

## ESTIMATION METHODS FOR THE STRUCTURAL EQUATION MODELS: MAXIMUM LIKELIHOOD, PARTIAL LEAST SQUARES AND GENERALIZED MAXIMUM ENTROPY

**Anna CRISCI**

Second University of Naples

E-mail: crisci.anna@virgilio.it

### ABSTRACT

The concept of Latent Variables (LVs or latent constructs) is, probably, one of the most charming and discussed of the last fifty years, although, even today, it is only possible to give a negative definition of it: what is not observable, lacking both of origin and of measurement unit. One of the difficulties for a researcher in the economic-social sciences in the specification of a statistical model describing the causal-effect relationships between the variables derives from the fact that the variables which are object of the analysis are not directly observable (i.e. latent), for example, the performance, the customer satisfaction, the social status etc. Although such latent variables cannot be directly observable, the use of proper indicators (i.e. manifest variables, MVs) can make the measurement of such constructs easy. Thanks to the SEM, it is possible to analyze simultaneously, both the relations of dependence between the LVs (i.e., Structural Model), and the links between the LVs and their indicators, that is, between the corresponding observed variables (i.e., Measurement Model). The different and proper methodologies of estimate of the dependence are topics of this work. In particular, the aim of this work is to analyze Structural Equation Models (SEM) and, in particular, some of the different estimation methods mostly adopted: the Maximum Likelihood-ML, the Partial Least Squares- PLS and the Generalized Maximum Entropy - GME, by illustrating their main differences and similarities.

**Keywords:** Structural Equation Models, Maximum Likelihood, Partial Least Squares, Generalized Maximum Entropy.

### 1. Introduction

The growing availability of the data in the present information- based- society has underlined the need to have at our disposal the proper tools for their analysis. The "data mining" and the applied statistics are suggested as privileged tools to get knowledge from big volumes of data.

In particular, the non-homogenous and extremely complex vision of reality has urged the researchers to make use of techniques of multivariate analysis in order to analyze the relationships existing between more variables, simultaneously.

Among the different methods of multivariate analysis Structural Equation Models-SEM largely satisfy this requirement. The SEM are tools elaborated at the beginning of 1970's, and they obtained, in that decade, a lot of appreciation, and more and more spread use of them. They are the reinterpretation, arrangement and- above all- generalization of those that, in the 1970's, were called causal models and that, in the first half of the same decade, had met a remarkable popularity thanks to the technique of the path analysis.

Thanks to the SEM, it is possible to analyze, simultaneously, both the relations of dependence between the LVs (i.e., Structural Model), and the links between the LVs and their indicators, that is, between the corresponding manifest variables, MVs (i.e., Measurement Model).

The LISREL (Jöreskog, 1970; Jöreskog & Sorbom, 1989; Byrne, Barbara, 2001) or Covariance Structural Analysis (CSA) is at the bottom of such models. The Lisrel was born at the beginning as a name of software and used to estimate the structural parameters of the factorial analysis by adopting the maximum likelihood method. For many years, the Maximum Likelihood method (SEM-ML) has been the only estimation method for SEM, while, today, different estimation techniques can be used for the estimation of the SEM.

In fact, in 1975 Wold developed a soft modeling approach, making it different from the hard modeling approach of Lisrel, in order to analyze the relationships among different blocks observed variables on the same statistics units.

The method, known as PLS for SEM (SEM-PLS) or as PLS-Path Modeling (PLS-PM), is distribution free, and it was developed as a flexible technique aimed at the causal predictive analysis when the high complexity and the low theoretical information are present.

A new technique for the estimation of the Structural Equation Models has been introduced recently. In 2003 Al Nasser suggested to extend the knowledge of the information theory to the SEM context by means of a new approach called Generalized Maximum Entropy (SEM-GME). This new method is still present in the PLS- approach since no distribution hypothesis is required.

These different and proper methodologies of estimate of the dependence are topics of this work.

The paper is organized as follows: in sections 2 the SEM- Maximum Likelihood is shown; in section 3 and section 4 the SEM-PLS and SEM-GME are shown. Finally, in section 5 a table illustrating the main different/similarities among the three estimation methods is shown.

## 2. The LISREL Approach (SEM- Maximum Likelihood, ML)

As mentioned before, on the basis of Structural Equation Models, the Covariance Structure Analysis (and, thus, LISREL modeling) can be found. The CSA is a "second generation" multivariate technique (Fornell, 1987) combining methodological contributions from two disciplines: the (confirmatory) factor analysis model from psychometric theory and structural equation model typically associated with econometrics. Its aim is to explain the

structure or pattern among a set of latent variables, each measured by one or more manifest and typically fallible indicators.

There are two parts into a covariance structure model (like other approaches analyzed later): the *structural model* and *measurement model*.

The *structural model* specifies the relationships between the latent variables themselves (reflecting substantive hypothesis based on theoretical consideration). The analysis is predominantly confirmative in nature, that is, it seeks to determine the extent to which the postulated structure is actually consistent with the empirical data at hand. This is carried out by computing the implied covariance matrix by means of the specified model, and comparing it to the (actual) covariance matrix based on the empirical data.

It follows that the first equation of the Lisrel model is:

$$\eta = B \eta + \Gamma \xi + \zeta \quad (1)$$

(mx1)      (mxm) (mx1)      (mxn) (nx1)      (mx1)

where  $\eta$  (*eta*),  $\xi$  (*ksi*) e  $\zeta$  (*zeta*) are three vectors of the endogenous (variables external to the model which always perform only as independent) and exogenous (variables internal to the model that at least in a one relation perform as a dependent variable) variables, and errors, respectively.  $B$  (*beta*) and  $\Gamma$  (*gamma*) are two matrix of structural coefficients between the endogenous variables, and between the exogenous and endogenous variables, respectively. The matrix  $B$  has  $mxm$  element, that is a square matrix whose size is equal to number of the endogenous variables  $\eta$ ; moreover, its diagonal is always composed of all zeros, since they concern with the coefficient regression of each variables with itself. The matrix  $\Gamma$ , instead, is  $mxn$  order. In order to be completely specified, this part of the model needs other two matrices:

1.  $\Phi$  (*phi*) a matrix containing the variances-covariances between the exogenous latent variables  $\xi$ ;
2.  $\Psi$  (*psi*) a matrix containing variances- covariances between the errors  $\zeta$ .

These matrices are squared and symmetric.

The *measurement model* describes how each latent variables is operationalized via the manifest variables, and provides information about the validities and reliabilities of the latter

The *measurement model* for endogenous variables is:

$$y = \Lambda_y \eta + \varepsilon \quad (2)$$

(p, 1)      (p, m) (m, 1)      (p, 1)

where  $y$ ,  $\eta$  (*eta*) e  $\varepsilon$  (*epsilon*) are three vectors of the observed endogenous variables, latent endogenous and errors, respectively.  $\Lambda_y$  (*lambda y*) is the matrix of the structural coefficients between the observed variables and the latent variables; this matrix contains  $p \times m$  elements.

The matrix of variance- covariance between errors  $\varepsilon$ , is indicated with  $\theta_\varepsilon$  (*theta-epsilon*).

This matrix is a squared and symmetric, of  $p \times p$  order ( $p$  is the number of errors  $\varepsilon$ , which is equal to that of the observed variables  $y$ ); it must be, in most cases, specified as a

diagonal, that is, the variances of the error are estimated, but the covariance between the errors are set equal to zero.

The measurement model for exogenous variables is:

$$\mathbf{x} = \boldsymbol{\Lambda}_{\mathbf{x}} \boldsymbol{\xi} + \boldsymbol{\delta} \quad (3)$$

$(q, 1) \quad (q, n)(n, 1)(q, 1)$

where  $\mathbf{x}$ ,  $\boldsymbol{\xi}$  e  $\boldsymbol{\delta}$  (delta) are three vectors of the observed exogenous, latent exogenous and errors, respectively.  $\boldsymbol{\Lambda}_{\mathbf{x}}$  (*lambda x*) is the matrix of the structural coefficients between the observed variables and latent variables. This matrix contains  $qxn$  elements.

The matrix of variance-covariance between errors  $\boldsymbol{\delta}$ , is indicated with  $\Theta_{\boldsymbol{\delta}}$  (*theta-delta*), it is a matrix squared and symmetric, of  $qxq$  order, and is also specified, in most cases, diagonal.

## 2.1 The Maximum Likelihood Estimation

Since the half of the 1960's Maximum Estimation-ML (Jöreskog 1970) has been the predominant estimation method. The ML is an estimation technique referred to Lisrel approach defined Covariance-Based, whose objective is to reproduce the covariance matrix of the MVs by means the model parameters. The ML estimation implies that the MVs follow a Multivariate Normal distribution. The analysis is predominantly confirmative in nature, that is, it seeks to determine the extent to which the postulated structure is actually consistent with the empirical data at hand. This is carried out by computing the implied covariance matrix produced by the specified model starting from parameter estimation ( $\hat{\boldsymbol{\Lambda}}$ ), and by comparing it to the (actual) covariance matrix based on the empirical data ( $S$ ).

Yet, in order to be able to continue, we need to compute the probability of obtaining  $S$  given  $\hat{\Sigma}$ . This is possible by means of the so-called Wishart distribution that defines such a probability. In the 1928 Wishart computed the probability function of the distribution  $S$ , hence called Wishart distribution. The sample covariance matrix  $S$  with general terms  $(s_{ij})$  is :

$$S = \frac{1}{N-1} \sum_{i=1}^N (x_{ij} - \bar{x}_j)(x_{ij} - \bar{x}_j)' = \frac{1}{n} \sum_{i=1}^N (x_{ij} - \bar{x}_j)(x_{ij} - \bar{x}_j)' \quad (4)$$

where:  $n=N-1$ , e  $\bar{x}_j = \frac{1}{N} \sum_{i=1}^N x_{ij}$ . where:  $n = N - 1$ , e  $\bar{x}_j = \frac{1}{N} \sum_{i=1}^N x_{ij}$

It can proved (Ghosh and Sinha,2002) that  $nS$  follows a Wishart distribution as:

$$W(\hat{\Sigma}, n) = \frac{\exp\left[-\frac{n}{2}\text{tr}(S\hat{\Sigma})\right]|nS|^{1/2(n-p-1)}}{|\hat{\Sigma}|^{n/2} 2^{np/2} \pi^{p(p-1)/4} \prod_{i=1}^p \Gamma(n+1-i)/2} \quad (5)$$

If we join all the constant terms (which do not depend on  $\hat{\Sigma}$ ) and we call this combined  $T$ , the equation 5 it may be re-written as follows:

$$W(S; \hat{\Sigma}, n) = \frac{\exp\left[-\frac{n}{2}\text{tr}(S\hat{\Sigma}^{-1})\right]}{|S|^{n/2}} T \quad (6)$$

A statistical function is developed using the likelihood ratio (Neyman and Pearson, 1928) that compares any theoretical model with a perfect fitting model, that is, the distance between the hypothesized model and theoretically perfect model.

The likelihood ratio is defined by:

$$\text{Likelihood ratio} = \frac{\text{likelihood for any given model under } H_0}{\text{likelihood with a perfectly fitting model under } H_1} \quad (7)$$

$$\text{Likelihood ratio} = \frac{\exp\left\{-\frac{n}{2}\text{tr}(S\hat{\Sigma}^{-1})\right\}|\hat{\Sigma}|^{-n/2}}{\exp\left\{-\frac{n}{2}\text{tr}(SS^{-1})\right\}|S|^{n/2}} \quad (8)$$

We can observed that  $\hat{\Sigma}$  has been replaced by  $S$  in the denominator of equation (8), because this represent a perfect model.

$$\text{Likelihood ratio} = \exp\left\{-\frac{n}{2}\text{tr}(S\hat{\Sigma}^{-1})\right\}|\hat{\Sigma}|^{-n/2} \exp\left\{-\frac{n}{2}\text{tr}(SS^{-1})\right\}|S|^{n/2} \quad (9)$$

Taking the natural logarithm of both sides of this equation:

$$\begin{aligned} \text{log likelihood ratio} &= \left\{-\frac{1}{2}\text{tr}(S\hat{\Sigma}^{-1})\right\} - \frac{n}{2}\log|\hat{\Sigma}| + \frac{1}{2}\text{tr}(SS^{-1}) + \frac{n}{2}\log|S| \\ &= -\frac{n}{2}\{\text{tr}(S\hat{\Sigma}^{-1}) + \log|\hat{\Sigma}|\} - \text{tr}(SS^{-1}) - \log|S| \end{aligned} \quad (10)$$

where  $SS^{-1}$  is identity matrix .

Therefore, following the ML approach for the estimation of the structural model of covariance- CSA o Lisrel (topic of the present study and that can be considered a generalization of the factorial analysis model inside a more complex and sophisticated system), the discrepancy function to be minimized (by means of derivation operation, that is by calculating the partial derivatees compared to the single parameters to be estimated) is:

$$F_{ML} = \text{tr}(S\hat{\Sigma}^{-1}) + \ln|\hat{\Sigma}| - \ln|S| - (p+q) \quad (11)$$

where  $q$  is the number of  $X$  variables and  $p$  is the number of  $Y$  variables.

It is important to underline that, as  $\hat{\Sigma}$  converges to  $S$ , the inverse  $\hat{\Sigma}$  approaches  $S$ , and  $S\hat{\Sigma}^{-1}$  approaches the identity matrix  $SS^{-1}$ . Since the identity matrix has the ones on the diagonal, the trace of  $S\hat{\Sigma}^{-1}$  will be equal to width of the matrix , i.e a  $(p+q)$ . The ML function to be minimized is distributed as follows:

$$(N-1)*F_{ML} \sim \chi^2 [(p+q)(p+q+1)-t] \quad (12)$$

where  $t$  is the number of free parameters (i.e to be estimated).

### 3. An approach based on Partial Least Squares-Path Modeling (PLS-PM or SEM-PLS)

The PLS Path Modeling is a statistical method which has been developed for the analysis Structural Models with latent variables. As opposed to the covariance-based approach (LISREL), the aim of PLS to obtain the scores of the latent variables for predicted purposes without using the model to explain the covariation of all the indicators. According to Chin (1988), the estimation of the parameters are obtained by basing on the ability of minimizing the residual variances of all dependent variables (both latent and observed).

PLS –Path Modeling aims to estimate the relationships among  $M$  blocks of variables, which are expression of unobservable constructs. Specifically, PLS-PM estimates the network of relations among the manifest variables and their own latent variables, and the latent variables inside the model through a system of interdependent equations based on simple and multiple regression.

Formally, let us usually assume  $P$  manifest variables on  $N$  units. The resulting data  $x_{n\times p}$  are collected in a partitioned table of standardized data  $\mathbf{X}$  :

$$\mathbf{X} = [X_1, \dots, X_m, \dots, X_M],$$

where  $X_m$  is the generic  $m$ -th block.

It is important to point out that in SEM literature there is no arrangement on the notation used to define the latent variables and all the other parameters of the model. As the matter of fact, as seem for Lisrel, the exogenous and endogenous latent variables, as the manifest variables and the parameters, are noted in different way, while in the PLS-PM all latent variables are expressed in the same way without considering their role in relation similar the regression. For this purpose, in this study no distinction un terms of notation between exogenous and endogenous construct is made, and all latent variables are defined as  $\xi$ 's.

The path models in the PLS involve three sets of relations:

1. **Inner Model or Model Structural**, which refers to the structural model and specifies the relationships between the latent variables LVs. Latent variable can play both predictand role an predictor one; a latent variable which is never predicted is called an exogenous variable, otherwise, it is called endogenous variable. The structural model can be expressed as :

$$\xi_m = B \xi_m + \zeta_m \quad (13)$$

where  $B$  is the matrix of all the path coefficients in the model. This matrix indicates the structural relationship between LVs.  $\zeta_m$  is the inner residual term, and the diagonal variance/covariance matrix among inner terms is indicated with  $\Psi$ .

2. **Outer Model or Measurement Model**, which refers to the measurement model and specifies the relationships between the constructs and the associated indicators MVs. Two ways to establish these links can be distinguished as follows:

- **Reflective way**: in which the indicators (manifest variables) are regarded to be reflections

or manifestations of their latent variables: a variation of the construct yields a variation in the measures. As a result, the direction of causality is from the construct to the indicator. Each manifest variables represents the corresponding latent variable, which is linked to the latent variable by means of a simple regression model. The reflective indicators of the latent construct should be internally consistent, and, as it is assumed that all the measures are indicators equally valid of a latent construct, they are interchangeable. The reflective measures are at the basis of the theory of the classical tests, of the reliability estimation, of and factorial analysis, each of them considers the manifest variable  $x_{pm}$  being a linear combination of its latent variable  $\xi_m$ .

$$x_{pm} = \lambda_{pm} \xi_m + \varepsilon_{pm} \quad (14)$$

where  $\lambda_{pm}$  is the generic loading coefficient associated to the  $p$ -th manifest variable in the  $m$  block, and we indicate with  $\Lambda$  the matrix containing all the loading coefficients in the block.  $\varepsilon_{pm}$  represents the generic outer residual term associated to the generic manifest variable and the corresponding diagonal variance/ covariance matrix is indicated with  $\Theta_\varepsilon$ .

- **Formative way**: in which the indicators are regarded as causes of their latent constructs: a

variation of the measures yields a variation in the construct. As a result, the direction of causality is from the indicator to the construct. The elimination of items that have low correlations compared with the overall indicators will compromise the construct validation, narrowing the domain.

This is one of the reasons by which the reliability measures of the internal consistency should not be used to estimate the fitting of the formative models. Moreover, the multi-collinearity between the indicators may be a serious problem for the parameter estimations of the measurement model when the indicators are formative, but it is a good point when the indicators are reflective. The latent variable  $\xi_m$  is assumed to be a linear combination of its manifest variables  $x_{pm}$ :

$$\xi_m = \sum_p \pi_{pm} x_{pm} + \delta_m \quad (15)$$

**3. Weight relations,** the specification of the relations between LVs and their set of indicators is carried out at a conceptual level. In other words, the outer relations refer to the indicator and the "true" LV, which is unknown. As a result, the weight relations must be defined for completeness. The estimation of LVs are defined as follows:

$$\widehat{\xi}_m = \sum_p \widehat{w}_{pm} x_{pm} \quad (16)$$

where,  $\widehat{w}_{pm}$  are the weights used to estimate the LV as a linear combination of their observed MVs.

In order to estimate the parameter, two double approximations for LVs are considered by PLS algorithm (Wold, 1982; Tenenhaus, 1999):

- **the outer approximation or external estimation**, called  $v_m$ , is used for the measurement model. In this stage we find an initial proxy of each LV,  $\xi_m$ , as a linear combination of its MVs  $x_{pm}$ . The external estimation is obtained as the product of the block of MVs and the outer weights  $\widehat{w}_{pm}$ ;

- **the inner approximation or internal estimation**, called  $\vartheta_m$ , is used for the structural model. The connections among LVs are taken into account in order to get a proxy of each LV worked out as weighted aggregate of its adjacent LVs. The internal estimation is obtained as the product of the external estimation  $v_m$  (of  $\xi_m$ ) and the so-called inner weights,  $e_{mm'}$ .

There are three ways to calculate the internal weights:

- **centroid scheme (Wold):** II centroid scheme is the scheme of the original algorithm by Wold. This scheme considers only the direction of the sign among the latent variables  $e_{mm'} = \text{sign}\{\text{cor}(v_m, v_{m'})\}$ .

- **factorial scheme (Lohmöller):** this scheme uses the correlation coefficients,  $e_{mm'} = \{\text{corr}(v_m, v_{m'})\}$ , as internal weights instead of using only the correlation sign and, therefore, it considers not only the direction of the sign but also the power of link of the paths in the structural model.

- **path weighting scheme:** in this case the latent variables are divided into predictors and followers according to the cause- effect relations between the two latent variables. A variable can be either a follower (if it is yielded by another a latent variable), or a predictor (if it is the cause of another latent variable).

$e_{mm'} = \text{simple/multiple OLS regression coefficient of } v_m \text{ on } v_{m'} \text{ if } \xi_m \text{ is explanatory of } \xi_{m'}$ ;  
 $e_{mm'} = r_{mm'} \text{ if } \xi_m \text{ is explanatory of } \xi_{m'}$ ;

Once a first estimation of the latent variables is obtained, the algorithm goes on by updating the outer weight. There are two ways to calculate the outer weights:

**1. Mode A:** is preferred when the indicators are linked to their latent variables by means of the reflective way, in which each weight  $w_{pm}$  is the coefficient regression of  $\vartheta_m$  in the simple regression of  $x_{pm}$  on  $\vartheta_m$ , that is the simple regression  $x_{pm} = w_{pm}\vartheta_m$  in which:

$$w_{pm} = (\vartheta_m' \vartheta_m)^{-1} \vartheta_m' x_{pm} = \text{cor}(x_{pm}, \vartheta_m) \quad (17)$$

**2. Mode B:** is preferred when the indicators are linked to their latent variables by means of the formative way, in which  $\vartheta_m$  is regressed on the block of indicators linked to the latent construct  $\xi_m$ , and the  $w_m$  of weight  $w_{pm}$  is the regression coefficients in the multiple regression:

$$\vartheta_m = \sum_p w_{pm} x_{pm}$$

and it is defined by means of:

$$w_m = (X_m' X_m)^{-1} X_m' \vartheta_m \quad (18)$$

The algorithm is iterated till convergence, which is demonstrated to be reached for one and two-block models. However, for multi-block models, convergence is always verified in practice. After convergence, structural (or path) coefficients are estimated through an OLS multiple regression among the estimated latent variable scores.

### 3.1 Summary

The PLS algorithm works on centered (or standardized) data, and it starts by choosing arbitrary weights (e.g 1,0..0). Chin (1999) suggested starting with equal weights for all indicators (the loadings are set to 1) to get a first approximation of the LVs as a simple sum of its indicators starts with arbitrary initial weights used to calculate an external approximation of the LVs. The inner relations among LVs are considered to estimate the internal approximation by choosing three options: centroid, factoring and path scheme. After obtaining the internal approximation, the algorithm turns around the external relations with the estimate of outer weights obtained by means of Mode A (Reflective) or by Mode B (Formative). The procedure is repeated until convergence of the weights is obtained. Once convergence of the weights is obtained and LVs are estimated, the parameters of the structural and measurement models are calculated by means of the Ordinary Least Squares (OLS).

## 4. An approach based on the Generalized Maximum Entropy or SEM-GME

Golan et al. (1996) suggested an alternative method to estimate the parameters for the regression models in case of ill-conditioned problem, as an extension of the measurement of entropy by Shannon and generalization of Maximum Entropy Principle (MEP) by Jaynes. The method is called Generalized Maximum Entropy (GME) and it is based on the re-parameterization and re-formulation of the generalized linear model  $y = X\beta + \varepsilon$  with  $n$  units and  $m$  variables, in such a way to estimate the parameters by means of the MEP developed by Jaynes, according to the following equation:

$$y_{(n,1)} = X_{(n,m)} \cdot \beta_{(m,1)} + \varepsilon_{(n,1)} = X_{(n,m)} \cdot Z_{(m,m,M)} \cdot p_{(m,M,1)} + V_{(n,n,N)} \cdot w_{(n,N,1)} \quad (19)$$

It is always possible to write the parameters  $\beta_k$  as a convex combination of the variables of finite support variables, in this case five of them in,  $\{z_{k1}, z_{k2}, z_{k3}, z_{k4}, z_{k5}\}$  (Paris,2001); this means:

$\beta_k = \{p_{k1} \cdot z_{k1} + p_{k2} \cdot z_{k2} + p_{k3} \cdot z_{k3} + p_{k4} \cdot z_{k4} + p_{k5} \cdot z_{k5}\}$ , where the probabilities  $0 \leq p_{kj} \leq 1$ ,  $j = 1, \dots, 5$

$$e \sum p_{kj} = 1.$$

Likely, each error term is dealt with as a discrete random variables.

The matrices  $Z$  e  $V$  are diagonal matrices, whose diagonal elements are vectors of support variables:

$$\beta = Z \cdot p \begin{bmatrix} z_1 & 0 & \dots & 0 \\ 0 & z_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & z_k \end{bmatrix} \begin{bmatrix} p_1 \\ p_2 \\ \vdots \\ p_k \end{bmatrix}; \varepsilon = V \cdot w \begin{bmatrix} v_1 & 0 & \dots & 0 \\ 0 & v_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & v_T \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_T \end{bmatrix} \quad (20)$$

The support variables are defined by the vectors  $z$  e  $v$ , whose dimension, usually from 2 to 7 elements, is identified by the numbers of fixed point  $M$  and  $N$ , respectively. The value "c" is a constant distributed in a symmetrical way around zero; in this application  $c=1$  and  $M$  and  $N$  are equal to 5.

$$z_k = \left[ -c \ -\frac{c}{2} \ 0 \ \frac{c}{2} \ c \right] v_k = \left[ -c \ -\frac{c}{2} \ 0 \ \frac{c}{2} \ c \right] \quad (21)$$

The vectors  $p$  and  $w$  are the probabilities associated with the regression coefficients  $\beta$  and with the error terms  $\varepsilon$ , respectively. The aim is to estimate these probabilities in such a way to represent the coefficients and the error terms as expected value of a discrete random variable.

$$p'_k = [p_{k1} \ p_{k2} \ p_{k3} \ p_{k4} \ p_{k5}] \quad w'_k = [w_{k1} \ w_{k2} \ w_{k3} \ w_{k4} \ w_{k5}] \quad (22)$$

The estimation of unknown parameters  $p$  and  $w$  is obtained through the maximization of entropy function by Shannon:

$$H(p, w) = -p'_{1,m,M} \cdot \ln p_{1,m,1} - w'_{1,n,N} \cdot \ln w_{1,n,1} \quad (23)$$

subjected to consistency and normalization constraints.

The former represent the information generated from the data, this implies a part of the model defined in the equation (17), while the latter identify the following conditions:

$$0 \leq p_{kj} \leq 1, \{j = 1, \dots, 5; k = 1, \dots, m\},$$

$$\sum p_{kj} = 1 \{k = 1, \dots, m\} \text{ e } 0 \leq w_{kj} \leq 1, \{j = 1, \dots, 5; k = 1, \dots, n\}, \sum w_{kj} = 1 \{k = 1, \dots, n\}.$$

#### 4.1 Generalized Maximun Entropy for SEM

The SEM based on the GME start from the classical formulation by Jöreskog, in which we can distinguish the equations of the structural and measurement models. In particular, this approach considers the re-parameterization of the unknown parameters and of the error terms, as a convex combination of the expected value of a discrete random variable. The equations (1), (2) and (3) can be re-formulated in just one function as:

$$Y_{(p,1)} = \Lambda_{(p,m)}^y (I_{(m,m)} - B_{(m,m)})^{-1} \{ \Gamma_{(m,n)} \Lambda_{(n,q)}^{x-1} (X_{(q,1)} - \delta_{(n,1)}) + \zeta_{(m,1)} \} + \varepsilon_{(p,1)} \quad (24)$$

where  $I$  is the identity matrix,  $\Lambda^{x-1}$  is the general inverse of  $\Lambda^x$ . The matrices of the coefficients  $B, \Gamma, \Lambda_x, \Lambda_y$  and the matrices of variance-covariance  $\Phi, \Psi, \Theta_\varepsilon, \Theta_\delta$ , are re-parametrised as an expected value of the random variables with  $M$  fixed points for the coefficients  $j$  and the errors.

GME provides a measurement of the normalized index that quantifies the level of the information generated from the model on the ground of the data collected, providing a

global measurement of the goodness of adaptation of the relation, assumed in the linear model of the simultaneous equations.

The normalized entropy measure can be expressed by means of the following expression:

$$S(\bar{p}) = \frac{-\bar{p}' \cdot \ln \bar{p}}{K \ln M} \quad (25)$$

This index of normalization is a measure of the reduction of the uncertain information, in which  $-\bar{p}' \ln \bar{p}$  is the entropy function by Shannon, while  $K$  the number of the predictors and  $M$  is the number of fixed points (FPs). The quantity  $K \cdot \ln M$  represents the maximum uncertainty. If  $S(\bar{p})=0$  there is no suitable to explain the data. The maximum entropy measure can be used as a method to select the explicative variables.

## 5. LISREL, PLS-Path Modeling e GME: differences and similarities

This section shows the main differences and similarities among the three estimation methods analyzed in the previous sections.

In particular, we have started out from the table suggested by Sanchez (2009) for Lisrel and PLS and, by keeping the same style by Sanchez, we have re-elaborated (in the past authors, such as Al-Nasser and Ciavolino Enrico, developed some aspects of GME) the same features for GME.

**Table 1: LISREL, PLS-Path Modeling e GME: differences and similarities**

	Lisrel (Covariance Structure Analysis)	PLS Path Modeling	GME
<b>Object</b>	<b>Parameter Oriented:</b> objective is to reproduce the covariance matrix of the MVs by means the model parameters.	<b>Description-Prediction Oriented:</b> obtain the scores of the latent variables for predicted purposes without using the model to explain the covariation of all the indicators	<b>Estimation precision-prediction oriented:</b> maximize the "objective function = <b>Shannon's entropy function</b> ", emphasizing both estimation precision and prediction
<b>Approach</b>	<b>Covariance-based:</b> the residual covariances are minimized for optimal parameter accuracy.	<b>Variance-based:</b> aims at explaining variances of dependent variables (observed and unobserved) in regression sense (i.e residual variances are minimized to enhance optimal predictive power).	<b>Theoretic Informantion- based:</b> under Jaynes' maximum entropy (uncertainty) principle, out of all those distribution consistent with the data-evidence we choose the one that maximizes the entropy function and thus maximizes the missing information, in order to get models based on real data..
<b>Optimality</b>	If the hypothesized model is correct in the sense of explaining the covariations of all indicator, CSA provides optimal estimates of the parameters (i.e offers statistical precision in the context of stringent assumptions).	PLS trades parameter efficiency for prediction accuracy, simplicity, and fewer assumptions.	- the GME provides the estimation in case of negative freedom degrees; -uses all the information in the data; - is robust relative to the underlying data generation process and to the limited-incomplete of economic data; -performs well relative to competing estimators under a squared error measure performance;

<b>Type of fitting algorithm</b>	Simultaneous estimation of parameters by minimizing discrepancies between observed and predicted Covariance/correlation matrix. <b>Full information method</b>	Multi-stage iterative procedure using OLS. Subset of parameters estimated separately <b>A limited information method.</b>	The estimation of the parameters is obtained by the maximization of the Shannon's entropy function subject to consistency and normalization constraints. <b>Full information method.</b>
<b>Conception</b>	Used more as an auxiliary tool for theory testing.	Used more as a decision making tool, with emphasis on parsimonious prediction.	Used as a tool to solve problems called ill-conditioned, where the lack of information and / or specific data about the problem at hand requires the recruitment of general assumptions as possible with respect to the parameters of the system under study.
<b>LV scores</b>	Indeterminate. Indirect estimation computed with the whole set of MVs.	LVs explicitly estimated as linear combination of their indicators.	Each a LV is re-parameterized as a convex combination of a discrete random variable.
<b>Relationships between the LVs and MVs</b>	Typically only with reflective indicators.	Reflective and formative indicators.	Reflective and formative indicators.
<b>Treatment of measurement residuals</b>	Combines specific variance and measurement error into a single estimate.	Separates out irrelevant variance from the structural portion of the model.	The variance/covariance matrix $\Psi, \Theta_{\varepsilon}, \Theta_{\delta}$ , are re-parametrization as a expected value of a discrete random variable.
<b>Manifest Variables</b>	Continuous and interval scaling	Continuous, interval scaling, categorical.	Continuous, interval scaling, categorical.
<b>Assumed distributions</b>	Multivariate normal if estimation through Maximum Likelihood.	No distribution assumptions	Semi-parametric
<b>Sample size</b>	High >200 unit	Medium 40 < unit < 200	Low 10 < unit < 40
<b>Model correctness</b>	To the extent that the theoretical model is correct it is able to explained the covariations of all indicators.	To the extent that the theoretical model is correct it is determined partly from the power of the relations of path between the LVs.	To the extent that the theoretical model is correct it is determined by the chance to obtain a set of consistent relations based on data.
<b>Consistency of estimators</b>	Consistent, given correctness of model and appropriateness of assumptions.	Bias estimators tend to manifest in higher loading coefficients and low path coefficients. The bias is reduced when both the size and the number of indicators for the LVs increase. <b>(consistency at large).</b>	Consistent and asymptotically normal under four mild conditions: <ol style="list-style-type: none"> <li>1. The error support spans a uniform and symmetrical around zero;</li> <li>2. The parameter support space contains the true realization of the unknown parameters;</li> <li>3. The errors are independently and identically distributed;</li> <li>4. The design matrix is of full rank (Golan 2003:5).</li> </ol>

<b>Missing Value</b>		Nipals algthm	Maximum Likelihood method
<b>Evaluation Model</b>	Evaluation Model by means hypothesis testing: <b>Chi-square:</b> the H0 hypothesis is: $S - \chi^2 = 0$	-R <sup>2</sup> for dependent LVs; -GoF(Amato et.al 2004) - resampling (jackknifing and bootstrapping) to examine the stability of estimation.	-Normalized index of entropy that quantified the level of information generated from the model on the bases of the collected data. -Pseudo R <sup>2</sup> .
<b>Applicability</b>	<ul style="list-style-type: none"> <li>• The phenomena analyzed are clear;</li> <li>• Low complexity of the model;</li> <li>• Presumes the use of reflective indicators;</li> <li>• Usually stringent assumptions about the distribution, independence, large sample size;</li> <li>• Treatment of hierarchical data, multi-group;</li> <li>• Comparison of models which come from different populations with a single objective function.</li> </ul>	<ul style="list-style-type: none"> <li>• Relatively new phenomena or mutant;</li> <li>• Relatively complex model with a large number of indicators and / or latent variables;</li> <li>• Epistemological need to model the relationship between LVs and indicators in different ways (formative and reflective);</li> <li>• Hypothesis normality, independence and the sample size is not met;</li> <li>• Multi-group.</li> </ul>	<ul style="list-style-type: none"> <li>• Complex model with incomplete data and small sample size;</li> <li>• Use both reflective and formative indicators;</li> <li>• It is easier to impose non-linear constraints;</li> <li>• Does not require distributional hypothesis;</li> <li>• Multi-group, hierarchical data;</li> <li>• Ability to insert a priori information on the model.</li> </ul>

## 6. Conclusion

This work has had the purpose of illustrating the structural equation models and, in particular, the three estimation methods mostly used in the econometric applications, showing the main differences and similarities of them.

The Covariance Structure Analysis (and, thus, Lisrel model) approach belonging to **Covariance-based approach**. The aim of Covariance-based techniques is to reproduce the sample covariance matrix by the model parameters. In other words, model coefficients are estimated in such a way to reproduce the sample covariance matrix. In the covariance based approach, the measurement model is typically considered as reflective, the multivariate normal must be respected if estimation is carried out by means of the ML and works on large sample.

The PLS approach is, instead, **Variance-based**, i.e strongly prevision oriented, whose aim is to obtain the scores of the latent variables for predicted purposes without using the model to explain the covariation of all the indicators. According to Chin (1988), the estimates of the parameters are obtained by basing on the ability of minimizing the residual variances of all dependent variables (both latent and observed). The PLS does not require items which follow a multivariate normal distribution and adopts both formative and reflective indicators and works on small samples properly.

Finally, the GME approach is **Theoretical Information-based** whose aim is to maximize the entropy function and, thus, maximizes the missing information, in order to

obtain model based on real data. This estimate technique remains in the optic the PLS approach since it does not require any distribution assumption (Ciavolino & Al-Nasser, 2006 demonstrated that the GME approach for SEM seems to work better than the PLS-PM when outliers are present.

## Reference

1. Al-Nasser, A. D., **Customer Satisfaction Measurement Models: Generalized Maximum Entropy Approach**. *Pak Journal of Statistics*, 19(2), 2003 213–226.
2. Amato, S., Esposito Vinzi, V., and Tenenhaus, M., **A global Goodness-of-Fit for PLS structural equation modelling. Oral communication to PLS Club**, HEC School of Management, France, March 24, 2004.
3. Byrne, Barbara M., **"Structural Equation Modeling with AMOS,EQS, and LISREL: Comparative Approaches to testing for Factory Validity of a Measuring Instrument"** *International Journal of Testing*, 2001, p.55-86.
5. Bagozzi, R.P, **"Structural Equation Models in Experimental Research"**, *Journal of Marketing Resarch*,14, 1977, pp 209-226.
6. Ciavolino, E. and Al-Nasser, A., **Comparing generalized maximum entropy and partial least squares methods for structural equation models**. *Journal of Non parametric Statistic* Vol. 21, No.8, 2009, 1017-1036.
7. Ciavolino, E., Al Nasser, A.D., and D'Ambra, A., **The Generalized Maximum Entropy Estimation method for the Structural Equation Models**, GFKL 2006, Berlino, marzo 2006.
8. Chin, W.W., and Newsted, P.R., **Structural Equation Modeling Analysis with Small Samples using Partial Least Squares**. In: *Statistical Strategies for Small Sample Research*, 1999, 307-341. Hoyle R. (Ed). London: Sage Publications.
9. Chin, W.W, **Partial Least Squares for research: An overview and presentation of recent advance using the PLS approach**, 2000, p.68.
10. Edwards, Jeffrey R. and Bagozzi, P.R, **"On the Nature and Direction of Relationships between Construct an Measures"**, *Psychological Method*,5(2), 2000, 155-74.
11. Esposito Vinzi, V., Chin, W.W., Henseler,J., and Wang. H, **Handbook of Partial Least Squares, Concepts, Methods and Application**, Springer Handbooks of Computational Statistics, New York, 2010.
12. Fornell, C., **"A Second Generation of Multivariate Analysis: Classification of Method and Implications for Marketing Research"**. In: *Review of Marketing*. (Ed) Houston, M.J. (Chicago), American Marketing Association, 1987, pp. 407-450

13. Fornell, C., and Bookstein, F.L., **Two Structural equation model: LISREL and PLS applied to consumer exit-voice theory.** *Journal of Marketing Research*, 19, 1982, 440-452.
14. Golan, Amos and Judge, G. and Miller, D., **The Maximum Entropy Approach to Estimation and Inference: An Overview**, Staff General Research Papers 1327, Iowa State University, Department of Economics, 1997.
15. Golan, A., Judge G., and Perloff J., **Estimation and inference with censored and ordered multinomial response data.** In: *Journal of Econometrics*, 79 (1997), 23-51.
16. Golan A, Judge G, Miller D, **Maximum entropy econometrics: robust estimation with limited data.** Wiley, New York, 1996.
17. Jaynes, E.T, **Information Theory and Statistical Mechanics.** *The Physical Review* 106(4), 620-630, May 15,1957.
18. Jöreskog, K.G., **Factor Analysis and Its Extensions. Factor Analysis at 100: Historical Developments and Future Decisions.** Department of Psychology, University of North Carolina. May 13-15, 2004.
19. Jöreskog, K.G., and Wold, H., **The ML and PLS Techniques for Modeling with Latent Variables: Historical and Comparative Aspects.** In: *Systems under indirect observation: Causality, structure, prediction. Part I*, 1982, 263-270. K.G.
20. Jöreskog, K.G., **"Testing Structural Equation Models".** In: *Testing Structural Equation Models*. (Ed) Bollen, K.A and Lang, J.S., Beverly Hills, Sage, 1993, pp. 294-316.
21. Jöreskog, K.G and Sorbon, D., **LISREL 7: A Guide to the Program and Application**, Chicago, IL, SPSS, Inc., 1989
22. Jöreskog, K.G and Sorbon, D., **LISREL 7: User's Reference Guide.** In: Scientific Software, Inc., Mooresville, 1989.
23. Tenenhaus, M., **L'Approche PLS**, Statistquè Applique XLVII(2), 1999, 5-40.
24. Tenenhaus, M., **Structural Equation Modeling for small samples.** HEC school of management (GRECHEC), 2008.
25. Tenenhaus, M., Esposito Vinzi, V., Chatelin,Y.M, and Lauro,C., **PLS Path Modeling**, *Computational Statistics & Data Analysis* 48, 2005, pp.159-205.
26. Wishart, J., **The Generalised Product Moment Distribution in Sample from Normal Multivariate Population.** *Biometrika*, Vol 20 A, No ½, Jul,1928, pp 32-52.
27. Wold, H., **Soft modeling: The Basic Design and Some Extensions.** In: *Systems under indirect observation: Causality, structure, prediction. Part II*, 1-54. K.G. Jöreskog and H. Wold (Eds). Amsterdam: North Holland, 1982.
28. Wold, H., **Non Linear estimation by iterative least squares procedures.** In: *Researche Paper Statistics, Festchrft for J. Neyman*, 411-444. F.David (Ed). New York: Wiley, 1966.

29. Wold, H., **Estimation of Principal Components an Related Models by Iterative Least Squares**, In: Multivariate Analysis, Proceedings o fan International Symposium, 391-430, P.R. Krishnaiah (Ed). New York: Academic Press, 1966.
31. Sanchez, G., **Partial Least Squares**, Phd thesis, 2009.
32. Shannon, C.E., **A mathematical theory of communication**. Bell system Technical Journal 27, 1948, 379-423.