STATISTICA, anno LXXIII, n.2, 2013

LEAST ORTHOGONAL DISTANCE ESTIMATOR AND TOTAL LEAST SQUARE

Alessia Naccarato
Dipartimento di Economia, Università di Roma Tre, Roma, Italia
Davide Zurlo
ISTAT – Ististuto Nazionale di Statistica, Roma, Italia
Luciano Pieraccini
Dipartimento di Economia, Università di Roma Tre, Roma, Italia

1. INTRODUCTION

As it is well known the relations between structural form (SF) and Reduced form (RF) parameters of simultaneous equation models are established in the so called identifying system of equations. These are nothing else but linear relations between variable affected by error once RF estimates are considered.
It may be worth to remember that T. W. Anderson (1976) in a very well known paper explicitly recognized it saying: "It turns out, however, that a problem investigated in great detail by econometricians can be transformed so that it is mathematically identical to the problem of fitting a straight line when both variables are subjected to error. In estimating a coefficient of an (endogenous) variable in one equation in a system of simultaneous equations, the first stage is to find the sample regression coefficients of the dependent (endogenous) variables on the independent (exogenous) variables. The sample regression coefficients are mathematically equivalent to the observations in the model described above, and the population regression coefficient satisfy a linear relationship". Along this line of thought Limited Information (LI) LODE has been produced initially (Pieraccini 1983, 1988) while more recently a first version of FI LODE was developed (Pieraccini and Naccarato, 2008).
A new version of FI LODE is presented here based on a new structure of the variance-covariance matrix that is employed in the estimation process. While in the previous version FI LODE was based on the variance-covariance matrix of error components related to the whole system of identifying equation, in the present one the variance-covariance matrix only refers to the error component of the so called over identifying equations.
Furthermore a new computational procedure is proposed using Singular Value Decomposition (SVD) instead of Spectral Decomposition (SD) used in the past. Even if results on SVD and SD for symmetric matrices are theoretically the same, the

computational algorithm based on SVD is numerically more robust with respect to the one based on SD; where robustness has to be understood as the greater probability to converge presented by the algorithm (Markovsky and Van Huffel 2007; Jennings 1980). In the light of these results a new computational procedure has been developed following the work of Gleser (1981) applying the Total Least Square procedure (Golub and Van Loan 1980; Van Huffel 1988, 1989, 2002, 2007).
The reason for the new version of FI LODE and for the use of a more robust estimation procedure has to be found in the results of a previous simulation experiment (Naccarato and Zurlo, 2008). The results showed that while FI LODE works usually better than other classic full information estimators in terms of bias, in terms of mean square error the estimates were affected by the presence of few very far outliers that weighted heavily on Mean Square Error (MSE).
A new Monte Carlo experiment (with the same structure of the one presented in the previous contribution) was then produced to evaluate the performance of the method with respect both to other classic full information methods and to the preceding FI LODE version.
In order to establish notation simultaneous equation models are briefly presented in paragraph 1 together with the original LI LODE, revisited in the light of SVD. In paragraph 2 the new version of FI LODE is shown and in paragraph 3 Total Least Square procedure is applied to. In paragraph 4 the complete design of the simulation experiment is presented. In paragraph 5 FILODE is compared with some usual methods of estimation like 3SLS (Theil and Zelner 1962) and FIML (Koopmans et al. 1950); the results are shown focusing on bias and MSE of estimators. In paragraph 6 few words of conclusion including a comparison between the previous version of FI LODE and the present one end the paper.

2. SIMULTANEOUS EQUATIONS SYSTEM

As it is well known the structural form (SF) of a simultaneous equation model can be defined as follows:

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

where Y is the $n \times m$ matrix of endogenous variables and Γ is the corresponding $m \times m$ matrix of structural parameters, X is the $n \times k$ matrix of exogenous variables and B is the $k \times m$ matrix of their structural parameters. Finally U is the $n \times m$ matrix of disturbances with

$$\begin{aligned} &\mathrm{E}(\mathrm{vec}\,U) = \mathbf{0} \\ &\mathrm{E}[\mathrm{vec}\,U(\mathrm{vec}\,U)^T] = \Omega \otimes I \end{aligned} \tag{2}$$

where

$$\Omega = \begin{bmatrix} \sigma_1^2 & \cdots & \sigma_{1m} \\ \vdots & \ddots & \vdots \\ \sigma_{m1} & \cdots & \sigma_m^2 \end{bmatrix}$$

is the constant variance-covariance matrix of the disturbances.
The reduced form (RF) of system (1), under non singularity condition for Γ, is

$$\underset{n,m}{Y} = \underset{n,k}{X}\,\underset{k,m}{\Pi} + \underset{n,m}{V} \tag{3}$$

where

$$\begin{aligned} \underset{k,m}{\Pi} &= -\underset{k,m}{\mathbf{B}}\,\underset{m,m}{\Gamma^{-1}} \\ \underset{n,m}{V} &= -\underset{n,m}{U}\,\underset{m,m}{\Gamma^{-1}} \end{aligned} \tag{4}$$

As it is well known the first of (4) gives the link between RF and SF parameters which post-multiplied by Γ gives

$$\underset{k,m}{\Pi}\,\underset{m,m}{\Gamma} = -\underset{k,m}{\mathbf{B}} \tag{5}$$

Introducing then exclusion conditions the following *identification system* for *i-th* equation is obtained:

$$\begin{cases} \underset{k_1,m_1}{\hat{\Pi}_{11}^i}\,\underset{m_1,1}{\Gamma_{1i}} + \underset{k_1,m_1}{B_{1i}} = \underset{}{\varepsilon_{1i}} \\ \underset{k_2,m_1}{\hat{\Pi}_{12}^i}\,\underset{m_1,1}{\Gamma_{1i}} = \underset{}{\varepsilon_{2i}} \end{cases} \tag{6}$$

where:

$$\hat{\Pi}_{11}^i = \begin{bmatrix} \underset{k_1,1}{\hat{\pi}_{01}^i} & \vdots & \underset{k_1,m_1-1}{\hat{\pi}_{11}^{ie}} \end{bmatrix}$$

refers to OLS estimates of reduced form parameters of endogenous and exogenous variables included in the *i-th* equation;

$$\hat{\Pi}_{12}^i = \begin{bmatrix} \underset{k_2,1}{\hat{\pi}_{02}^i} & \vdots & \underset{k_2,m_2-1}{\hat{\pi}_{12}^{ie}} \end{bmatrix}$$

refers to OLS estimates of reduced form parameters of endogenous included and exogenous excluded ones; Γ_{1i} and $\mathbf{B}_{1i}$ are the structural form endogenous and exogenous var-

iables' coefficients included in the *i-th* equation; finally ε_{1i} and ε_{2i} are the error components of the system.

In the original paper in which the limited information version of LODE was proposed (Pieraccini 1988), the estimation was based on the second subsystem of (6) and not on the whole system as it was in the more recent paper (Pieraccini and Naccarato 2008). We have decided to go back to the original version of LODE and to derive the full information version starting from that point. As it will be seen in the conclusion (§ 6) the goodness of this choice has been somehow confirmed by the simulation experiment.

To simplify the exposition of the new version of FI LODE and the computational procedure let us now introduced LI LODE (Pieraccini 1988) in the light of SVD.
It can be shown (Pieraccini 1969) that in the second of (6) it is

$$\varepsilon_{2i} = R_{2i} X_i^T U_{1i} \tag{7}$$

where R_{2i} comes from

$$\left(X_i^T X_i \right)^{-1} = \begin{bmatrix} R_{1i} \\ k_{1i},k \\ R_{2i} \\ k_{2i},k \end{bmatrix} = \begin{bmatrix} R_{11ii} & R_{12ii} \\ k_{1i},k_{1i} & k_{1i},k_{2i} \\ R_{21ii} & R_{22ii} \\ k_{2i},k_{1i} & k_{2i},k_{2i} \end{bmatrix} \tag{8}$$

so that it is $\mathrm{E}(\varepsilon_{2i}) = 0$, $\mathrm{E}(\varepsilon_{2i}\varepsilon_{2i}^T) = \sigma_i^2 R_{22ii}$.

According to the Spectral Decomposition theorem LI LODE estimators of structural parameters Γ_i's are based on the eigenvector corresponding to the minimum eigenvalue of variance-covariance matrix $\hat{\Pi}_{12}^i R_{22ii}^{-1} \hat{\Pi}_{12}^{i\ T}$.

Since $R_{22ii} = C\Lambda C^T$, with C and Λ the matrices of eigenvectors and eigenvalues of R_{22ii}, it is possible to define $Q_{22i} = C\Lambda^{-\frac{1}{2}} C^T$.

Premultiplying the elements of the second equation of (6) for Q_{22i} it will become

$$Q_{22i}\hat{\Pi}_{12}^i \Gamma_{1i} = Q_{22i}\varepsilon_{2i} \tag{9}$$

LI LODE method is based on the minimization , with respect to Γ_{1i}, of the variance covariance matrix $Q_{22i}\varepsilon_{2i}$. Using singular value decomposition (9) becomes

$$Q_{22i}\hat{\Pi}_{12}^i = S\Lambda D^T \tag{10}$$

with $S = \lfloor S_1, \ \ldots, \ S_{m_{1i}} \rfloor$, $\Lambda = \mathrm{diag}\left(\lambda_1, \ \ldots, \ \lambda_{m_{1i}}\right)$ and $D = \lfloor D_1, \ \ldots, \ D_{m_{1i}} \rfloor$ if $\lambda_{m_{1i}} \neq 0$.

An approximation of $\mathcal{Q}_{22i}\hat{\Pi}_{12}^{i}$ denoted $\left(\mathcal{Q}_{22i}\hat{\Pi}_{12}^{i}\right)'$ is needed together with a vector P such that

$$\left(\mathcal{Q}_{22i}\hat{\Pi}_{12}^{i}\right)' P_{m_{1i}} = 0 \tag{11}$$

According to Eckart and Young (1936), the best rank $m_{1i} - 1$ approximation $\left(\mathcal{Q}_{22i}\hat{\Pi}_{12}^{i}\right)'$ of $\mathcal{Q}_{22i}\hat{\Pi}_{12}^{i}$, is given by $\left(\mathcal{Q}_{22i}\hat{\Pi}_{12}^{i}\right)' = S\Lambda D^T$ where it is $\Lambda' = \mathrm{diag}\left(\lambda_1, \ \square, \ \lambda_{m_{1i}-1}, 0\right)$ and the minimal correction is

$$\min_{rank(L)=m_{1i}-1} \left\| \mathcal{Q}_{22i}\hat{\Pi}_{12}^{i} - \left(\mathcal{Q}_{22i}\hat{\Pi}_{12}^{i}\right)' \right\|_F^2 = \lambda_{m_{1i}} .$$

The solution of equation (11) is given by the m_{1i}*-th* vector of the matrix D, called $d_{m_{1i}}$, that belongs to $N\left(\mathcal{Q}_{22i}\hat{\Pi}_{12}^{i}\right)'$ (the null space of the approximation matrix).

The estimate of the parameters entering the *i-th* equation is the normalized right singular vector of $\mathcal{Q}_{22i}\hat{\Pi}_{12}^{i}$, namely the eigenvector of the matrix $\hat{\Pi}_{12}^{i\,'}\mathcal{Q}_{22i}'\mathcal{Q}_{22i}\hat{\Pi}_{12}^{i}$, associated to the smallest eigenvalue of this latter matrix.

The estimates of structural parameters Γ_{1i} for the *i-th* structural equation are defined as $\hat{\Gamma}_{1i} = -\frac{1}{\nu_{0i}} d_{m_{1i}}$, where ν_{0i} is the element of the characteristic vector associated with right hand side endogenous variable.

It has to be noticed that $\left(\mathcal{Q}_{22i}\hat{\Pi}_{12}^{i}\right)' d_{m_{1i}} = 0$ and $\left(\mathcal{Q}_{22i}\hat{\Pi}_{12}^{i}\right)'$ represents the $(m_{1i} - 1)$ dimensional subspace spanned by the first $(m_{1i} - 1)$ principal axis that minimize the sum of squares orthogonal distance between the observed points and the subspace itself.

3. THE NEW VERSION OF FI LODE

The second equation of system (6) for the whole system of equation can be written as

$$\underset{r,z}{\hat{\Pi}_{12}} \underset{z,m}{\Gamma_1} = \underset{r,m}{\Xi} \tag{12}$$

where

$$\underset{r,m}{\Xi} = \begin{bmatrix} \varepsilon_{21} & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \varepsilon_{2m} \end{bmatrix}$$

and, once the dependent endogenous variable is chosen in each equation, the Γ_1 matrix can be written as

$$\underset{\tilde{z},m}{\Gamma_1} = \begin{bmatrix} & -I_m & \\ \underset{m_{11}-1,1}{\Gamma^e_{11}} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \underset{m_{1m}-1,1}{\Gamma^e_{1m}} \end{bmatrix}$$

with $\tilde{z} = \sum_{i=1}^{m} m_{1i}$ where m is the number of structural form's equations. The identity matrix $-I_m$ contains the m coefficients of the dependent endogenous variables and $\underset{m_{1i}-1,1}{\Gamma^e_{1i}}$ is the matrix of the coefficients of the endogenous explanatory variables of *i-th* equation after the normalization rule.
Let:

$$\underset{r,\tilde{z}}{\hat{\Pi}_{12}} = \begin{bmatrix} \hat{\pi}^1_{02} & 0 & 0 & \hat{\Pi}^{1e}_{12} & 0 & 0 \\ 0 & \ddots & 0 & 0 & \ddots & 0 \\ 0 & 0 & \hat{\pi}^m_{02} & 0 & 0 & \hat{\Pi}^{me}_{12} \end{bmatrix}$$

$r = \sum_{i=1}^{m} k_{2i}$ and

$$E\left[\operatorname{vec}\Xi\left(\operatorname{vec}\Xi\right)^T\right] = \begin{bmatrix} \sigma^2_{11}R_{2211} & \cdots & \sigma_{1i}R_{221i} & \cdots & \sigma_{1m}R_{221m} \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ \sigma_{i1}R_{22i1} & \cdots & \sigma^2_{ii}R_{22ii} & \cdots & \sigma_{im}R_{22im} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sigma_{m1}R_{22m1} & \cdots & \sigma_{mi}R_{22mi} & \cdots & \sigma^2_{mm}R_{22mm} \end{bmatrix} \tag{13}$$

In equation (13) while the block diagonal elements are defined in (8), the extra-diagonal elements are given by

$$\left(\underset{k,k}{X_i^T X_j} \right)^{-1} = \begin{bmatrix} \underset{k_{1i},k_{1j}}{R_{11ij}} & \underset{k_{1i},k_{2j}}{R_{12ij}} \\ \underset{k_{2i},k_{1j}}{R_{21ij}} & \underset{k_{2i},k_{2j}}{R_{22ij}} \end{bmatrix}$$

X_i and X_j being the matrix of exogenous variables ordered according to exogenous included and excluded variables in *i-th* and *j-th* structural equation.

It has to be stressed the difference between (13) and the variance-covariance matrix considered in the previous version of LI LODE (Pieraccini and Naccarato 2008).

4. TOTAL LEAST SQUARE PROCEDURE

Total Least Square procedure has been applied to equation (12) to estimate endogenous variables' coefficients. Let it be

$$Q_2 Q_2^T = \begin{bmatrix} \sigma_{11}^2 R_{2211} & \cdots & \sigma_{1i} R_{221i} & \cdots & \sigma_{1m} R_{221m} \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ \sigma_{i1} R_{22i1} & \cdots & \sigma_{ii}^2 R_{22ii} & \cdots & \sigma_{im} R_{22im} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sigma_{m1} R_{22m1} & \cdots & \sigma_{mi} R_{22mi} & \cdots & \sigma_{mm}^2 R_{22mm} \end{bmatrix} \tag{14}$$

where

$$Q_2 = V \Lambda^{-\frac{1}{2}} V^T \tag{15}$$

and V and Λ are – respectively – the matrix of eigenvectors (the first) and the diagonal matrix of eigenvalues of (14) (the second).

The endogenous variables parameters estimates are given by the matrix $\underset{z,m}{D_2}$ of the last m singular right vectors that correspond to the m smallest singular values of $Q_2 \hat{\Pi}_{12}$. According to Eckart and Young (1936) the best rank $z-m$ matrix approximation of $Q_2 \hat{\Pi}_{12}$ is defined as $\left(Q_2 \hat{\Pi}_{12} \right)' = S \Lambda' D^T$ where $\Lambda' = diag\left(\lambda_1, \ldots, \lambda_{z-m}, 0, \ldots, 0 \right)$ is the singular values diagonal matrix of $Q_2 \hat{\Pi}_{12}$ with the last m elements equal to zero. S and D are the left and right matrices of $Q_2 \hat{\Pi}_{12}$, namely the respective eigenvectors of the matrix $Q_{22,i} \hat{\Pi}_{12}^i \hat{\Pi}_{12}^{i\,'} Q_{22,i}{}'$ and of the matrix $\hat{\Pi}_{12}^{i\,'} Q_{22,i}{}' Q_{22,i} \hat{\Pi}_{12}^i$.

The minimal correction is then

$$\min_{rank(L)=\tilde{z}-m} \left\| Q_2\hat{\Pi}_{12} - L \right\|_F^2 = \left\| Q_2\hat{\Pi}_{12} - \left(Q_2\hat{\Pi}_{12}\right)' \right\|_F^2 = \sqrt{\sum_{i=\tilde{z}-m+1}^{\tilde{z}} \lambda_i^2}$$

where λ_i^2 are the last m eigenvalues of $\hat{\Pi}_{12}^T Q_2 Q_2 \hat{\Pi}_{12}$.

Given the matrix D_2 of the last m right singular vectors $\underset{\tilde{z},m}{D_2} = \begin{bmatrix} \underset{m,m}{D_{12}} \\ \underset{\tilde{z}-m,m}{D_{22}} \end{bmatrix}$, the equation $\left(Q_2\hat{\Pi}_{12}\right)' D_2 = \underset{r,m}{0}$ is true.

Total Least Square procedure (Golub and Van Loan 1980; Van Huffel 2002) is then applied

$$-D_2\left(D_{12}\right)^{-1} = \begin{bmatrix} & -I_m & \\ \hat{\Gamma}_{11}^e & p_{1i} & p_{1m} \\ p_{i1} & \ddots & p_{im} \\ p_{m1} & p_{mi} & \hat{\Gamma}_{1m}^e \end{bmatrix}$$

and the estimates of endogenous parameters for *i-th* equation are the sub-vector $\hat{\Gamma}_{1i}^e$ of $-D_2\left(D_{12}\right)^{-1}$.

The vector of the estimates of endogenous variables' parameters $\hat{\Gamma}_1$ is

$$\underset{\tilde{z},1}{\hat{\Gamma}_1} = \begin{bmatrix} \underset{m_{11},1}{\hat{\Gamma}_{11}} \\ \underset{m_{12},1}{\hat{\Gamma}_{12}} \\ \vdots \\ \underset{m_{1m},1}{\hat{\Gamma}_{1m}} \end{bmatrix} \tag{16}$$

The exogenous coefficients matrix is then obtained as

$$\hat{\mathbf{B}}_1 = \left(X_1^T X_1\right)^{-1} X_1^T \left(Y_0 - Y_1\hat{\Gamma}_1\right) \tag{17}$$

where X_1 is the block diagonal matrix of exogenous variables included in each equation, Y_0 is the left side endogenous variables vector' for the m equations and Y_1 is the right side endogenous variables block diagonal matrix.

Equation (14) defines the matrix Q_2 as a function of the elements σ_{ii}^2 and σ_{ij} i.e. the variance of *i-th* equation's error component and the covariance between *i-th* and *j-th*

equations' ones, which are both of them unknown. It is then necessary to estimate them.

The disturbances variance-covariance matrix $\hat{\Omega}$ is obtained as in the previous version of FI LODE (Pieraccini and Naccarato 2008) through a two stage procedure.

Let $\hat{\Gamma}_{1i}$ be the first stage LI LODE estimates of the SF parameters and $\hat{V}$ the matrix of RF equations' residuals. Then the matrix of SF disturbances $\hat{U} = -\hat{V}\hat{\Gamma}$ is obtained in order to get

$$\hat{\Omega} = \begin{bmatrix} \hat{\sigma}_{11}^2 & \cdots & \hat{\sigma}_{1i} & \cdots & \hat{\sigma}_{1m} \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ \hat{\sigma}_{i1} & \cdots & \hat{\sigma}_{ii}^2 & \cdots & \hat{\sigma}_{im} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \hat{\sigma}_{m1} & \cdots & \hat{\sigma}_{mi} & \cdots & \hat{\sigma}_{mm}^2 \end{bmatrix} = G^{-\frac{1}{2}} \hat{U}^T \hat{U} G^{-\frac{1}{2}} \tag{18}$$

where

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

with $g_i = n - m_{1i} - k_{1i}$. The second stage structural parameters estimates are then obtained introducing $\hat{\sigma}$ in equation (15).

5. THE DESIGN OF THE EXPERIMENT

As in the previous simulation experiment (Naccarato and Zurlo, 2008), the new one has been performed using the three equation model proposed by Cragg in 1967:

$$\begin{cases} y_1 = -0.89\, y_2 - 0.16\, y_3 + 44 + 0.74 x_2 + 0.13 x_5 \\ y_2 = -0.74\, y_1 + 62 + 0.70 x_3 + 0.96 x_5 + 0.06 x_7 \\ y_3 = -0.29\, y_2 + 40 + 0.53 x_3 + 0.11 x_4 + 0.56 x_6 \end{cases}$$

Following the same scheme three sample sizes $n = 20, 30, 100$ have been considered. As in the previous experiment we have not taken into consideration very large samples'

sizes (let's say with $n = 1000$ or even $n = 10000$) on the ground that in the econometric practice the number of observations is generally not bigger than 30. Hence the following steps have been performed:

1. *Exogenous variables generation.* For each sample size exogenous variables have been generated from uniform distribution in the following intervals:

$$X_2 = [10 - 20], \quad X_3 = [15 - 27], \quad X_4 = [3 - 12], \quad X_5 = [3 - 7],$$

$$X_6 = [20 - 50], X_7 = [7 - 13].$$

Exogenous variables have been kept constant for each sample size.

2. *Error component variance covariance matrix generation.* The matrix Ω has been chosen in the following way:
   a) diagonal elements have been obtained as a proportion of the variance of
   $$Y\Gamma = Z \text{ i.e. } \omega_{ii} = \sigma_Z^2 S_i$$
   where S_i are proportionality coefficients randomly chosen from a uniform distribution in three intervals: $[0.2 - 0.25]$, $[0.4 - 0.5]$, $[0.75 - 0.8]$.
   b) extra diagonal elements have been obtained at first generating randomly $m(m-1)/2$ correlation coefficients ρ_{ij} in $[0.1 - 0.2]$, $[0.4 - 0.5]$, $[0.8 - 0.9]$ assigning them a random sign. Then covariance between error components in equation i and in equation j has been computed as $\omega_{ij} = \rho_{ij}\left(\omega_{ii}\,\omega_{jj}\right)^{1/2}$.
3. *Normal and Uniform error distribution.* For each sample of n observations, m series of random numbers have been generated independently from a standardized Normal distribution and from a Uniform distribution in the interval $\left(-\sqrt{3}, \sqrt{3}\right)$ to have zero mean and variance one. To evaluate the performance of the estimation methods considered in non standard situation, the experiment has been extended to the case in which the error component is Uniformly distributed in $(-10, 10)$.

Table 1 shows the structure of the experiment as it was in Naccarato and Zurlo (2008). For each scenario 500 samples have been performed both for Normal and Uniform error components' distribution.

To synthesize results of the simulation experiment two indicators have been considered: as it was in Naccarato and Zurlo (2008),

$$\varphi = \left(\hat{\theta} - \theta\right)/\theta \tag{20}$$

where $\hat{\theta}$ is the average over the 500 samples of parameters estimates and θ is the original parameter;

$$\psi = RMSE/\theta \tag{21}$$

where RMSE is the Root Mean Square Error of $\hat{\theta}$.
For both of them ratios between values presented by 3SLS and FIML with respect to FILODE are also computed.

TABLE 1
Simulation Scenarios

ρ_{ij}	S_i		
	0.20-0.25	0.4-0.5	0.75-0.80
0.1-0.2	N=20	N=20	N=20
	N=30	N=30	N=30
	N=100	N=100	N=100
0.4-0.5	N=20	N=20	N=20
	N=30	N=30	N=30
	N=100	N=100	N=100
0.8-0.9	N=20	N=20	N=20
	N=30	N=30	N=30
	N=100	N=100	N=100

6. RESULTS OF SIMULATION EXPERIMENT

Leaving to the next paragraph the comparison between the two versions of FI LODE we present here the results of the experiment with regard to the performance of FI LODE in the new version (using SVD) in comparison with that of two classical estimators like 3SLS and FIML.

We consider the results of the three methods first with regard to bias and then with regard to mean square error.

6.1. Bias

As just pointed out, to evaluate the bias of the three methods under comparison we will make use of the indicator φ defined in (20). The analysis for the three situations foreseen with regard to error components, is separately presented.

Normal error component

For Normal error component FI LODE performs better in terms of bias (φ) almost everywhere: in 20 cases out of 27 FI LODE presents the higher percentage of lowest bias. For small samples, $n = 20$, this good performances becomes more evident since 8 times out of 9 FI LODE has the higher percentage of lower bias.

With regard to average bias 16 times out of 27 FI LODE is performing better then FIML and 3SLS. Here too the situation improves for small samples (8 out of 9 cases) and increasing correlation.

To better compare the results, let's consider Table 2 in which the ratios between average bias both of 3SLS and FIML with respect to FI LODE are presented.

FI LODE's bias is very frequently smaller (values of the ratio greater than 1.05) or at most equal (values of the ratio between 0.95 and 1.05) to the one presented by FIML. It has to be stressed that this situation improves for small samples and increasing values of ρ. 3SLS always presents very high values of the ratio showing a very bad performance of the method with respect to average bias.

TABLE 2

Average bias ratio by S_i, ρ_i and sample size - Normal error component

		ρ_i					
		0.1-0.2		0.4-0.5		0.8-0.9	
S_i	Sample size	FIML/ FILODE	3SLS/ FILODE	FIML/ FILODE	3SLS/ FILODE	FIML/ FILODE	3SLS/ FILODE
0.2-0.25		1.42	5.97	2.16	3.19	1.92	5.74
0.4-0.5	20	3.88	2.21	2.66	7.00	1.72	0.71
0.75-0.8		0.76	1.42	1.68	2.19	4.82	1.91
0.2-0.25		1.18	6.64	0.97	8.72	1.03	12.55
0.4-0.5	30	1.08	5.48	0.42	4.00	14.16	7.00
0.75-0.8		0.85	5.98	6.21	10.86	23.71	13.62
0.2-0.25		0.68	25.46	0.57	21.85	0.95	30.21
0.4-0.5	100	0.96	14.18	0.50	11.19	0.40	9.40
0.75-0.8		0.57	21.85	1.14	23.49	1.21	24.11

Uniform error component $\left(-\sqrt{3}, \sqrt{3}\right)$

When Uniform $\left(-\sqrt{3}, \sqrt{3}\right)$ distribution of error component is considered results do not change substantially, even if its percentage of lowest bias reduces to 13 times out of 27. Again FI LODE is almost always performing better then FIML for high values of ρ (in the interval $0.8-0.9$) and for small samples ($n = 20, 30$). As before FILODE performs very much better then 3SLS whose bias is very high.

Uniform error component (-10, 10)

As we have already said, in order to evaluate the effect of non standard situation characterized by more scattered error components, a second Uniform distribution in the interval $\left(-10, 10\right)$ has been considered.

In this situation, FI LODE performs almost everywhere better than FIML: the percentage of times a lower bias is presented is largely in favor of FI LODE; its average

bias is largely lower than FIML's one in most of the scenarios (Table 4). Moreover, it has to be noticed that FIML average bias is substantially bigger than the FILODE one even when $n = 100$.

TABLE 3

Average bias ratio by S_i , ρ_i and sample size - Uniform error component in $\left(-\sqrt{3},\sqrt{3}\right)$

		ρ_i					
		0.1-0.2		**0.4-0.5**		**0.8-0.9**	
S_i	**Sample size**	**FIML/ FILODE**	**3SLS/ FILODE**	**FIML/ FILODE**	**3SLS/ FILODE**	**FIML/ FILODE**	**3SLS/ FILODE**
0.2-0.25		0.31	1.50	0.57	3.29	1.57	5.86
0.4-0.5	*20*	0.55	3.55	0.60	2.47	1.40	4.60
0.75-0.8		0.45	1.79	0.43	3.54	1.44	6.19
0.2-0.25		1.29	10.12	0.72	5.50	1.14	8.07
0.4-0.5	*30*	1.81	11.42	1.44	10.80	1.11	7.56
0.75-0.8		0.62	3.15	0.61	5.11	0.59	4.75
0.2-0.25		0.83	31.17	0.67	7.17	0.50	15.43
0.4-0.5	*100*	0.96	17.91	1.42	28.00	0.71	23.82
0.75-0.8		1.03	14.18	0.65	16.52	0.15	2.16

The average bias presented by 3SLS is almost everywhere bigger than the one presented by FI LODE with differences reaching 100%.

TABLE 4

Average bias ratio by S_i , ρ_i and sample size - Uniform error component in $(-10,10)$

		ρ_i					
		0.1-0.2		**0.4-0.5**		**0.8-0.9**	
S_i	**Sample size**	**FIML/ FILODE**	**3SLS/ FILODE**	**FIML/ FILODE**	**3SLS/ FILODE**	**FIML/ FILODE**	**3SLS/ FILODE**
0.2-0.25		2.74	2.00	1.53	1.13	0.69	1.09
0.4-0.5	*20*	1.66	1.06	1.14	1.52	1.01	1.08
0.75-0.8		5.32	0.85	2.71	1.29	1.02	0.98
0.2-0.25		14.36	1.81	1.44	1.98	1.71	1.21
0.4-0.5	*30*	8.29	0.82	2.90	1.20	6.60	1.00
0.75-0.8		9.25	1.24	2.73	1.16	2.47	0.97
0.2-0.25		200.32	1.38	221.66	1.00	88.75	1.79
0.4-0.5	*100*	288.59	1.11	421.31	1.13	316.46	1.30
0.75-0.8		2019.69	1.48	243.43	1.55	431.06	1.26

6.2. *Mean Square Error*

To evaluate the performance of the three methods with respect to RMSE we will make use of ψ defined in (21): relative frequency distribution of lowest values will be considered and the actual values presented. Here too, the ratio between RMSE both of 3SLS and FIML with respect to FILODE is given.

Normal error component

Looking at RMSE the situation is somehow different from the one seen for bias: The estimator that shows more frequently the lowest RMSE is FIML with 21 cases out of 27.

It has nevertheless to be noticed (Table 5) that when the differences between FIML and FI LODE are in favor of the first one they are almost always small and frequently less than or equal to 5%. As ever, FI LODE seems to work better for small samples and high correlation. Almost the same happens with regard to 3SLS that only sometimes performs better than FILODE.

TABLE 5

Average RMSE ratio by S_i, ρ_i and sample size - Normal error component

		ρ_i					
		0.1-0.2		0.4-0.5		0.8-0.9	
S_i	Sample size	FIML/ FILODE	3SLS/ FILODE	FIML/ FILODE	3SLS/ FILODE	FIML/ FILODE	3SLS/ FILODE
0.2-0.25		1.28	1.24	0.92	1.10	1.19	1.22
0.4-0.5	20	1.03	0.36	1.95	1.08	3.87	0.90
0.75-0.8		0.69	0.54	0.33	0.39	0.98	0.28
0.2-0.25		1.00	1.11	0.79	1.02	0.75	1.06
0.4-0.5	30	1.46	0.92	0.57	0.63	7.64	1.03
0.75-0.8		1.02	0.97	1.91	0.55	1.86	0.51
0.2-0.25		0.95	1.21	0.84	1.11	0.75	1.43
0.4-0.5	100	0.97	1.21	0.95	1.22	0.70	1.16
0.75-0.8		0.98	1.14	0.83	1.19	0.71	1.30

Uniform error component $\left(-\sqrt{3},\sqrt{3}\right)$

The results are very similar to those seen for Normal distribution.

While the number of times in which FIML shows a lowest value of RMSE is almost as before, the differences with FI LODE are generally higher than the ones seen in the preceding case. Furthermore it has to be stressed that the RMSE shown by 3SLS estimator is frequently the second best particularly for high values of the correlation coefficient.

Uniform error component $(-10, 10)$

As far as estimators' RMSE is concerned the comparison has to be made only between

TABLE 6

Average RMSE ratio by S_i, ρ_i *and sample size - Uniform error component in* $(-\sqrt{3}, \sqrt{3})$

		ρ_i					
		0.1-0.2		**0.4-0.5**		**0.8-0.9**	
S_i	**Sample size**	**FIML/ FILODE**	**3SLS/ FILODE**	**FIML/ FILODE**	**3SLS/ FILODE**	**FIML/ FILODE**	**3SLS/ FILODE**
0.2-0.25		0,81	0.90	0.79	1.07	0.64	0.59
0.4-0.5	*20*	0.78	0.75	0.61	0.85	0.70	0.57
0.75-0.8		0.63	0.66	0.53	0.65	0.62	0.84
0.2-0.25		0.95	1.11	0.77	1.02	0.74	0.68
0.4-0.5	*30*	0.99	1.04	0.97	1.03	0.74	0.65
0.75-0.8		0.69	0.61	0.44	0.65	0.53	0.82
0.2-0.25		0.97	1.08	0.90	1.02	0.79	0.65
0.4-0.5	*100*	0.96	1.15	0.77	1.19	0.69	0.59
0.75-0.8		0.93	1.11	1.04	1.24	0.53	0.62

LODE and 3SLS, since FIML estimators always produce higher RMSE than the other two methods. In general 3SLS is the one that presents the lowest RMSE, in accordance with what is established in literature. Only in few cases FI LODE performs better than 3SLS.

TABLE 7

Average RMSE ratio by S_i, ρ_i *and sample size - Uniform error component in* $(-10, 10)$

		ρ_i					
		0.1-0.2		**0.4-0.5**		**0.8-0.9**	
S_i	**Sample size**	**FIML/ FILODE**	**3SLS/ FILODE**	**FIML/ FILODE**	**3SLS/ FILODE**	**FIML/ FILODE**	**3SLS/ FILODE**
0.2-0.25		7.31	1.28	2.20	0.86	2.75	0.63
0.4-0.5	*20*	5.40	0.89	1.40	0.58	2.24	0.76
0.75-0.8		7.16	0.73	4.83	0.63	5.11	0.54
0.2-0.25		8.38	0.48	2.02	0.49	4.49	0.49
0.4-0.5	*30*	10.71	0.47	2.56	0.44	12.68	0.42
0.75-0.8		15.45	0.42	4.25	0.37	6.44	0.48
0.2-0.25		1212.57	1.05	826.84	0.43	154.21	0.48
0.4-0.5	*100*	1018.79	0.34	491.36	0.31	579.59	0.52
0.75-0.8		1443.01	0.35	421.59	0.44	793.35	0.38

While with respect to FIML the ratio between RMSE (Table 7) is showing a very strong prevalence of FI LODE, with respect to 3SLS the situation is almost everywhere in favour of the last one even if with very much smaller differences.

7. CONCLUSION

In this work we present a LODE estimator inspired by Gleser (1981) and by the Total Least Square procedure introduced by Golub and Van Loan (1980) and Van Huffel (2002), in which estimates of structural form parameters is obtained using the last m smallest singular right vectors associated to the m smallest singular values. The performances of this estimator are very good both in terms of bias and RMSE. In particular has to be notice that it works better than 3SLS and FIML in case of small samples and high correlation. A quite substantial improvements of the performances of FI LODE is shown in standard situation like the one considered in the Monte Carlo experiment presented here.

Few general considerations about LODE's performance can be made:

1) with regard to bias it can be said that LODE seems to perform better than the other two methods with which it is compared with. Both for Normal and Uniform distribution, LODE's bias seems to be lower than FIML almost for all scenarios; it has in particular to be stressed its good performances for small samples and increasing correlation. The same happens with 3SLS which almost in all the situations present a bias greater than FILODE
2) similar consideration can be made with regard to RMSE even if the good performance of FI LODE is a little less stringent than in the preceding case. In particular it has to be stressed its very good performance for Uniform $(-10, 10)$ errors' distribution if compared with FIML estimator which performs quite badly. It is not the same with 3SLS average RMSE which is almost everywhere the lowest one.

From these considerations it seems to appear that FI LODE performs at least as well as FIML both with respect to bias and to mean square error. With the very strong exception of the case of Uniform distribution in (-10,10) in which the latter is out-performed by the former.

If compared with 3SLS, FI LODE performs always better both for bias and mean square error. Only exception is the better performance of the first one with regard to RMSE in the case of Uniform $(-10, 10)$.

Let's now go back to the many times postponed comparison between the two versions of FILODE: the one proposed in (Pieraccini and Naccarato 2008): and the one presented here. The first one is based on SD of the errors' variance-covariance matrix related to the whole system of identifying equations, while the second one is based on SVD of errors' variance-covariance matrix related only to over identifying equations, The comparison will then take into account both the difference between variance-covariance matrices considered in the two version and the computational procedure applied (SVD and SD).

To this extent we will consider in detail only the case of Normal error component since those for both Uniform distributions do not change substantially. For the sake of simplicity the two versions of LODE will be indicated as *1st* Version the one published in

2008 and 2^{nd} Version the current one.

As ever, relative frequency distribution of lowest values will be considered and the actual values presented. Here too, the ratio between bias and mean square error of the two versions will be given.

In terms of bias the 2^{nd} *Version* of FI LODE presents the best results 19 times out of 27 with respect to the previous version of FI LODE, frequently with very high percentages.

TABLE 8

Average bias ratio by S_i, ρ_i and sample size - Normal error component

		ρ_i		
		0.1-0.2	**0.4-0.5**	**0.8-0.9**
S_i	**Sample size**	$1^{st}/2^{nd}$	$1^{st}/2^{nd}$	$1^{st}/2^{nd}$
0.2-0.25	*20*	2.00	1.90	2.25
0.4-0.5		0.53	1.63	0.31
0.75-0.8		0.59	0.77	1.04
0.2-0.25	*30*	5.00	2.50	4.33
0.4-0.5		2.67	1.85	1.64
0.75-0.8		1.67	3.67	5.33
0.2-0.25	*100*	3.67	2.75	7.75
0.4-0.5		3.17	7.00	4.83
0.75-0.8		5.80	11.33	7.00

TABLE 9

Average RMSE ratio by S_i, ρ_i and sample size - Normal error component

		ρ_i		
		0.1-0.2	**0.4-0.5**	**0.8-0.9**
S_i	**Sample size**	$1^{st}/2^{nd}$	$1^{st}/2^{nd}$	$1^{st}/2^{nd}$
0.2-0.25	*20*	1.36	1.19	1.11
0.4-0.5		0.39	1.25	1.02
0.75-0.8		1.01	0.42	0.39
0.2-0.25	*30*	1.31	1.22	1.03
0.4-0.5		1.34	0.76	1.14
0.75-0.8		1.43	0.83	0.64
0.2-0.25	*100*	1.11	1.11	1.32
0.4-0.5		1.36	1.28	1.03
0.75-0.8		1.63	1.47	1.50

The average bias of *2nd Version* of FILODE is almost everywhere lower than the other one (Table 8) sometimes presenting a very high value of the ratio (especially for $n = 20$) showing a very strong reduction of the bias. A generalized improvement of the new version with respect to bias has then to be recognized.

Also with regard to RMSE the *2nd Version* of FILODE presents a generalized improvement: here too 19 times out of 27 it is the one with the lowest average (Table 9); the ratio between average RMSE of the two versions is almost every time in favor of the new one.

From these consideration it becomes evident that the *2nd Version* of FI LODE presents a strong improvement with respect to the preceding one.

REFERENCES

T. W. ANDERSON (1976). *Estimation of linear functional relationship: approximate distribution and connections with simultaneous equation in econometrics*. Journal of the Royal Statistics Society, series B, 38, pp. 1-20.

J. G. CRAGG (1967). *On Relative Small Sample Properties of Several Structural Equation Estimator*. Econometrica, 35.

C. ECKART, G. YOUNG (1936). *The approximation of one matrix by another of lower rank*. Psychometrika, 1, no. 3.

L.GLESER, (1981). *Estimation in a multivariate "errors in variable model" regression model: Large sample results*. Ann. Statistics, 9, pp. 525-534.

G. H. GOLUB, C. VAN LOAN (1980). *An analysis of the total least square problem*. SIAM J., Number Anal. 17.

L. S. JENNINGS (1980). *Simultaneous Equation Estimation computational aspects*. Journal of Econometrics, 12, pp. 23-39.

T. C. KOOPMANS, H. RUBIN, R. B. LEIPNIK (1950). *Measuring the Equation System of Dynamic Economics*. In Statistical Inference in Dynamic Economic Models, T. C. Koopmans (eds.), New York John Wiley & Sons, Inc., Chapter 2, pp. 53-237.

I. MARKOVSKY, S. VAN HUFFEL (2007). *Overview on total least square*. Signal Processing, 87, pp. 2283-2303.

A. NACCARATO, D. ZURLO (2008). *A Monte Carlo Study on Full Information Methods in Simultaneous Equation Models*. Quaderni di Statistica, 10, pp. 115-144.

L. PIERACCINI, A. NACCARATO (2008). *Full information Least Orthogonal Distance Estimator of Structural Parameters in Simultaneous Equation Models*. Statistica, Vol. spec., pp. 52-76.

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

L. PIERACCINI (1983). *The Estimation of Structural Parameters in Simultaneous Equation Models*. Quaderni di statistica e Econometria, V, pp. 1-21.

L. PIERACCINI (1969). *Su un'interpretazione alternativa del metodo dei minimi quadrati a due stadi*. Statistica, 4 XI, pp. 786-802.

H. THEIL, A. ZELLNER (1962). *Three stage least squares: simultaneous estimation of simultaneous equation*. Econometrica, 30.

S. VAN HUFFEL (2007). *Total least square and error in variable model*. Computational Statistics and Data Analysis, 52, pp. 1076-1079.

S. VAN HUFFEL (2002). *Total Least Squares and Errors-In-Variables Modeling: Analysis, Algorithms and Applications*. Kluwer Academic Publications.

S. VAN HUFFEL (1989). *The partial total least square algorithm*. J. Computational Analysis, 93, pp. 149-162.

S. VAN HUFFEL (1988). *Analysis an properties of the generalized total least square problem.* SIAM J., Matrix Anal. Appl., 10, pp. 294-315.

SUMMARY

Least Orthogonal Distance Estimator and Total Least Square for Simultaneous Equation Models

Least Orthogonal Distance Estimator (LODE) of Simultaneous Equation Models' structural parameters is based on minimizing the orthogonal distance between Reduced Form (RF) and the Structural Form (SF) parameters. In this work we propose a new version – with respect to Pieraccini and Naccarato (2008) – of Full Information (FI) LODE based on decomposition of a new structure of the variance-covariance matrix using Singular Value Decomposition (SVD) instead of Spectral Decomposition (SD). In this context Total Least Square is applied. A simulation experiment to compare the performances of the new version of FI LODE with respect to Three Stage Least Square (3SLS) and Full Information Maximum Likelihood (FIML) is presented. Finally a comparison between the FI LODE new and old version together with few words of conclusion conclude the paper.