AN ALTERNATING LEAST SQUARE APPROACH FOR THE ESTIMATION OF A SEM BASED ON ORDINAL VARIABLES*

D. Nappo, M.G. Grassia

Department of Mathematic and Statistic-University "Federico II" of Naples

Abstract

The aim of this paper is to propose an approach to quantify the qualitative variables, within Structural Equation Models (SEM), and in particular of PLS-PM.

We propose a new algorithm, called Partial Alternating Least Squares Optimal Scaling- Path Modeling (PALSOS-PM), which through an iterative procedure, computes an optimal quantification, for qualitative variables, and structural parameters of the model chosen.

1. INTRODUCTION

For data, in the social sciences, there are, usually, four scales of measurement that must be considered: nominal, ordinal, interval and ratio (Stevens, 1951). However, choice of the statistical analyses typically rests on a more general or cruder classification: "categorical" scale for qualitative variables and "continuous" scale for quantitative variables. Ordinal scales with few categories (2,3, or possibly 4) are often classified as categorical, whereas ordinal scales, with many categories (5 or more), are classified as continuous. Although Likert-type items are technically ordinal scales, but most researchers treat them as continuous variables.

When you want to jointly analyze data measured on different scales is necessary to homogenize variables. Thus, it is quite common practice to recode quantitative variables into qualitative ones. One needs to be careful to converting variables measured by continuous scale into categorical or dichotomous ones also because it can be as problematic if you want to use them for the analysis of Statistical Models which were developed for quantitative variables.

* *This paper is supported by a PRIN project 2006 (Multivariate statistical models for the ex-ante and the ex-post analysis of regulatory impact) coordinated by Carlo Lauro*

So, it is possible saying that the dichotomy, between categorical and continuous scales, is an oversimplification and the recoding, of all data in qualitative variables, is a mistake mainly in the marketing researches, in public-opinion surveys and in social researches, where data are often measured by ordinal scales and analyzed with Structural Equation Models (SEM).

We propose, in this paper, a method of quantification of these ordinal variables: we have developed in a Structural Equation Model, based on the Partial Least Squares method (PLS), an original algorithm that pursues the optimal quantification of ordinal variables and nominal variables according to an Alternating Least Squares (ALS) logic.

To validate the algorithm and verify the advantages of this procedure that uses a quantification for each ordinal variable we used the dataset "mobile" published in the paper of Tenenhaus et al (2005). The results are compared with those obtained with the traditional algorithm PLS-PM in which ordinal variables are processed as quantitative variables.

2. THE QUANTIFICATION OF THE ORDINAL VARIABLES

A quantification for ordinal and nominal variables based on an *optimal scaling* technique is largely used in literature to analyze the data through statistical methods developed for continuous variables.

The *optimal scaling* is defined by Bock as "*the assignment of numerical values to alternatives or categories, so as to discriminate optimally among the objects, in some sense. Usually it is the least squares sense, and the values are chosen so that the variance between objects after scaling is a maximum with respect to that within objects*".

The *optimal scaling* techniques can be classified in three categories:

- methods of *optimal scaling* drawn through *scale construction* models
- methods of *optimal scaling* drawn through an *objective function*
- methods of *optimal scaling* obtained simultaneously with the estimation of parameters.

The first, the *scale construction,* is characterized and based on the definition and construction of a scale of values, as Likert, Guttman and Rash scale. A prior quantification based on external *optimal scaling* technique is applied. The simplicity and applicability of these methods have rendered them the tools more used to measure and quantify the ordinal/nominal variables.

Despite their simplicity, they have some disadvantages. In particular only the approach of Rash produces, using a logistic model, numerical variables.

Moreover, it is important to bear in mind that, in presence of categorical and continuous variables, this approach does not allow to have a unique function to optimize. The Rash model does not permit, finally, to estimate simultaneously the quantification and the parameters of the model: these ones could be estimated only after the process of quantification and by another algorithm. The second approach is based on the definition of an *objective function* coherent with the analysis to be develop. The *optimal scaling* is integrated inside the methodology of data analysis or modeling. This approach has the advantage to obtain the quantification across the maximization of a criterion, but, if it is used in a Structural Equation Model, the function optimized does not express a casual model, and so this method does not allow to estimate in a unique function the parameters of a model and the optimal quantification.

The third approach is based on the *Alternating Least Squares Optimal Scaling* (ALSOS) algorithms, in which, according to the analysis to be develop, the *optimal scaling* is a step useful to maximize the relationship between the variables optimally quantified. The ALSOS algorithms are based on the Kruskal proposal (1964), that transforms the qualitative variables with assumption that the relationships, among the quantified variables (dependent and independent), are linear. This approach has these properties:

- the optimal scaling is used to quantify nominal or ordinal variables, contextualizing the process in the general analysis to be develop;
- the estimation of the parameters and the quantification are two different steps, alternated, that take inspiration from the non metric ANOVA of Kruskal (1964) and the Factorial Analysis (Kruskal-Shepard, 1974);
- the starting point is the algorithm HOMALS that develops a Multiple Correspondence Analysis, from which other methods are derived, adding some constraints on the parameters.

What kind of *optimal scaling techniques* are used in SEM and in particular in the Partial Least Squares-Path Modeling (PLS-PM)?

In the PLS-PM, the basic idea is to assume the continuity for the ordinal variables, treating them as numerical, and to dichotomize the nominal variables, increasing the dimension of the raw matrix.

Sometimes are used *scale construction* models or, in the specific case of the ordinal variables, an equidistribution linear normalization.

In the literature there is a proposal of E. Jakobowicz and C. Derquenne (2006) for the estimation of a SEM, with the algorithm of PLS-PM, in presence of qualitative manifest variables. They propose a new algorithm, called Partial Maximum Likelihood (PML), based on Generalized Linear Models (GLM) . They modify the first step of the PLS-PM algorithm, according to the nature of manifest variables (nominal or ordinal). The authors introduce the concept of reference variable as the initial estimation of the latent variable: it is a manifest variable of any latent block associated to the *j-th* block that is supposed to better explain the latent concept. The vector of the initial weights will be equal to (Lohmöller has demonstrated that for any initial vector of weights the algorithm of PLS-PM converges):

$$w_{jh}^{0} = \mathrm{cov}(x_{jh}, x_{i1}) \quad (1)$$

where x_{i1} is the reference variable chosen between the blocks associated to *j*-ith block. The authors propose a series of statistical methodologies well known in literature, as the Logistic and ANOVA model, whose differences between themselves are related to the nature of the variable x_{jh} and the reference variable x_{i1}. The *inner* estimation is the same as in the classical PLS-PM algorithm, while for the *outer* estimation it is important to consider the nature of the manifest variables; in particular if the manifest variable x_{jh} is quantitative, the algorithm proceeds in the classical way, while if it is qualitative, it is used, to obtain a new estimation of the latent variable, the variance model. In this case for each category of the manifest variable is computed a weight, that corresponds to the coefficient of the variance analysis.

This approach has the advantage to make the quantification of the qualitative variables by an internal procedure to the classical algorithm of PLS-PM and it is possible to introduce all kind of variables, and for the qualitative ones, each of them quantified according to its nature.

On the other hand there are two disadvantages:

- the properties[1] of the algorithm of PLS-PM are lost;
- the weights of qualitative and quantitative variables are not estimated in the same way[2].

Another proposal in the literature is of P.G. Lovaglio (2002), who developed an algorithm for the estimation of a multivariate regression model with mixed variables (dependent and independent): this algorithm computes a regression

---

[1] *The properties of the algorithm of PLS-PM are: absence of distributional assumption, the possibility to apply the technique to matrixes with a number of individuals minor than of number of variables.*

[2] *Where there are mixed variables, the weights of quantitative variables are computed as the covariance between manifest and latent variables, instead of the case of ordinal or nominal variables the weights are either the means of the values of these variables or the coefficient of logistic regression.*

model in which there are a set of manifest variables *X*, that are explicative, and a set *Y* of manifest variables that are dependent and that define a latent variable. The algorithm estimates, alternating two steps, the best quantification for the variables X and Y (in the case in which the sets are composed by qualitative variables) and the best estimation of the parameters of the model. So the algorithm belongs to the family of ALSOS and, in particular, the proposal of Lovaglio is based on the join between two approaches: the first is the Non Linear Regression of the set Y on X to obtain the optimal quantification for both variables, and the second step is the Principal Component Analysis to obtain the estimation of the latent variable as the first component of $\hat{\boldsymbol{Y}}'\hat{\boldsymbol{Y}}$. These two steps are alternating until the convergence and the results are the estimation of the regression coefficients and the optimal quantification for the qualitative variables. This algorithm allows the possibility to introduce in a regression model all kind of variables and, therefore, it can be considered an alternative proposal to the LISREL approach to estimate a latent concept, measured by indicators and causes, using the Non Linear Principal Component Analysis (NPCA). The fundamental characteristic of this algorithm is the simultaneous estimation of the vector of scaling and of the parameters of the model (regression coefficients).

3. PALSOS-PM ALGORITHM

Aim of this paper is presenting a method of quantification of ordinal variables: we have developed in a Structural Equation Model, based on the Partial Least Squares Path Modeling method (PLS-PM), an original algorithm that pursues the optimal quantification of ordinal variables, and also nominal variables, according to an Alternating Least Squares (ALS) logic. The algorithm is called Partial Alternating Least Squares Optimal Scaling- Path Modeling (PALSOS-PM), because it has the structure of the algorithm of PLS-PM, but it uses, as method of *Optimal Scaling*, an Alternating Least Squares algorithm to obtain the optimal quantification for manifest variables and the inner and outer estimation of the model.

The principal characteristic of this approach is the absence of distributional assumption and the possibility to introduce all kind of variables in same block of manifest variables. PALSOS-PM algorithm has some characteristics of PLS-PM and some of Alternating Least Square algorithms: of the first, it has the basic structure (*inner* and *outer* estimation and the Path Analysis) and the inner estimation of the latent variables, of the second, it has the process of quantification, modifying the estimation of the outer weights, because it takes into account the nature of the variables. The PALSOS-PM algorithm obtains the best coefficients of the model and the best quantification for the variables by the use of Morals (Young et al. 1976) algorithm, that computes simultaneously the parameters of a regression model and the optimal scaling vectors for the manifest variables.

PLS-PM approach (Wold, 1982) is based on alternated simple and multiple regressions steps. Scores are determined by alternating an external estimate of the *Latent Variable* (LV) and an internal one until convergence.

In the external estimate ($\boldsymbol{v}_q$) each LV (ξ_q) is obtained as a linear combination of its *Manifest Variable* (MV) $\boldsymbol{x}_{pq}$ $(p = 1, \ldots, P_q)$:

$$v_q \propto \pm\left(\sum_{p=1}^{P_q} w_{pq} x_{pq}\right) \tag{2}$$

Then, basing on this external estimate $\boldsymbol{v}_q$, the internal estimate $\mathbf{z}_q$ of each LV ξ_q in relation to the other LVs is obtained:

$$z_q \propto \sum_{q=1}^{k} e_{qq'} v_{q'} \tag{3}$$

where $\boldsymbol{e}_{qq'}$ are the internal weights usually set equal to the sign of the correlation coefficient between the external estimates of the q^{th} and the q'^{th} LVs. The symbol $\propto$ means that internal and external estimates of LVs are standardized.

Then, the next step is the estimate of the external weights ($\boldsymbol{w}_{pq}$) to be assigned to each MV (in the new external estimate step) of the corresponding LV. They can be computed according the reflective (mode A) or formative (mode B) scheme. In particular, in a reflective block, each relationship can be expressed as a simple regression model where the weights are the simple regression coefficients between each MV and the LV; in a formative block, one LV depends on its MVs, then its relations define a multiple regression model, where $\mathbf{X}_\mathrm{q}$ is the explanatory MV matrix and the LV is the response and where the weights are estimated like multiple regression coefficients.

Partial results for external weights are used for the new external estimate of LV. The estimation algorithm alternates these 3 steps until convergence between internal and external estimates is reached. Finally, LV estimates are used in a set of multiple and/or simple regression analyses for determining the structural relations (among LV), or *path-coefficient*.

The Morals algorithm optimizes the multiple correlation between a single criterion variable and a set of predictor variables where any of the variables (criterion included) may be measured by nominal, ordinal, interval and ratio scales.

The variables do not all have to be measured at the same level nor does the process, which is assumed to have generated data, may be either discrete or continuous. Morals obtains an optimal scaling for each variable within the restrictions imposed by the regression model, the measurement level, and the generating process. The scaling is optimal in the Fisher sense of optimal scaling: the multiple correlation is maximized. It is based on the minimization of a quadratic function in respect to three parameters.

The independent qualitative variables (nominal and ordinal) are specified as the product between an indicator matrix $\mathbf{G_j}$ ($n*k_j$) and a vector of the scaling parameters $\mathbf{y_j}$ (k_j*1), that after estimation defines the variables $\mathbf{x_j^{os}} = \mathbf{G_j y_j}$. This procedure is made also for the dependent variable **z**, that becomes $\mathbf{z^{os}} = \mathbf{G_z t}$, where **t** is the vector of scaling of the dependent variable **z**. The loss function to optimize is:

$$\min_{\beta,t,y_j} SSQ(G_z t - Gy\beta) \tag{4}$$

with the constraints $u'G_z t = 0, t'G_z'G_z t = 1$ with $y_j \in C_j$ and $t \in C_z$, where u is a vector of 1 and C_j and C_z are the spaces of admissible transformation (they are closed convex cones) for the categories of each variable, taking into account the level of measurement. In particular, if the variables are nominal there are not constraints on the values of quantification, while if the variables are ordinal there are the constraints of order between the categories.

So the final objects of this technique are to obtain the best quantification of the qualitative variables and to optimize the regression parameters; obtained, in fact, the first estimation of the vectors $\mathbf{z^{os}}$ and $\mathbf{x^{os}}$ (*os* is the acronym of optimal scaling) the parameters of multiple regression are updated using as variables the new obtained to the previous step, reiterating the steps until the convergence.

In PALSOS-PM this procedure is done for each block of latent variables, separately.

The algorithm starts with an arbitrary quantification of the manifest variables (the typical coding of a questionnaire). The algorithm proceeds with the inner estimation of latent variables, and then it updates the outer weights by Morals. So in the external estimate:

$$v^{os}{}_q = \pm\left(\sum_{p=1}^{P_q} w^{os}{}_{pq} x_{pq}\right) \tag{5}$$

The internal estimate $\mathbf{z}_q$ of each LV ξ_q in relation to the other LVs is obtained:

$$z^{os}_{q} \propto \sum_{q'} e_{qq'} v^{os}_{q'} \tag{6}$$

We have chosen to use this iterative algorithm for three reasons: the first is the possibility to estimate the relationship between variables in the reflective and formative case; the second is its capability to treat simultaneously different kind of variables, because the quantification step is individually; the third is the simultaneous estimation of the relationship between the manifest and latent variables and the best quantification.

The steps of PALSOS-PM algorithm are described in the figure 1.

Consider the matrix X of manifest variables and define the path diagram

Step1

Compute a first casual vector of weights w_{pq}

repeat

Step2

for (q *in* $1{:}P_q$)

$$v_q = \left(\sum_{p=1}^{P_q} w_{pq} x_{pq} \right)$$

endfor

Step3

for (q in 1:k)

$$z_q = \sum_{q} e_{qq'} v_{q'}$$

The weights $e_{pp'}$ *can be computed by the Centroid, Factorial or Path scheme, as in the PLS-PM*

endfor

Step4

Update the estimation of weights w_{ij} *by Morals*

$W_{pq} = cov(x_{pq}, z_q)$ ***reflective***

$W_{pq} = (X'X)^{-1} X'Z$ ***formative***

Step5

Ceck the convergence

$|w^{old}_{pq} \quad w^{new}_{pq}| <= 10 \; ^{5}$ break

Fig. 1: PALSOS-PM algorithm.

When the algorithm returns to the outer estimation, a new quantification is obtained by Morals. The PALSOS-PM algorithm besides to estimate a SEM model with ordinal or qualitative variables, allows to estimate a model with all quantitative variables: in this case it computes the parameters by the classical PLS-PM algorithm.

For the validation of the *outer* and *inner* model, we use the bootstrap technique to create suitable interval confidence, because the quantification procedure is not based on distributional assumptions. Therefore, information about the variability of the parameter estimates and hence their significance is generated by means of resampling procedures.

PALSOS-PM takes into account, during the estimation of the parameters, the problem of the signs and, as in the PLS-PM, it solves it using the comparison of the signs of the eigenvectors.

4. PALSOS-PM AT WORK: THE MOBILE DATASET

The PALSOS-PM algorithm is here applied, for comparative aims, to a dataset used in the work of Tenenhaus et al. (2005), in which an ECSI model is estimated to evaluate the customer satisfaction. The European Costumer Satisfaction Index (ECSI) is an economic indicator that measures customer satisfaction. A model has been derived specifically for the ECSI. In this model, seven interrelated latent variables are introduced. It is based on well-established theories and approaches in customer behaviour and it is to be applicable for a number of different industries. ECSI is an adaptation of the Swedish customer satisfaction barometer (Fornell, 1992) and is compatible with the American customer satisfaction index. The entire model is important for determining the main target variable, being Customer Satisfaction Index. The ECSI model is described in figure 2.

A set of manifest variables is associated with each of the Latent Variables. This model is applied to a sample of 250 customers of a mobile society, to evaluate their satisfaction respect to the services received.

The manifest variables are 24 and are so subdivided in the latent blocks:

- five manifest variables for the block Image;
- three manifest variables for the block Expectation;
- seven manifest variables for the block Perceived quality;
- two manifest variables for the block Perceived value;
- three manifest variables for the block Customer Satisfaction;
- one manifest variable for the block Complaints;
- three manifest variables for the block Loyalty.

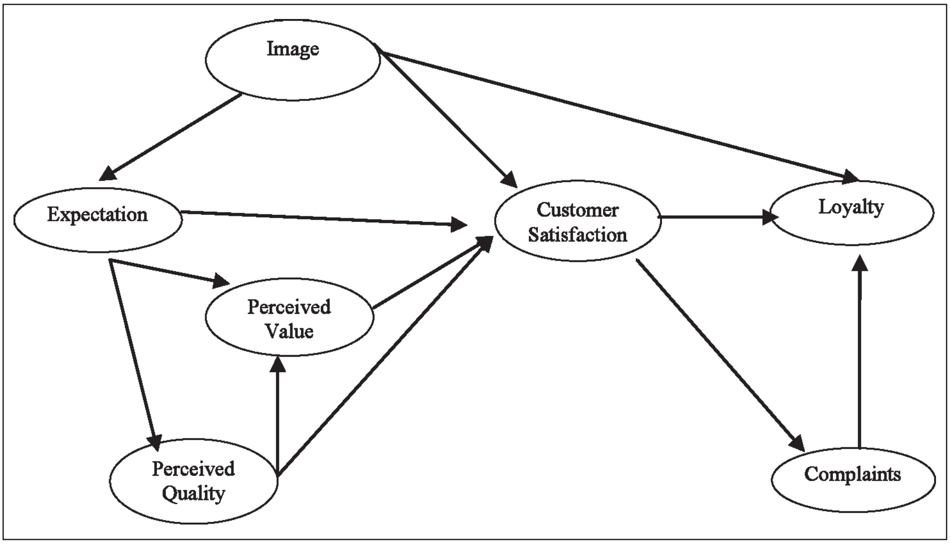

Fig. 2: The model for the evaluation of Customer Satisfaction.

All variables are ordinal and express on a scale of ten values, so the PALSOS-PM algorithm is used to obtain an optimal quantification for these variables and to estimate the structural parameters of the model. This model was estimated with the classical algorithm of PLS-PM with the software X-LSTAT, and with the algorithm PALSOS-PM with the quantification of the manifest variables.

In the figures 3 and 4 the results of the inner estimation are reported, allowing to compare the PALSOS-PM with those of the classical PLS-PM.

We can see that, respect to the model estimated with PLS-PM algorithm, only one parameter assumes a negative sign, and so it is no-significant: *Expectation* on *Customer Satisfaction* (this parameter has a negative value for the T-Statistic), while with the PLS-PM algorithm also the relationship between *Expectation* and *Perceived Value* is no-significant. Besides, the variable with a good impact on the *C. Satisfaction* is the *Perceived Quality* (0.595) (as also with the PLS-PM), followed by *Image* (0.228) and *Perceived Value* (0.194): the difference between the two algorithms is in the estimation of parameters that in PALSOS-PM is optimized by Morals.

It is confirmed the strong impact of *Perceived Quality* on the *Perceived Value* (0.758), like it is interesting to note as the *C. Satisfaction*, with PASOS-PM algorithm, has a strong impact on the *Loyalty* (0.570), respect to PLS-PM (0.195). So the quantification has produced an improvement in the estimation of the inner relationship between the latent variables, strengthening existing relationships and validating other universally rejected.

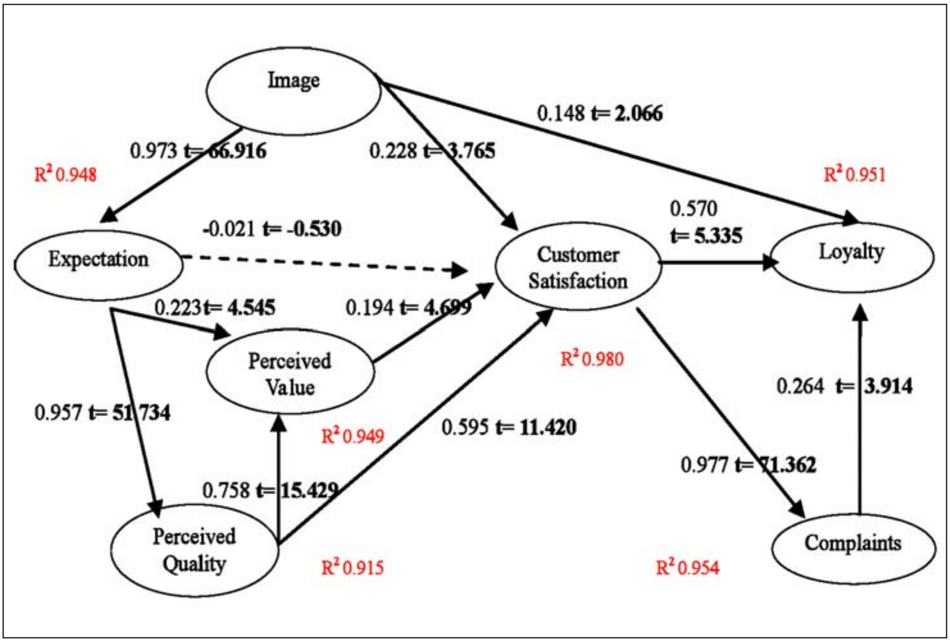

Fig. 3: The model with PALSOS-PM.

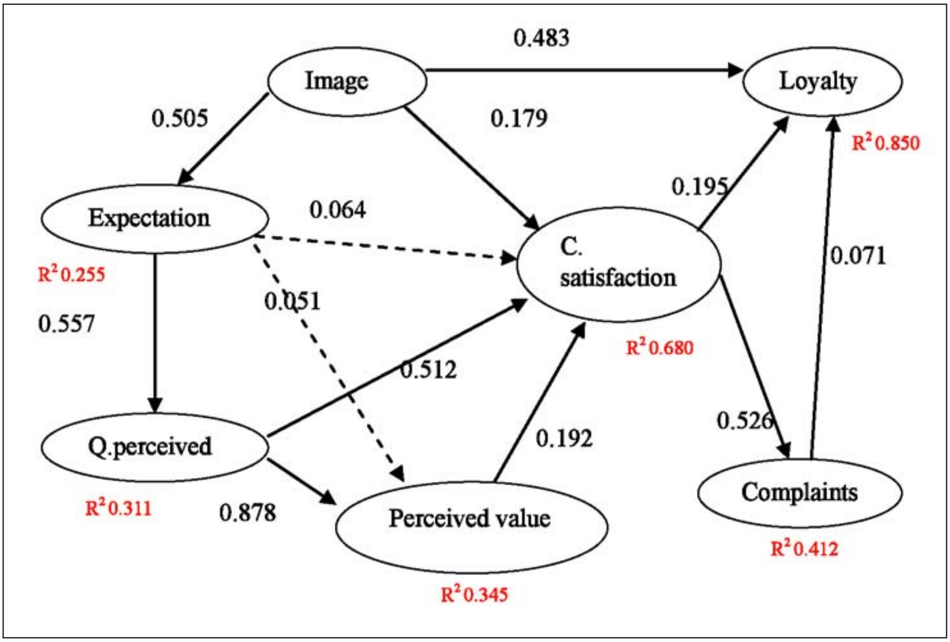

Fig. 4: The model with PLS-PM (XLSTAT).

In the table 1 there are the R^2 of the inner regressions and the Average Communality (this index is the mean of the correlations between the manifest and latent variables) for each block. In the last row we have reported the value of Goodness of Fit index (GoF is the index used to evaluate the global fit of the model to the data) for the two models. Respect to the values of R^2, the PALSOS-PM algorithm improves significantly them, because the quantification step; we have the same result for the Average Communality, because the correlations are optimized by Morals.

Tab. 1: The results of validation indexes with the PALSOS-PM.

Latent variable	R2	A. communality	A. redundancy	Gof index
Image		0,943		
Expectation	0,948	0,957	0,902	
P.quality	0,915	0,960	0,886	
P.value	0,949	0,984	0,935	
C.Satisfaction	0,980	0,975	0,956	
Complaints	0,954	1,000	0,934	
Loyalty	0,951	0,889	0,833	
				0,951

We can see as the variables are strictly correlated and the latent variables are the best obtainable from a given set of manifest variables (the maximization of the correlation coefficients). In particular, for the Average Communality, we have a significative improvement for the latent block *Image*, for which the value passes from 0.48 to 0.94 and for the latent block *Loyalty* (0.52 vs 0.88).

For these latent variables the quantification has produced a significative improvement in their definition. Concerning the value of the redundancy index, its values for each block, except for the block *C. Satisfaction*, are higher than the one of the model estimated with PLS-PM: the manifest variables and the exogenous latent variables are able to explain more variability of the manifest variables of endogenous latent blocks.

This improvement is reflected also in the computation of the GoF index, that depends from the R^2 and Communality. As a consequence of these two results the value of GoF is higher than of PLS-PM (0.95 against of 0.61).

Tab. 2: The results of validation indexes with the PLS-PM (XLSTAT).

Latent variable	R^2	A. communality	A. redundancy	Gof index
Image		0,478		
Expectation	0,255	0,480	0,122	
P.quality	0,311	0,577	0,179	
P.value	0,345	0,849	0,292	
C.Satisfaction	0,680	0,693	0,472	
Complaints	0,277	1,000	0,277	
Loyalty	0,457	0,517	0,238	
				0,471

4. CONCLUSION AND FUTURE PERSPECTIVES

In the previous sections we have discussed briefly the characteristics of the ALS algorithms, and in particular of Morals, that we have used to quantify the ordinal variables used in a SEM model. The problem of quantification is due to the strong assumption of continuity of these variables made in the PLS-PM algorithm: sometimes the relationships between the variables are not significant because they have a small scale of values. Our proposal allows to introduce all kinds of variables in a SEM model, and to quantify them respect to their nature (nominal, ordinal, interval, ratio), and this step of quantification is iterate in the algorithm until the convergence. The advantage of our proposal is the possibility to use mixed variables, in the same latent block, because Morals estimates the regression coefficients in presence also of mixed variables, each of them is treat respect its nature. The future work is oriented to verify the influence of the scale of measure (what does happen with a scale of 3, 4 or 5 values?) on the estimation of the model across PALSOS-PM, and what is the impact of the number of categories of nominal variables on the estimation of the model.

BIBLIOGRAPHY

ANDERSEN, E.B. [1995]. Polytomous Rasch models and their estimation. *In* G. H. Fischer, & I. W. Molenaar (Eds.), *Rasch models: Foundations, recent developments, and applications*. New York: Springer-Verlag, 271-292.

BARLOW, R.E., BARTHOLOMEW, D.J., BREMER, J.M. & BRUNK, H.D. [1972]. *Statistical inference under order restrictions.* New York: John Wiley.

de Leeuw, J., Young, F. & Takane, Y. [1976]. Additive structure in qualitative data: an alternating least squares method with optimal scaling features. *Psychometrika*, **41**, 471-503

DE LEEUW J. [1977]. Correctness of Kruskal's algorithms for monotone regression with ties. *Psycometrika*, **42**, **1**, 141-144.

DE LEEUW J. & van Rijckevorsel J.[1980]. HOMALS and PRINCALS some generalizations of Principal Components Analysis. *In* E. Diday, M. Jambu, L. Lebart, J.P. Pages, Y. Schektman, R. Tomassone (Eds.), *Data Analysis and Informatics*. Amsterdam (NL), 231-241.

DE LEEUW, J. [1984]. The GIFI system on nonlinear multivariate analysis. In Data Analysis and Informatics. . *In* E. Diday, M. Jambu, L. Lebart, J.P. Pages, Y.

DE LEEUW, J. [1984]. The GIFI system on nonlinear multivariate analysis. In Data Analysis and Informatics. . *In* E. Diday, M. Jambu, L. Lebart, J.P. Pages, Y. Schektman, R. Tomassone (Eds.), *Data Analysis and Informatics*. Amsterdam (NL), **3**, 415-424.

GIFI A. [1981]. *Nonlinear multivariate analysis*. Department of data Theory, University of Leiden, The Netherlands.

HWANG, H. & TAKANE, Y. [2004]. Generalized structured component analysis. *Psychometrika*, **69**, 81-99.

JAKOBOWICZ E. & C. DERQUENNE [2007]. A modified PLS Path Modeling algorithm handling reflective categorical variables and a new model building strategy. *Computational Statistics & Data Analysis* **51,8,** 3666-3678

KRUSKAL J.B. [1964]. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psycometrika*, 29, 1-27

KRUSKAL J.B. & SHEPARD R.N. [1974]. A nonmetric variety of linear factor analysis. *Psycometrika*, **39**, **2**, 123-157

LOVAGLIO P.G. [1997]. Un Algoritmo di regressione multipla con dati categoriali. *Quaderni di Statistica e Matematica Applicata alle Scienze Economico-Sociali*, Trento, **XIX**, **3**, 281-291.

LOVAGLIO P.G. [2002]. La stima di Variabili Latenti da variabili osservate miste. *Statistica*, Bologna, **LXII**, **2**, 203-213.

MUTHÉN B. [1984]. A general structural equation model with Dichotomous, Ordered categorical, and Continuous Latent Variable indicators. *Psychometrika*, **49**, **1**, 115-132.

SAPORTA G. [1990]. Simultaneus analysis of qualitative and quantitative data. *Proceedings of the 35th Scientific Meeting of the Italian Statistical Society*, 63-72.

STEVENS S.S. [1951]. Mathematics, Measurement and Psychophysics. *In* S.S. Stevens (Ed.) *Handbook of Experimental Psychology*. New York: John Wiley & Sons, 1-49.

TENENHAUS M. & YOUNG F.W. [1985]. An Analysis and Synthesis Of Multiple Correspondence Analysis Optimal Scaling, Dual Scaling, Homogenety Analysis and Other Methods for Quantifying Categorical Multivariate Data. *Psychometrika*, **50**,**1**, 91-119

YOUNG F.W. [1981]. Quantitative analysis of qualitative data. *Psychometrika*, **46, 4**, 357-388.

YOUNG F. W., TAKANE Y & DE LEEUW J. [1976]. Regression with qualitative and quantitative variables: an Alternating Least Square method with optimal scaling features. *Psycometrika*, **41**, **4**, 505-529.

WOLD H. [1982]. Soft modeling: The basic design and some extensions. *In* K. G. Joreskog, and H. Wold (Eds.), *Systems Under Indirect Observation*. Amsterdam (NH) 2, 1-54.

UN APPROCCIO ALTERNATING LEAST SQUARE PER LA STIMA DI UN SEM BASATO SU VARIABILI ORDINALI

Riassunto

In questo lavoro viene proposta una possibile soluzione al trattamento delle variabili ordinali nel contesto dei Modelli ad Equazioni Strutturali (MES), ed in particolare per il caso del PLS-PM.

Dopo aver evidenziato i limiti delle soluzioni proposte in letteratura, viene esplicitato la logica alla base dell'algoritmo PALSOS-PM da noi sviluppato in ambiente R. L'algoritmo inserendo come metodo di Optimal Scaling (O.S.) la metodologia Alternating Least Squares (Als), consente di ottenere nel PLS-PM, la migliore quantificazione per ogni tipo di variabile e la stima dei parametri del modello.

In particolare viene introdotto,per la stima esterna del modello PLS-PM, l'algoritmo Morals che consente di effettuare una regressione multipla con variabili dipendenti miste (quantitative, ordinali e/o nominali). L'algoritmo proposto è applicato ad un dataset *noto in letteratura, al fine di confrontare i risultati ottenibili con quelli derivanti dall'applicazione del modello classico PLS-PM.*