BETWEEN EXPLORATIVE AND CONFIRMATIVE ESTIMATION METHODS FOR THE STRUCTURAL EQUATION MODELS

Enrico Ciavolino,
Università del Salento, Dipartimento di Filosofia e Scienze Sociali
enrico.ciavolino@ateneo.unile.it

Abstract

The aim of this work is to show an integrated approach for using the Structural Equation Model (SEM) with an information theoretic proposal when the classic approach doesn't work.
The estimation methods used are the Partial Least Squares (PLS), for a non-parametric and explorative analysis and the Maximum Likelihood Estimation (MLE) method, for parametric and confirmative analysis.
In this respect, the innovatory aspects are the follows: (i) the estimation methods are used as an statistical integrated approach, in order to apply the potentiality and the main characteristics of both of them; (ii) the extension of the SEM model in case both of the estimation methods cannot perform well the data, by considering the information theory of the Generalized Maximum Entropy (GME).

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

1. Introduction

In this paper the *Partial Least Squares* (PLS) and the *Maximum Likelihood Estimation* (MLE) are used as an *statistical integrated approach*, in order to use the potentiality and the main characteristics of both of them.
The methodology, called *Generalized Maximum Entropy* (GME), that can integrate some of the main characteristics of the PLS and MLE is here presented.
The paper is organized as follows. In section 2, it is presented the PLS and MLE estimation methods as integrated approach. Section 3 illustrates the GME formulation for avoiding the problem as small and possibly ill-behaved, noisy data, or in case of endogeneity.

2. The PLS & MLE as Integrated Estimation Methods for SEM

The following paragraphs reports the main characteristics of the two traditional approaches in estimating SEM parameters and a proposal for using them as a integrated approach.

2.1 The Partial Least Squares Estimation Method

The main idea of the PLS for the structural equation models, is an iterative combination of path analysis, for give a measure of the relationships among the theoretical constructs (*Structural Model* or *Inner Model*, equation 1), and factorial analysis, for measuring the latent construct (*Measurement Model* or *Outer Model*, equation 2):

$$\boldsymbol{\xi}_{(m,1)} = \mathbf{B}_{(m,m)} \cdot \boldsymbol{\xi}_{(m,1)} + \boldsymbol{\tau}_{(m,1)} \tag{1}$$

$$\mathbf{x}_{(q,1)} = \boldsymbol{\Lambda}_{(q,m)} \cdot \boldsymbol{\xi}_{(m,1)} + \boldsymbol{\delta}_{(q,1)} \tag{2}$$

PLS (Chin, 2000, Fornell & Bookstein, 1982) can be a powerful estimation method of analysis in case of small sample size, strong correlation among the items, missing data and no residual distribution assumption.
In the *Structural Model* (1), ξ is the vector of the m latent variables and **B** is the path coefficients matrix, with zeros on its diagonal, of the causal effect among the latent variables.

The *Measurement Model* (2) contains the **x** vector of the *q* manifest variables and the coefficient matrices **Λ** of the relationships between the latent constructs and the observed variables.
The vectors **τ** and **δ**, are the structural and the measurement errors vectors and the **Ψ** and $\mathbf{\Theta}^{\delta}$ are respectively the diagonal matrices variance, of the structural error term **τ** and the measurement error term **δ**.
For the parameters estimation, the PLS algorithm (Wold, 1982; Tenenhaus, 1999) considers two double approximation for the latent variables ξ_j:

- the *external estimation*, called $\mathbf{y}_j$, obtained as product between the block of manifest variables $\mathbf{X}_j$, of the j^{th} latent variable ξ_j, and the so called *outer weights* $\mathbf{w}_j$.
- the *internal estimation*, called $\mathbf{z}_j$, obtained as product between the external estimation $\mathbf{y}_i$ (of ξ_i) and the so called *inner weights* $e_{j,i}$.

The PLS algorithm starts with the initialization of the *outer weights* $\mathbf{w}_j$, equal for the first step, to 1 for the first manifest variable and 0 to the remaining ones. After the initialization, the routine estimates the other parameters in according to the following iterative three estimation steps: (*i*) *external estimation*; (*ii*) *internal estimation*; (*iii*) *outer weights estimation*.
The procedure is repeated until the convergence between the internal (y_t-y_{t-1}→0) and the external estimates (z_t-z_{t-1}→0). The estimation of the parameters matrix **B** of the causal path relationships, it is accomplished using Ordinary Least Squares (OLS) method.

2.2 Maximum Likelihood Estimation

Jöreskog (1970) developed the maximum likelihood estimation (MLE) method for the SEM parameters. This method, also called *LiSRel* (Linear Structural Relationships), is based on the comparison between a theoretical Covariance Matrix ∑, defined by theoretical hypothesis, and a Covariance Matrix **S,** observed on the sample.
The Structural Equation Model is defined by two main parts: the first part is the *Structural Model* (4), and represents the linear relationships among the latent variables; the second part, is the *Measurement Model* and represents the relationships between the manifest and the latent variables, endogenous (5) and exogenous (6).

$$\boldsymbol{\eta}_{(m,1)} = \mathbf{B}_{(m,m)}\boldsymbol{\eta}_{(m,1)} + \boldsymbol{\Gamma}_{(m,n)}\boldsymbol{\xi}_{(n,1)} + \boldsymbol{\tau}_{(m,1)} \quad (4)$$

$$\mathbf{y}_{(p,1)} = \boldsymbol{\Lambda}^{y}_{(p,m)}\boldsymbol{\eta}_{(m,1)} + \boldsymbol{\varepsilon}_{(p,1)} \quad (5)$$

$$\mathbf{x}_{(q,1)} = \boldsymbol{\Lambda}^{x}_{(q,n)}\boldsymbol{\xi}_{(n,1)} + \boldsymbol{\delta}_{(q,1)} \quad (6)$$

The model specification has some assumptions: The variables are centered, E(**η**)= E(**ξ**)= E(**τ**)= 0; E(**y**)= E(**ε**)= 0; E(**x**)= E(**δ**)= 0; The independent variables and the error terms are un-correlated, E(**ξ τ'**)= 0; E(**η ε'**)=0; E(**ξ δ'**)= 0; E(**η δ'**)= 0; E(**ξ ε'**)= 0; The error terms are un-correlated, E(**τ ε'**)= 0; E(**τ δ'**)= 0; E(**ε δ'**)= 0; The matrix **B** is not singular
The eight above matrices defined, the coefficient matrices, **B**, **Γ**, $\mathbf{\Lambda}^{y}$ and $\mathbf{\Lambda}^{x}$, and the co-variance matrices, **Φ**, **Ψ**, $\mathbf{\Theta}^{\varepsilon}$, $\mathbf{\Theta}^{\delta}$, identify the theoretical Covariance Matrix ∑.
The parameters estimation is made by comparing ∑ and **S** through numerical maximization of a fit criterion, as provided by likelihood function:

$$Log \quad L = -\left[\frac{N-1}{2}\right]\left[Log\left|\boldsymbol{\Sigma}\right| + tr(\mathbf{S}\boldsymbol{\Sigma}^{-1})\right] \quad (7)$$

The likelihood function (7) is maximized numerically by using the Davidon-Fletcher-Powell method. For measuring the fit of the estimated model and the observed covariance matrix, a common measure used is the Chi-Square, where the correspondence p-value is usually more than the 10%.

2.3 A proposal for an integrated use of the PLS and MLE

The PLS, the non parametric estimation method, can be used in order to carry out preliminary theoretical hypothesis to explore the potential relationships among Latent Variables.
The main characteristic of the PLS method is to find a local optimal solution by a series of interdependent OLS regression, minimizing residual variance under the so called Fixed Point (FP) method (Wold, 1965, 1981, Fornell and Bookstein, 1982). This characteristic is one of the limit of the PLS, because it doesn't optimize any global scalar function, so that, it doesn't provide an index for evaluating the global validation of the model.
On the other hand this can suggest also an approach that can be seen as the selection of sub-models from the main model analyzed, by considering the PLS and MLE in complementary way, where the FP criterion can suggest the best sub-model based on local measure of fit (Local R^2).
The *Sub-Models* so suggested, can illustrate particular aspects of the SEM. They can be studied and validated by using the parametric MLE method, considering all the constraints (error distribution, covariance structure) that a parametric approach needs, obtaining representative population models.

3. The GME Estimation Method

The GME method represents a new *semi-parametric* estimation method for the Structural Equation Model (Al-Nasser, 2003). The GME for the SEM can be seen as an extension of the GME application for the *simultaneous equations system* (Zellner, 1962) already developed by Golan, Judge and Miller (1996, Wiley).
With respect to the PLS and MLE has some desirable properties which can be briefly summarized in the following points:

- The GME approach uses all the data points and does not require restrictive moments or distributional error assumptions.
- Thus, unlike the MLE estimator, the GME is robust for a general class of error distributions.
- The GME estimator may be used when the sample is small, where there are many covariates, and when the covariates are highly correlated.
- Moreover, using the GME method, it is easy to impose nonlinear and inequality constraints.
- The GME provides the measure of the *normalized entropy measure* (Golan *et al*, 1996) that quantifies the level of information in the data, giving a global goodness of fit index.

The objective of the GME is to recover the unknown parameters of the SEM, with minimal distributional assumptions. The GME approach for the SEM considers the *Re- Parameterization* of the unknown parameters and the disturbance terms, as a convex combination *of expected value of a discrete random variable*. The three equations [4, 5, 6] can be *Re-Formulated* as a unique function model as:

$$\mathbf{Y}_{(p,1)} = \mathbf{\Lambda}^{y}_{(p,m)}\left(\mathbf{I}_{(m,m)} - \mathbf{B}_{(m,m)}\right)^{-1}\left\{\mathbf{\Gamma}_{(m,n)}\mathbf{\Lambda}^{x-1}_{(n,q)}\left(\mathbf{X}_{(q,1)} - \boldsymbol{\delta}_{(q,1)}\right) + \boldsymbol{\tau}_{(m,1)}\right\} + \boldsymbol{\varepsilon}_{(p,1)} \qquad [8]$$

The coefficient matrices, **B**, **Γ**, $\mathbf{\Lambda}^{y}$, $\mathbf{\Lambda}^{x}$, and the co-variance matrices, **Φ**, **Ψ**, $\mathbf{\Theta}^{\varepsilon}$, $\mathbf{\Theta}^{\delta}$, defined in the MLE session, are all *Re-Parameterized* as expected values of discrete random variable with *M* fixed points for the coefficients and *J* for the errors. The Re-Parameterized coefficients are so defined: $\mathbf{B}_{(m,m)} = \mathbf{Z}_{(m,m\cdot M)} \cdot \mathbf{P}^{\mathrm{B}}_{(m\cdot M,m)}$; $\mathbf{\Gamma}_{(m,l)} = \mathbf{Z}_{(m,l\cdot M)} \cdot \mathbf{P}^{\Gamma}_{(l\cdot M,l)}$; $\mathbf{\Lambda}^{y}_{(r,m)} = \mathbf{Z}_{(r,m\cdot M)} \cdot \mathbf{P}^{\Lambda^{y}}_{(m\cdot M,m)}$; $\mathbf{\Lambda}^{x}_{(q,l)} = \mathbf{Z}_{(q,l\cdot M)} \cdot \mathbf{P}^{\Lambda^{x}}_{(l\cdot M,l)}$ and the *J* fixed points for the error terms: $\boldsymbol{\zeta}_{(m,1)} = \mathbf{V}_{(m,j\cdot J)} \cdot \mathbf{G}^{\zeta}_{(j\cdot J,1)}$; $\boldsymbol{\varepsilon}_{(r,1)} = \mathbf{V}_{(r,j\cdot J)} \cdot \mathbf{G}^{\varepsilon}_{(j\cdot J,1)}$; $\boldsymbol{\delta}_{(q,1)} = \mathbf{V}_{(q,j\cdot J)} \cdot \mathbf{G}^{\delta}_{(j\cdot J,1)}$.
Given the *Re-Parameterization* and the *Re-Formulation*, the GME system can be expressed as a non-linear programming problem subject to linear constraints. The coefficients and the error terms are estimated by recovering the probability distribution of the discrete random variables set. The vectors $\mathbf{p}^{\mathbf{B}}$=*vec*($\mathbf{P}^{\mathbf{B}}$), $\mathbf{p}^{\mathbf{\Gamma}}$=*vec*($\mathbf{P}^{\mathbf{\Gamma}}$), $\mathbf{p}^{\mathbf{\Lambda y}}$=*vec*($\mathbf{P}^{\mathbf{\Lambda y}}$), $\mathbf{p}^{\mathbf{\Lambda x}}$=*vec*($\mathbf{P}^{\mathbf{\Lambda x}}$), are obtained by using the *vec*

operator of the matrices $\mathbf{P^B}$, $\mathbf{P^\Gamma}$, $\mathbf{P^{\Lambda y}}$, $\mathbf{P^{\Lambda x}}$. The vectors: $\mathbf{p^B}$, $\mathbf{p^\Gamma}$, $\mathbf{p^{\Lambda y}}$, $\mathbf{p^{\Lambda x}}$, $\mathbf{G^\zeta}$, $\mathbf{G^\varepsilon}$, $\mathbf{G^\delta}$, are calculated by the maximization of the following entropy function:

$$H(\mathbf{p}^{B},\mathbf{p}^{\Gamma},\mathbf{p}^{\Lambda^y},\mathbf{p}^{\Lambda^x},\mathbf{G}^{\zeta},\mathbf{G}^{\varepsilon},\mathbf{G}^{\delta}) = -\mathbf{p}^{B'}\cdot\ln\mathbf{p}^{B}-\mathbf{p}^{\Gamma'}\cdot\ln\mathbf{p}^{\Gamma}-\mathbf{p}^{\Lambda^{y}'}\cdot\ln\mathbf{p}^{\Lambda^y}-\mathbf{p}^{\Lambda^{x}'}\cdot\ln\mathbf{p}^{\Lambda^x}-\mathbf{G}^{\zeta'}\cdot\ln\mathbf{G}^{\zeta}-\mathbf{G}^{\varepsilon'}\cdot\ln\mathbf{G}^{\varepsilon}-\mathbf{G}^{\delta'}\cdot\ln\mathbf{G}^{\delta} \quad [9]$$

Subjected to the *consistency* and *normalization constraints*. This method can be seen as a intermediated approach (semi-parametric) that can be used in same case. Based on some simulation studies (Ciavolino & Al-Nasser, 2006) it is showed that in comparing the PLS and MLE estimation methods for small sample sizes, the GME estimation method outperforms the PLS for most cases in terms of MSE. However, for moderate sample sizes the presence of outliers make both methods rather equivalent. The best simulation result is obtained in the multicollinearity experiments, where, for two degrees of association, the GME estimation outperforms PLS, either for different sample sizes.

Bibliography

Al-Nasser, A. D., 2003, *Customer Satisfaction Measurement Models: Generalized Maximum Entropy Approach*, Pak Journal of Statistics, 19(2): 213–226.

Bollen, K. A. 1989, Structural Equations with Latent Variables. New York: John Wiley and Sons.

Chin W., 2000, "Partial Least Squares for Researchers: An Overview and Presentation of Recent Advances Using the PLS Approach", http://disc-nt.cba.uh.edu/chin/indx.html.

Ciavolino E., Al Nasser A.D., D'Ambra A., 2006, The Generalized Maximum Entropy Estimation method for the Structural Equation Models, *GFKL 2006*, Berlino, marzo 2006.

Ciavolino E., Dahlgaard J.J. (2007), ECSI - Customer Satisfaction Modelling and Analysis: a case study, in press on Total Quality Management & Business Excellence, Volume 18, Issue 5, 545 – 554.

Duncan, O.D., 1975, *Introduction to Structural Equation Models*. New York: Academic Press.

Fornell C., Bookstein F. L., 1982, Two structural equation models: LISREL and PLS Applied to Consumer Exit-Voice Theory, Journal of Marketing Research, Vol 19, n. 4, pp 440-452.

Golan A, Judge G, Miller D ,1996, Maximum entropy econometrics: robust estimation with limited data. Wiley, New York.

Jöreskog, K. G., 1973, *A general Method of Estimating a Linear Structural Equation system*. In Goldberg, S. A. & Duncan, D. O (Eds). Structural Equation Models in the Social Sciences. New York: Seminar Press. 85-112.

Tenenhaus, M., 1999, L'Approche PLS, Statistique Appliquèe XLVII (2): 5-40.

Wold, H., 1965, A fixed-point theorem with econometric background, I-II. Arkiv for Matematik, 6, 209–240.

Wold, H., 1981, The Fixed Point Approach to Interdependent Systems. Amsterdam: North Holland.

Zellner, A., 1962, *An efficient method of estimating seemingly unrelated regressions and tests of aggregation bias*. Journal of the American Statistical Association 57, 348-368.