i} \end{bmatrix} + [V_{1i} \;\vdots\; V_{2i}],$$

so that the Π matrix comes out to be partitioned in four blocks related to endogenous and exogenous variables both included and excluded from the equation.

In the same way, multiplying both sides of the second equation of (5) by the matrix Γ it is

$$V\Gamma = -U ,$$

and for *i-th* equation

$$V\Gamma_i = -U_i$$

i.e.

$$V_{1i}\Gamma_{1i} = -U_i\,,$$

which will represent the relationship between RF and SF disturbances.

3. CONDITION FOR IDENTIFICATION

Usually, rank conditions for the identification of a simultaneous equation system, as well as order conditions, are obtained after applying the normalization rule: in our case, this will not happen so that identifiability conditions have to be redefined. With respect to the *i-th* structural equation the system of relation (7) can be written as

$$\begin{bmatrix} \underset{k_{1i},m_{1i}}{\Pi_{1i}} & I_{k_{1i}} \\ \underset{k_{2i},m_{1i}}{\Pi_{2i}} & \underset{k_{2i},k_{1i}}{0} \end{bmatrix} \begin{bmatrix} \underset{m_{1i},1}{\Gamma_{1i}} \\ \underset{k_{1i},1}{B_{1i}} \end{bmatrix} = 0\,, \tag{8}$$

and setting

$$\underset{m_{1i}+k_{1i},m_{1i}+k_{1i}}{\Pi_*^i} = \begin{bmatrix} \underset{k_{1i},m_{1i}}{\Pi_{1i}} & I_{k_{1i}} \\ \underset{k_{2i},m_{1i}}{\Pi_{2i}} & \underset{k_{2i},k_{1i}}{0} \end{bmatrix},$$

it can be shown that the rank condition for solving the system (8) with respect to γ's and β's takes the following form

Condition 1 – System (8) admits a unique solution – up to a proportionality constant – if and only if

$$r(\Pi_*^i) = m_{1i} + k_{1i} - 1\,. \tag{9}$$

Condition 2 – The rank of matrix Π_*^i *is equal to* $m_{1i} + k_{1i} - 1$ *if and only if*

$$r(\Pi_{2i}) = m_{1i} - 1\,. \tag{10}$$

Where equation (10) is the usual form for rank condition.

When the parameters Π of reduced form are substituted with their OLS estimates $\hat{\Pi}$, the system (8) becomes

$$\begin{cases} \underset{k_{1i},m_{1i}}{\hat{\Pi}_{1i}} \; \underset{m_{1i},1}{\Gamma_{1i}} + \underset{k_{1i},1}{\mathrm{B}_{1i}} = \underset{k_{1i},1}{\varepsilon_{1i}} \\ \\ \underset{k_{2i},m_{1i}}{\hat{\Pi}_{2i}} \; \underset{m_{1i},1}{\Gamma_{1i}} = \underset{k_{2i},1}{\varepsilon_{2i}}, \end{cases} \tag{11}$$

so that in both equations an error component occurs. In this situation the rank conditions cannot be verified and it is no more possibile to use the rank of Π_{2i} as identification criterion. The so-called "order conditions" have then to be defined.

Condition 3 – If rank condition (9) is satisfied, the matrix $\hat{\Pi}_*^i$ *has to be of order* $k \times (k_{1i} + m_{1i} - 1)$ *where it has to be*

$$k \geq k_{1i} + m_{1i} - 1,$$

i.e.

$$k_{2i} \geq m_{1i} - 1, \tag{12}$$

where equation (12) is the formulation generally used for order condition.

The generalized least squares estimators of γ's and β's parameters of the second the system of equations (11) – after having applied the normalization rule – give rise to Two Stage Least Squares Estimators (Pieraccini, 1969).

4. LIMITED INFORMATION LODE

Defining

$$\underset{k,m_{1i}+k_{1i}}{\hat{\Pi}_*^i} = \begin{bmatrix} \underset{k_{1i},m_{1i}}{\hat{\Pi}_{1i}} & I_{k_{1i}} \\ \underset{k_{2i},m_{1i}}{\hat{\Pi}_{2i}} & \underset{k_{2i},k_{1i}}{0} \end{bmatrix}, \quad \underset{m_{1i}+k_{1i},1}{\delta_i} = \begin{bmatrix} \underset{m_{1i},1}{\Gamma_{1i}} \\ \underset{k_{1i},1}{B_{1i}} \end{bmatrix}, \; \underset{k,1}{\varepsilon_i} = \begin{bmatrix} \underset{k_{1i},1}{\varepsilon_{1i}} \\ \underset{k_{2i},1}{\varepsilon_{2i}} \end{bmatrix},$$

system (12) can be written as

$$\hat{\Pi}_*^i \, \delta_i = \varepsilon_i, \tag{13}$$

or

$$\begin{bmatrix} \hat{\Pi}_{1i} & I_{k_{1i}} \\ \hat{\Pi}_{2i} & 0 \end{bmatrix} \begin{bmatrix} \Gamma_{1i} \\ B_{1i} \end{bmatrix} = \varepsilon_i,$$

Since it is

$$\underset{k,1}{\varepsilon_i} = (X^T X)^{-1} X^T V_{1i} \Gamma_{1i} = (X^T X)^{-1} X^T U_i \,, \tag{14}$$

the variance-covariance matrix of ε_i comes out to be

$$E(\varepsilon_i \varepsilon_i^T) = (X^T X)^{-1} X^T E(U_i U_i^T) X (X^T X)^{-1} = \sigma_i^2 (X^T X)^{-1} \,,$$

which is the variance-covariance matrix of $\hat{\Pi}_i$.

Let us now set

$$(X^T X)^{-1} = T \Lambda^{-1} T^T \,,$$

the matrix Λ being the diagonal matrix of characterstic roots of $X^T X$ and the matrix T the one of characteristic vectors. Defining

$$Q = T \Lambda^{\frac{1}{2}} T^T \,,$$

so that it is

$$QQ = T \Lambda T^T = X^T X \,,$$

Applying to the error component the following transformation

$$\omega_i = Q \varepsilon_i \,, \tag{15}$$

it will be

$$E(\omega_i^T \omega_i) = E(\varepsilon_i^T X^T X \varepsilon_i) = k \sigma_i^2 \,,$$

so that the sample estimate of σ_i^2 will be

$$\hat{\sigma}_i^2 = \tfrac{1}{k} \omega_i^T \omega_i = \tfrac{1}{k} \delta_i^T \hat{\Pi}_*^{i\,T} X^T X \hat{\Pi}_*^i \delta_i \,. \tag{16}$$

The LODE method is based on the minimization of (16) i.e. on finding the vector δ which minimizes the sample residual variance for the *i-th* structural equation.

Since it can be easily shown that

$$\begin{aligned} \underset{m_{1i}+k_{1i},k}{\hat{\Pi}_*^{iT}} \underset{k,k}{(X^T X)} \underset{k,m_{1i}+k_{1i}}{\hat{\Pi}_*^i} &= \begin{bmatrix} \underset{m_{1i},m_{1i}}{\hat{\Pi}_i^T X^T X \hat{\Pi}_i} & \underset{m_{1i},k_{1i}}{\hat{\Pi}_i^T X^T X_{1i}} \\ \underset{k_{1i},m_{1i}}{X_{1i}^T X \hat{\Pi}_i} & \underset{k_{1i},k_{1i}}{X_{1i}^T X_{1i}} \end{bmatrix} = \\ &= \begin{bmatrix} \underset{m_{1i},m_{1i}}{Y_{1i}^T X (X^T X)^{-1} X^T Y_{1i}} & \underset{m_{1i},k_{1i}}{Y_{1i}^T X_{1i}} \\ \underset{k_{1i},m_{1i}}{X_{1i}^T Y_{1i}} & \underset{k_{1i},k_{1i}}{X_{1i}^T X_{1i}} \end{bmatrix} = \underset{m_{1i}+k_{1i},m_{1i}+k_{1i}}{A_{ii}} \,, \end{aligned} \tag{17}$$

then, disreguarding the constant $1/k$, the quadratic form to be minimized becomes

$$\delta_i^T A_{ii} \delta_i \,, \tag{18}$$

where the reasons for using the symbol A_{ii} will become clear when treating the full information version of LODE method.

LODE estimator has then to be proportional to the vector, say P, such that

$$P^T A_{ii} P = min \,, \tag{19}$$

where, to make the solution univocally determined, the condition

$$P^T P = 1 \,, \tag{20}$$

has to be added.

As it is well known, to find the minimum of (19) under condition (20) one has to minimize the function

$$G = P^T A_{ii} P - \lambda (P^T P - 1) \,,$$

with respect to p_i $(i = 1, \dots, m_{1i} + k_{1i})$ and to the Lagrange multiplier λ.

The system obtained equating to zero the partial derivatives with respect to P and λ will then be

$$\begin{aligned} \frac{\partial G}{\partial P} &= 2 A_{ii} P - 2\lambda P = 0 \\ \frac{\partial G}{\partial \lambda} &= P^T P - 1 = 0 \,, \end{aligned} \tag{21}$$

whose solutions will be obtained solving the system

$$(A_{ii} - \lambda I) P = 0 \,, \tag{22}$$

under condition given by the second of (21).

Let us remember that to obtain a solution for (22), λ has to be the solution of the determinantal equation

$$\left| A_{ii} - \lambda I \right| = 0 \,, \tag{23}$$

which, being a polynomial of degree $s \le m_{1i} + k_{1i} - 1$ in λ, gives raise to s roots such that

$$\lambda_1 \ge \lambda_2 \ge \dots \ge \lambda_s \ge 0 \,,$$

The vector P_s associated to the smallest root λ_s of equation (23) is then the solution of the problem.

As a consequence the equation

$$X\hat{\Pi}_*^i P_s = 0\,, \tag{24}$$

is the expression of the $(s-1)$ dimensional subspace spanned by the first $(s-1)$ principal axis, i. e. the one which minimizes the sum of squares of the orthogonal distances between the observed points and the subspace itself. In other words (24) will be the last principal component.

Introducing at this point the normalization rule for *i-th* structural equation, least orthogonal distance estimator of δ_i are defined as

$$\hat{\delta}_i = \begin{bmatrix} \hat{\Gamma}_{1i} \\ \hat{B}_{1i} \end{bmatrix} = -\frac{1}{p_{oi}} P_s\,, \tag{25}$$

where p_{oi} is the element of the characteristic vector associated with the right hand side endogenous wariable in the *i-th* structural equation.

The estimate of *i-th* structural equation variance of disturbances will be as a consequence

$$\hat{\sigma}_i^2 = \frac{1}{kp_{oi}^2}\lambda_s\,, \tag{26}$$

Notice that when the *i-th* equation is exactly identified equation (23) will have $(m_{1i} + k_{1i} - 1)$ roots the last one being $\lambda_{m1i+k1i-1} = 0$ so that equation (18) will have a unique solution that coincides with ILS estimator.

On the contrary when *i-th* equation is under identified $(k_{2i} < m_{1i} - 1)$ the carateristic root equal to zero will have multiplicity equal to $r = m_{1i} - 1 - k_{2i}$ and the system (22) will have infinite to the r solutions.

5. FULL INFORMATION LODE

Relation (13) between reduced and structural form parameters for the whole system of equation can be written as

$$\begin{bmatrix} \underset{k,m_{11}+k_{11}}{\hat{\Pi}_*^1} & \underset{k,m_{12}+k_{12}}{0} & \cdots & \underset{k,m_{1m}+k_{1m}}{0} \\ \underset{k,m_{11}+k_{11}}{0} & \underset{k,m_{12}+k_{12}}{\hat{\Pi}_*^2} & \cdots & \underset{k,m_{1m}+k_{1m}}{0} \\ \vdots & \vdots & \ddots & \vdots \\ \underset{k,m_{11}+k_{11}}{0} & \underset{k,m_{12}+k_{12}}{0} & \cdots & \underset{k,m_{1m}+k_{1m}}{\hat{\Pi}_*^2} \end{bmatrix} \begin{bmatrix} \underset{m_{11}+k_{11,1}}{\delta_1} \\ \underset{m_{12}+k_{12,1}}{\delta_2} \\ \vdots \\ \underset{m_{1m}+k_{1m,1}}{\delta_1} \end{bmatrix} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_m \end{bmatrix}, \tag{27}$$

or in a more compact form, using a self evident notation

$$\underset{mk,s}{\hat{\Pi}_*}\ \underset{s,1}{\delta} = \underset{mk,1}{\varepsilon}\ , \tag{28}$$

where it is

$$s = \sum_{i=1}^{m} (m_{1i} + k_{1i})\ ,$$

and

$$\underset{km,1}{\varepsilon} = \left[I_m \otimes \underset{k,n}{(X^T X)^{-1} X^T} \right] \underset{nm,1}{vec(V\Gamma)} = \left[I_m \otimes \underset{k,n}{(X^T X)^{-1} X^T} \right] \underset{nm,1}{vec(U)}\ ,$$

Because of (14) applied to the vector ε defined in (28) the variance-covariance matrix of the error component can be written in the following way

$$E(\varepsilon\varepsilon^T) = \underset{mk,mk}{\Sigma} = \underset{m,m}{\Omega} \otimes \left(\underset{k,k}{X^T X} \right)^{-1}, \tag{29}$$

and taking into account that it is

$$E(\varepsilon^T \varepsilon) = tr\{\Omega\} tr\{(X^T X)^{-1}\} = \sum_{i=1}^{m} \sigma_i^2 \sum_{j=1}^{k} d^{jj}\ , \tag{30}$$

where d^{jj} are the diagonal elements of $(X^T X)^{-1}$, to obtain full information LODE it is necessary to minimize the quadratic form

$$\underset{1,s}{\delta^T}\ \underset{s,mk}{\hat{\Pi}_*^T} \underset{mk,mk}{(\Omega \otimes (X^T X)^{-1})^{-1}} \underset{mk,s}{\hat{\Pi}_*}\ \underset{s,1}{\delta} = \delta^T \hat{\Pi}_*^T (\Omega^{-1} \otimes (X^T X)) \hat{\Pi}_* \delta\ , \tag{31}$$

i.e. to consider the characteristic vector associated with the smallest characteristic root of the matrix

$$\underset{s,s}{A} = \hat{\Pi}_*^T (\Omega^{-1} \otimes (X^T X)) \hat{\Pi}_*\ , \tag{32}$$

where its explicit form is the following

$$A = \begin{bmatrix} \hat{\Pi}_*^{1T} \sigma^{11} (X^T X) \hat{\Pi}_*^1 & \cdots & \hat{\Pi}_*^{1T} \sigma^{1i} (X^T X) \hat{\Pi}_*^i & \cdots & \hat{\Pi}_*^{1T} \sigma^{1m} (X^T X) \hat{\Pi}_*^m \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ \hat{\Pi}_*^{iT} \sigma^{i1} (X^T X) \hat{\Pi}_*^1 & \cdots & \hat{\Pi}_*^{iT} \sigma^{ii} (X^T X) \hat{\Pi}_*^i & \cdots & \hat{\Pi}_*^{iT} \sigma^{im} (X^T X) \hat{\Pi}_*^m \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \hat{\Pi}_*^{iT} \sigma^{m1} (X^T X) \hat{\Pi}_*^1 & \cdots & \hat{\Pi}_*^{mT} \sigma^{mi} (X^T X) \hat{\Pi}_*^i & \cdots & \hat{\Pi}_*^{mT} \sigma^{mm} (X^T X) \hat{\Pi}_*^m \end{bmatrix} \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}\ ,$$

with σ^{ij} being the element of the matrix Ω^{-1}.

While the block diagonal elements of A are of the form (17) – now it is clear the reason for using the proposed notation – the extra diagonal block elements are

$$A_{ij} = \sigma^{ij} \begin{bmatrix} \underset{m_{1i},k}{\hat{\Pi}_i^T} \underset{k,k}{X^T X} \underset{k,m_{1j}}{\hat{\Pi}_j} & \underset{m_{1i},k}{\hat{\Pi}_i^T} \underset{k,n}{X^T} \underset{n,k_{1j}}{X_{1j}} \\ \underset{m_{1i},k}{X_{1i}^T} \underset{n,k}{X} \underset{k,m_{1j}}{\hat{\Pi}_j} & \underset{m_{1i},k}{X_{1i}^T} \underset{n,k_{1j}}{X_{1j}} \end{bmatrix}, \tag{33}$$

that come out to be

$$A_{ij} = \sigma^{ij} \begin{bmatrix} \underset{m_{1i},m_{1j}}{Y_{1i}^T X (X^T X)^{-1} X^T Y_{1j}} & \underset{m_{1i},k_{1j}}{Y_{1i}^T X_{1j}} \\ \underset{k_{1i},m_{1j}}{X_{1i}^T Y_{1j}} & \underset{k_{1i},k_{1j}}{X_{1i}^T X_{1j}} \end{bmatrix}, \tag{34}$$

The characteristic vector associated with the smallest characteristic root of matrix $\underset{S,S}{A}$ minimizes the quadratic form (31).

Let a be the smallest characteristic root of $\underset{S,S}{A}$ and P_a be the associated characteristic vector. The characteristic vector P_a multiplied by m suitable constants gives FI LODE.

Defining C as the block diagonal matrix

$$C = \begin{bmatrix} c_1 I_{m_{11}+k_{11}} & & \\ & \ddots & \\ & & c_m I_{m_{1m}+k_{1m}} \end{bmatrix}, \tag{35}$$

in which c_i are defined as follows

$$c_i = -\frac{1}{p_{0i}}, \tag{36}$$

with p_{0i} being the characteristic vector's element corresponding to the endogenous variable y_{oi} chosen to be at left hand side in *i-th* structural equation.

The FI estimator is then

$$\hat{\delta} = C P_a, \tag{37}$$

6. ESTIMATION OF THE VARIANCE-COVARIANCE MATRIX

Equation (31) which defines explicitly the quadratic form to be minimized is a function of disturbances variance-covariance matrix Ω which is unknown. It is then necessary to estimate it.

As usual, it is possible to go through a two steps procedure: in the first step estimates of the SF parameters are obtained using LI LODE which are then used to calculate the matrix $\hat{U}$ of SF disturbances

$$\hat{U} = -\hat{V}\hat{\Gamma},$$

$\hat{V}$ been the matrix of RF equations' OLS residuals.

The matrix $\hat{\Omega}$ is then computed in the following way

$$\hat{\Omega} = G^{-1/2}\hat{U}^T\hat{U}G^{-1/2},$$

Where

$$G^{-1/2} = \begin{bmatrix} 1/\sqrt{g_1} & 0 & \cdots & \cdots & 0 \\ 0 & \ddots & \cdots & \cdots & \vdots \\ 0 & \cdots & 1/\sqrt{g_1} & \cdots & 0 \\ \vdots & \cdots & \cdots & \ddots & 0 \\ 0 & \cdots & \cdots & 0 & 1/\sqrt{g_m} \end{bmatrix},$$

with

$$g_i = n - m_{1i} - k_{1i},$$

It has to be notice that consistency of limited information SF parameters estimators implies the consistency of the variance covariance matrix estimators.

The second stage structural parameters estimates are then obtained introducing $\hat{\Omega}$ in equation (31). Full Information LODE is then proportional to the characteristic vector associated to the smallest characteristic root of

$$\hat{A} = \hat{\Pi}_*^T(\hat{\Omega}^{-1} \otimes (X^T X))\hat{\Pi}_*.$$

from which $\hat{\delta}$ follows according to (37).

7. CONSISTENCY OF FULL INFORMATION LODE

Generalizing a result already obtained for LI LODE (Perna, 1988) it is possible to prove consistency of FI LODE.

Let us assume that conditions (2) and (3) are true and – in addition – that the exogenous variables matrix $\underset{n,k}{X}$ is of full rank

$$r(\underset{n,k}{X}) = k ,$$

and that its elements are non random. Under these conditions the following Lemma is true.

Lemma: The characteristic vector associated to the smallest characteristic root of the matrix $\Pi_*^T(\Omega^{-1} \otimes (X^T X))\Pi_*$ is proportional to the parameters vector δ according to m constants of proportionality.

Proof: Considering equation (8) for the whole system, the following expression can be written

$$\Pi_*^T(\Omega^{-1} \otimes (X^T X))\Pi_*\delta = 0 ,$$

or

$$H\delta = 0 ,$$

where H is the matrix $\Pi_*^T(\Omega^{-1} \otimes (X^T X))\Pi_*$.

Let α be the smallest characteristic root of matrix H and Ψ_α the characteristic vector associated to it, then it is possible to write

$$\Psi_\alpha^T \Pi_*^T(\Omega^{-1} \otimes (X^T X))\Pi_*\Psi_\alpha = 0 . \tag{38}$$

Since the matrix Ω is assumed to be positive definite and matrix X of full rank, condition (38) is true if and only if

$$\Pi_*\Psi_\alpha = 0 , \tag{39}$$

It follows that Ψ_α has to be proportional to δ - since δ is the vector of parameters for which (39) is true.

Let $\Psi_{\alpha i}$ be the sub-vector of Ψ_α corresponding to *i-th* equation and ξ_i be the corresponding constant of proportionality, the δ vector can be written as follows

$$\underset{S,1}{\delta} = \begin{bmatrix} \xi_1 \underset{m_{11}+k_{11}}{I} & & & & \\ & \ddots & & & \\ & & \xi_i \underset{m_{i1}+k_{i1}}{I} & & \\ & & & \ddots & \\ & & & & \xi_1 \underset{m_{m1}+k_{m1}}{I} \end{bmatrix} \begin{bmatrix} \underset{m_{11}+k_{11},1}{\Psi_{\alpha 1}} \\ \vdots \\ \underset{m_{i1}+k_{i1},1}{\Psi_{\alpha i}} \\ \vdots \\ \underset{m_{m1}+k_{m1},1}{\Psi_{\alpha m}} \end{bmatrix} ,$$

where

$$\xi_i = \frac{1}{\psi_{0i}},$$

and ψ_{0i} is the element corresponding to the endogenous variable at left hand side of the *i-th* structural equation, with respect to which the normalization rule has been made.

Theorem: Full Information LODE consistently estimates structural form parameters.

Proof: Taking into account OLS estimator consistency of RF parameters, it is

$$p\lim_{n\to\infty} \hat{\Pi}_*^T (\Omega^{-1} \otimes X^T X)\hat{\Pi}_* = \Pi_*^T (\Omega^{-1} \otimes X^T X)\Pi_* . \tag{40}$$

Since the characteristic roots of a matrix are differentiable functions of its elements (Kato, 1982), if a is the smallest characteristic root of $\hat{\Pi}_*^T (\Omega^{-1} \otimes X^T X)\hat{\Pi}_*$, it follows that

$$p\lim_{n\to\infty} a = \alpha ,$$

and consequently

$$p\lim_{n\to\infty} P_a = \Psi_\alpha . \tag{41}$$

According to the preceeding Lemma the vector of SF parameters δ is proportional to the characteristic vector Ψ_α. It is then

$$p\lim_{n\to\infty} \hat{\delta} = \delta . \tag{42}$$

8. CONCLUSIONS

The extension of Least Orthogonal Distance Estimators to a Full Information context, is the occasion for the authors to give a complete illustration of a new version of LODE method, the principles upon which the method is based, the formal derivation of its properties and to illustrate some very impressive resultus obtained in the simulation experiment.

With regard to this last point one has always to take in mind that the number of equations in the system, the different degree of parameters' over-identification and the characteristics of the simulation experiment influence the results, that hence have to be taken with caution. Yet the simulation experiment that has been recently produced (Naccarato and Zurlo, 2007) has highlighted some very good features of the method – both in terms of bias and MSE – that we think it is worthwhile to mention.

The simulation experiment has been made using the same three equation model proposed by Cragg in 1967. The simulation, starting from a so called *basic experiment* (characterized by a variance of the error component between 20-25% of the variance of the endogenous variable to which it refers), considers increasing levels of variance of the error components in each equation and of the correlation between them. Increasing sample sizes are also considered.

With regard to the *basic experiment* it has to be stressed that, apart from one exception, both Limited Information and Full Information LODE feature a lower bias than other estimators reaching up to more than 90% bias reduction with respect to 2SLS. The exception is given by the third equation's coefficients (for samples of size less than 50) for which the greater proportion of cases with lower bias is taken by 2SLS.

At increasing values of the variance of error components and of the correlation, LODE method – both LI and FI – always shows lower bias than other methods in at least the 60% of the parameters reaching sometimes 80-90% of the cases.

With respect to MSE in the *basic experiment* LI LODE is better than 2SLS for sample size greater than 30. On the contrary FI LODE presents a positive or negative variation with respect to 3SLS of about 30% for sample size of 20. Things get better by increasing sample size when there are gains up to the order of 70% in the case of size equal to 100.

At increasing values of variances and covariances, LI LODE is more efficient than 2SLS only for samples size greater or equal to 50 as in the *basic experiment* while FI LODE is frequently more efficient than 3SLS obtaining gains up to 30%. In particular, it performs better (in more than 60% of the cases) when the variance of the error component is lower than fifty percent of the endogenous variable's one.

Finally it has to be noticed that FI LODE could have strong computational advantages with respect to FIML.

Departments of Economics LUCIANO PIERACCINI
Roma Tre University ALESSIA NACCARATO

REFERENCES

G. CAU (1990), *Stimatori alternativi nei modelli ad equazioni simultanee: un esperimento di simulazione*, Unpublished thesis, Università "La Sapienza", Rome.

A. S. GOLDBERGER (1964), *Econometric Theory*, John Wiley & Sons, Inc.

G. G. JUDGE, W. E. GRIFFITH, R. C. HILL, H. LUTKEPOHL, T. C. LEE (1985), *The Theory and Practice of Econometrics*, J. Wiley, New York, II Edition.

T. KATO (1982), *A short Introduction to Perturbation Theory for Linear Operation*, Springer, New York.

A. NACCARATO (2007), *Full Information Least Orthogonal Distance Estimator of Structural Parameters in Simoultaneous Equation Models*, Quaderni di Statistica, 9, 87-105.

A. NACCARATO, D. ZURLO (2007), *Least Orthogonal Estimation of Simulataneous Equations: a Simulation Experiment*, Collana del Dipartimento di Economia, Università degli Studi Roma Tre, Working Paper n° 90.

K. PEARSON (1901), *On Lines and Planes of Closest Fit to System of Points in Space*, University College, London.

L. PIERACCINI (1969), *Su di una interpretazione alternativa del metodo dei minimi quadrati a due stadi*, Statistica, IV, 786-802.

L. PIERACCINI (1978), *Su alcuni metodi di stima dei parametri strutturali di un sistema di equazioni simultanee*, Quaderni di Statistica e Econometria, I, 1-30.

L. PIERACCINI (1983), *The estimation of Structural Parameters in Simultaneous Equations Models*, Quaderni di Statistica ed Econometria, V, 3-20.

L. PIERACCINI (1988), *Il metodo LODE per la stima dei parametri strutturali di un sistema di equazioni simultanee*, Quaderni di Statistica e Econometria, X, 5-14.

L. PIERACCINI (1992), *Metodi di stima dei parametri strutturali nei modelli lineari ad equazioni simultanee: una rassegna ed una proposta*, Atti della XXXI Riunione Scientifica della Società Italiana di Statistica, 435-446.

C. PERNA (1988), *La consistenza dello stimatore LODE nei sistemi ad equazioni simultanee*, Quaderni di Statistica e Econometria, X, 15-24.

C. PERNA (1989), *Un confronto tra due metodi di stima: il 2SLS ed il LODE*, Quaderni di Statistica e Econometria, XI, 23-43.

G. SBRANA (2001), *Una generalizzazione del metodo LODE per la stima dei parametri strutturali di un sistema di equazioni simultanee*, Quaderni di Statistica, 3, 107-125.

D. ZURLO (2006), Un esperimento Montecarlo per la stima dei parametri strutturali di modelli ad equazioni simultanee, Unpublished thesis, Università Roma Tre.

SUMMARY

Least orthogonal distance estimator of structural parameters in simultaneous equation models

The aim of this paper is to present a consistent estimator of parameters in simultaneous equation model, based on characteristic roots and vectors of a matrix derived from the so called over-identifying restrictions. The Least Orthogonal Distance Estimator presented here is a more recent development of its original limited information version. The occasion, for reviewing it, has been given by its extension to a full information context which is here completely formalized and by the very encouraging results of recent simulation experiments.