A nonlinear PLS path modeling based on monotonic B-spline transformations

Emmanuel Jakobowicz[1,2] and Gilbert Saporta[1]

[1]*CEDRIC, Conservatoire National des Arts et Métiers, 292 rue Saint Martin, 75141 Paris Cedex 03, France*

[2]*EDF R&D, 1 avenue du Général de Gaulle, 92121 Clamart Cedex, France*

INTRODUCTION

PLS path modeling is widely used in marketing applications. It is based on linear equations. However, in practical applications, many relations cannot be regarded as linear. For example, the relations between satisfaction and its attributes are nonlinear (Mittal et al., 1998). In this paper, we present a two step approach in order to include nonlinear relationships between manifest and latent variables.

We use a certain type of data transformation, often called optimal scaling, with monotonic B-spline optimal transformations of the manifest variables prior to PLS path modeling.

Combining optimal scaling methods and structural equation models is not new; LISREL users already perform these kinds of variables transformation using maximum likelihood estimation (Meijernick, 1995). De Leeuw (1988) advocated a two-step approach and Hwang and Takane (2002) presented an alternative approach to PLS path modeling which facilitates the use of optimal scaling but nothing has been made in the PLS path modeling framework. This can be explained by the lack of global optimization criterion. In this paper, we decided to stay in a general context with no global criterion. We focused on the outer model thinking in term of nonlinear principal components analysis (nonlinear PCA, Young et al., 1978). PLS path modeling is based on the estimation of latent variables as composite variables of their own manifest variables. They are created as linear combination in the space of their own manifest variables which makes it relevant to use nonlinear PCA. In the practical applications; we present a comparison between classical PLS path modeling and B-spline transformed PLS path modeling on customer satisfaction data.

A TWO-STEP APPROACH

We introduce prior to PLS path modeling algorithm, an optimal scaling technique, to obtain parameters of the transformations of the manifest variables. Let $\mathbf{X}_k$ be the manifest variables associated to latent variable ξ_k. We try to find a vector of function $\mathbf{B}_k$ with $\mathbf{B}_k\left(\mathbf{X}_k\right)=\tilde{\mathbf{X}}_k$ maximises the correlation with the *m* first principal components of the PCA on that block (Coolen et al., 1982). The loss function to be minimised for block *k* is:

$$tr\left(\left(\tilde{\mathbf{X}}_k-\mathbf{G}_k\mathbf{F}_k\right)'\left(\tilde{\mathbf{X}}_k-\mathbf{G}_k\mathbf{F}_k\right)\right),$$

over $\tilde{\mathbf{X}}_k$, $\mathbf{G}_k$ and $\mathbf{F}_k$ with $\tilde{\mathbf{X}}_k'\tilde{\mathbf{X}}_k=1$. Where $\mathbf{G}_k$ is a $m\times r$ matrix of *m* component' scores on *r* components, and $\mathbf{F}_k$ is a $n\times r$ matrix of *n* loadings on the *r* components. For identification purposes it is conventional to impose $\frac{\mathbf{G}_k'\mathbf{G}_k}{m}=1$ and $\mathbf{F}_k'\mathbf{F}_k=\mathbf{D}$ (diagonal).

In order to obtain the parameter estimation, we use alternating least squares estimation method based on Young (1981). First the parameters of the PCA are estimated with a fixed transformation of $\mathbf{X}_k$; second the parameters of the transformation $\mathbf{B}_k$ are estimated with fixed PCA parameters. This process is iterated until convergence. We, thus, maximise the correlation between the transformed variable and the first principal component which is usually very close to the latent variables score.

We focus on monotonic B-splines because of their interesting properties (De Boor, 1978). B-splines can be computed with simple recursive formulas, they form an orthonormal basis and they are attractive for the nonlinear analysis of continuous numerical variables. We use monotonic transformations to facilitate their interpretation in the PLS path modeling framework. Once the transformations estimated, transformed manifest are included in the model and PLS path modeling is applied.

APPLICATIONS

We present an application in the marketing domain, on customer satisfaction survey data. The structural equation model is defined in figure 1 and manifest variables are on scales between 1 and 10.

We obtain parameter estimations of the transformations by alternating least squares and nonlinear PCA. These data are included in the PLS path modeling treatment of the data. In order to obtain the knots and degrees of the B-splines, we conduct a search based on a hierarchical clustering algorithm (as in Durand, 2001) in order to maximize the *GoF* (Tenenhaus et al., 2005). It appears that the best transformations of the variables are obtained with quadratic splines with 2 knots.

We transform all the manifest variables of our model and present the results in table 1 comparing them to the classical PLS path modeling case.

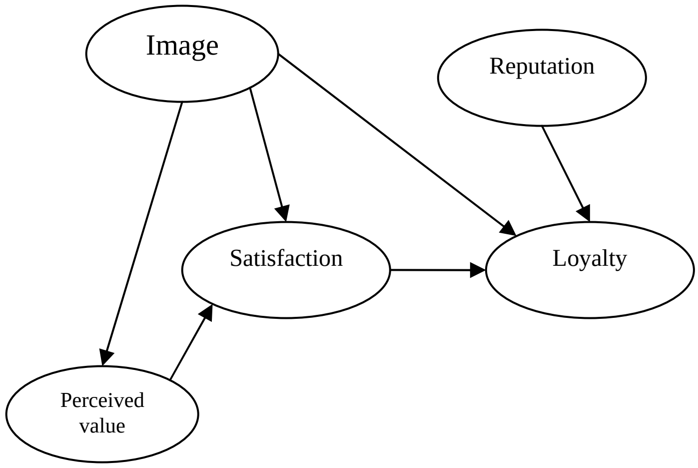

Figure 1: Structural model

Method	*GoF*	$\overline{R^2}$	$\overline{H^2}$	$\overline{F^2}$	Mean of Cronbach's α
B-spline PLS-PM	0.523	0.464	0.591	0.283	0.786
Classical PLS-PM	0.507	0.449	0.573	0.270	0.777

Table 1: Main results with B-spline transformed manifest variables

It appears that most indexes (see Tenenhaus et al. (2005) for a presentation) are improved when using transformed data. It should be pointed out that manifest variables associated to the perceived value and the reputation are not changed because relations appeared to be already linear. A simulation study should be applied in order to validate these results.

DISCUSSIONS

This procedure makes it possible to use transformed data in the PLS path modeling framework. Using monotonic transformation makes it possible to interpret structural coefficients of the model. Furthermore, the visualization of the transformation helps the researcher in finding the global behaviour of the customers on that variable.

This short presentation shows that including nonlinear transformation in structural equation models can be simple and can add quality of explanation to the model. However important shortcomings have to be pointed out, first using a nonlinear PCA does not include the structural model in the estimation of the transformation. It focuses even more on the outer model. Second, a global criterion should have been used in order to obtain a global optimal transformation. However, global criterions exist but depend on the choice of the outer and inner weights schemes. Our approach can be applied in all case to understand how manifest variables are connected to their own latent variables. Of course, many more research should be done in order to improve results and understand properties of this approach.

REFERENCES

Coolen H., Van Rijckevorsel, J. and De Leeuw, J. (1982) An algorithm for Nonlinear Principal Components Analysis with B-Splines by means of Alternating Least Squares. In Caussinus, H., Ettinger, P. and Mathieu, J.R. (Eds.), *COMPSTAT 1982, Part 2,* Physica-Verlag, Wien, 61-63.

De Boor, C. (1978) *A practical Guide to Splines*, Springer.

De Leeuw, J. (1988) Multivariate analysis with linearizable regressions. *Psychometrika*, 53(4):437-454.

Durand, J.-F. (2001) Local polynomial additive regression through PLS and splines: PLSS. Chemometrics and Intelligent Laboratory Systems, 58-235-246.

Hwang, H. and Takane, Y. (2002) Structural Equation Modeling by Extended Redundancy Analysis. In Nishisato, S., Baba, Y., Bodzogan, H. and Kanefuji, K. (Eds.), *Measurement and Multivariate Analysis*, Springer, 115-124.

Meijerink, F. (1995) *A Nonlinear Structural Relations Model*. DSWO Press.

Mittal, V., Ross, W., Baldasare, P. (1998) The asymmetric impact of negative and positive attribute-level performance on overall satisfaction and repurchase intentions. *Journal of Marketing*, 62:33–47.

Tenenhaus, M., Esposito Vinzi, V., Chatelin, Y.M. and Lauro, C. (2005) PLS path modeling. *Computational Statistics and Data Analysis*, 48(1), 159–205.

Young, F.W. (1981), Quantitative Analysis of Qualitative Data. *Psychometrika*, 46:357-388.

Young, F.W., Takane, Y. and De Leeuw, J. (1978) The Principal Components of Mixed Measurement Level Multivariate Data: an Alternating Least Squares Method with Optimal Scaling Features. *Psychometrika*, 45:279-281.