Latent variable transformation using monotonic B-splines in PLS Path Modeling

E. Jakobowicz [1] [2]

[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

Keywords: PLS path modeling - Nonlinear transformation - Marketing.

Abstract

PLS Path Modeling is a widely used approach in marketing. In that field, relationships between latent variables are frequently nonlinear. This nonlinearity is usually defined by a piecewise linear function. In this talk, we present an approach to include non linear transformations of the latent variables in the PLS path modeling algorithm with monotonic B-splines. We use optimal scaling methods to obtain an optimal transformation of the latent variables explaining a target latent variable. Applications on customer satisfaction data and on simulated data are presented.

Introduction

We introduce a modification of the PLS path modeling algorithm to include non linear transformations in the structural model. In many applications, relations between concepts cannot be stated to be linear. For example, in customer behaviour analysis, the relation between satisfaction and loyalty is not supposed to be linear. It is usually defined by a piecewise linear function. In order to include these kinds of nonlinearity in a structural equation model, we use optimal scaling theories. Latent variables are transformed using optimal monotonic B-splines transformations. These transformations are estimated using alternating least squares estimator for every iteration of the PLS path modeling algorithm. In the first section, we present the B-spline transformation using alternating least squares and introduce PLS path modeling. In the second section, we introduce the modified PLS path modeling algorithm. Applications on customer satisfaction and simulated data follow. We finish with some global remarks and further research interests.

1 Optimal scaling using monotonic B-splines and PLS path modeling

Optimal scaling techniques are widely used in social sciences; for a review, you can refer to De Leeuw (2005). Optimal scaling extends the ordinary general linear model by providing optimal variable transformations that are iteratively derived using the method of alternating least squares. The transformations of the variables are optimal in the sense that transformed variables have linear relationships with their dependant variable in term of linear regressions (De Leeuw, 1988). In this paper, we focus on that kind of transformations in the framework of PLS path modeling. We use an algorithm based on the maximization of the target latent variable squared multiple correlation (Young et al., 1976). To do so, an alternating least squares estimation is necessary (Young, 1981). This means that the algorithm proceeds in alternating steps, where in one step the loss function is minimized with respect to the weights of the linear problem with fixed parameters for the transformation of the data and in the other step the loss function is minimized with respect to the parameters of the transformation of the data with the weights of the linear problem fixed.

We focus on B-spline transformations because of their interesting properties and in particular on monotonic B-spline in order to be able to interpret the resulting structural coefficients. For a review on spline, you can refer to De Boor (1978). A spline can be defined as a piecewise polynomial. It is defined by its degree and its knots. A B-spline is defined on a whole interval $[a, b]$ and is not zero on a number of subinterval, this number being equal to $(degree) - 1$.

Combining optimal scaling with B-splines is not new and has been made by Coolen et al. (1981) for nonlinear principal component analysis. Monotonic B-splines are furthermore interesting because of their easy interpretation properties.

PLS path modeling was introduced in the early 80s by Wold (1982) and widely used in recent years. The basic approach is based on an iterative algorithm alternating outer model (or measurement model) estimation and inner model (or structural model) estimation. For a review, you can see Tenenhaus et al. (2005). PLS path modeling is a confirmatory method which has predictive properties. Indeed, it is not based on any distribution assumptions (this is why it is called soft modeling) and model fit cannot be estimated using a parametric χ^2 test. Instead of using goodness of fit indexes, predictive quality indexes like R^2, *communality* or *redundancy* are used. Finally, it has a major drawback; there is no general optimization function for the algorithm (specific cases exist depending on the inner and outer weights estimation schemes).
Nonlinear transformations of PLS path modeling are still marginal, however Wold (1982) proposed to use transformed manifest variables and Kramer (2005) introduced an algorithm based on the kernel trick. Using nonlinear relations in path models is an important point to better illustrate the "reality" of the relations.

2 B-spline transformed PLS path modeling

Including monotonic B-spline transformations in PLS path modeling requires a few basic hypothesis. First, a target latent variable has to be chosen in order to select explaining latent variables to be transformed. Second, a new step in the PLS algorithm has to be added. We focus on the case where all the explaining latent variables of the target variable are transformed; questions about interactions, convergence and effect on other latent variable scores will be treated later on.

Let $\mathbf{y}_K = \mathbf{X}_K\mathbf{w}_K$ be the outer estimation of the target latent variable $\boldsymbol{\xi}_K$. Let $\mathbf{y}_j = \mathbf{X}_j\mathbf{w}_j$ be the outer estimations of the latent variables $\boldsymbol{\xi}_j$ explaining the target latent variable $\boldsymbol{\xi}_K$ in the model, let

$$\mathbf{Y}_J = \begin{pmatrix} \mathbf{y}_1 \ \mathbf{y}_2 \ ... \ \mathbf{y}_P \end{pmatrix} \tag{1}$$

where $\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, ..., \boldsymbol{\xi}_P$ are explaining $\boldsymbol{\xi}_K$. We would like to find a monotonic B-spline transformation $B_j(.)$ of the $\mathbf{y}_j$s which maximises the squared multiple correlation of $\mathbf{y}_K$. Let $\tilde{\mathbf{Y}}_J = B_J(\mathbf{Y}_J)$, our goal is to minimize:

$$\lambda^2 = (\tilde{\mathbf{Y}}_J\boldsymbol{\alpha} - \mathbf{y}_K)'(\tilde{\mathbf{Y}}_J\boldsymbol{\alpha} - \mathbf{y}_K) \tag{2}$$

where $\boldsymbol{\alpha}$ is the multiple regression coefficient between the $\mathbf{y}_j$s and $\mathbf{y}_K$. We suppose that all $\mathbf{y}_i$ are standardised.

This estimation is done using an alternating least squares algorithm developed by Young et al. (1976). We focus on regression transformations because in the final step of the PLS path modeling algorithm, structural coefficients are estimated using multiple OLS regressions.

We now present the basic steps of the modified PLS path modeling algorithm:

1. Initialisation of the outer weights $\mathbf{w}_i$;

2. Outer estimation of all the latent variables $\boldsymbol{\xi}_i$:

$$\mathbf{y}_i = \mathbf{X}_i\mathbf{w}_i. \tag{3}$$

3. Optimal transformation of the latent variable outer estimation:

   - For all latent variables $\boldsymbol{\xi}_j$ connected to the target latent variable $\boldsymbol{\xi}_K$:

     (a) Estimation of the monotonic B-spline optimal transformation of the explaining latent variable using $\mathbf{y}_j$ and $\mathbf{y}_K$ by alternating least squares to maximise

     $$R^2 = \mathbf{y}'_K \tilde{\mathbf{Y}}_J (\tilde{\mathbf{Y}}'_J \tilde{\mathbf{Y}}_J)^{-1} \tilde{\mathbf{Y}}'_J \mathbf{y}_K \tag{4}$$

     (b) Transformation of latent variable outer estimation using B-spline $B_j(.)$

     $$\tilde{\mathbf{y}}_j = B_j(\mathbf{y}_j). \tag{5}$$

   - For latent variables not connected to $\boldsymbol{\xi}_K$ and for $\boldsymbol{\xi}_K$, the outer estimation is unchanged

     $$\tilde{\mathbf{y}}_j = \mathbf{y}_j. \tag{6}$$

4. Inner weight estimation using any estimation scheme, for example the centroid scheme:

   $$e_{ij} = sgn(\tilde{\mathbf{y}}'_i \tilde{\mathbf{y}}_j). \tag{7}$$

5. Inner estimation of the latent variables using transformed outer estimation of the variables:

   $$\mathbf{z}_j = \sum_{m:\boldsymbol{\xi}_m \leftrightarrow \boldsymbol{\xi}_l} \tilde{\mathbf{y}}_m. \tag{8}$$

6. Update of the outer weights using mode A:

   $$\mathbf{w}_j = \frac{1}{\mathbf{z}'_j \mathbf{z}_j} \mathbf{X}'_j \mathbf{z}_j. \tag{9}$$

7. Iterate (2) to (6) until convergence.

8. Estimate outer model coefficients by OLS regression between $\mathbf{x}_{ij}$ and $\mathbf{y}_i$, estimate inner model coefficients by OLS regression between $\tilde{\mathbf{y}}_i$s.

A few observations can be drawn from this algorithm. The new step adds complexity to the algorithm. To fully understand the estimated coefficients, a visualisation of the transformation is necessary. However as long as we use monotonic transformation, we can make relevant remarks on the coefficients without any visualisation. It is in the same time an advantage and a drawback of this approach. Convergence of the algorithm is an important point, the ALS algorithm has well known convergence properties but the convergence of PLS path modeling for more than two blocks has not been proved but is observed in most cases. Practical applications have shown that this modified algorithm converges when PLS path modeling converges. Because of the complexity of the algorithm, basic properties cannot be obtained analytically and we try to estimate these properties and the efficiency of this approach in an empirical framework using practical applications.

3 Applications

3.1 Customer satisfaction data

We present an application in the marketing domain, on customer satisfaction survey data from the French electricity market. The structural equation model is defined in figure 1 and manifest variables are on a scale between 1 and 10. Each latent variable has between 3 and 11 manifest variables. The sample size is 1988. We focus on a complex target latent variable: loyalty.

The transformed variables are image, satisfaction and reputation. Transformations are estimated in order to make the relation between loyalty and its explaining latent variables linear

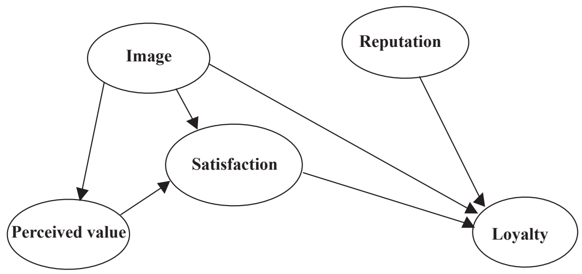

FIGURE 1. Structural model

and to maximise loyalty's score estimation squared multiple correlation coefficients. Using Durand (2001) method, we obtain optimal values of the number of knots and of the degrees of the B-spline function. We use a monotonic B-spline with 2 knots and 3 degrees. Perceived value is unchanged.

We apply path analysis using first the classical PLS path modeling and second the nonlinear PLS path modeling. Transformations of the latent variables image, satisfaction and reputation appear in figure 2. We summarize in table 1 the major results for the two approaches (focusing on loyalty's predictive quality).

Method	Nonlinear PLS-PM	PLS-PM
$R^2_{loyalty}$	0.410	0.406
$H^2_{loyalty}$	0.590	0.590
$F^2_{loyalty}$	0.242	0.239
Struc. coef. (Ima./Sat./Reput.)	0.419/0.208/0.118	0.411/0.205/0.106
GoF	0.505	0.507

TABLE 1. Comparison between nonlinear PLS-PM and classical PLS-PM on real data

Results are slightly better with the new approach. However these differences are quiet small even with important transformations (see figure 2). These figures show that the relations between loyalty and its explaining latent variables are not linear in this case. It highlights the weak effect of the inner model on the global estimation procedure.

In order to understand the basic properties of this approach, simulated data should be used.

3.2 Simulated data

Let simulate data according to a specific structural equation nonlinear model based on the ECSI model with 3 manifest variables for each latent variable (figure 3). The relations between the two endogenous latent variables is non linear following a piecewise quadratic transformation with 2 knots.
First the exogenous latent variable is simulated using a Beta(4,4) distribution between 1 and 10 and all error terms are simulated using a Beta(3,3) distribution between -1.5 and 1.5, the other variables are computed using the structural equations.

Results are summarised in figure 4 and table 2. They show that the identification of the transformation is well estimated but smoothed. However final results are close to the results with the classical PLS approach.

These results confirm older ones, it appears that nonlinear transformations are well identified by our approach but global and local predictive quality indexes are only slightly modified when introducing the transformation in the PLS algorithm. The major reasons of this weak effect are:

- Transformations highlighted using monotonic B-splines are rarely very different from the linear case. They stress out local transformations that can be interpreted. It is logical that

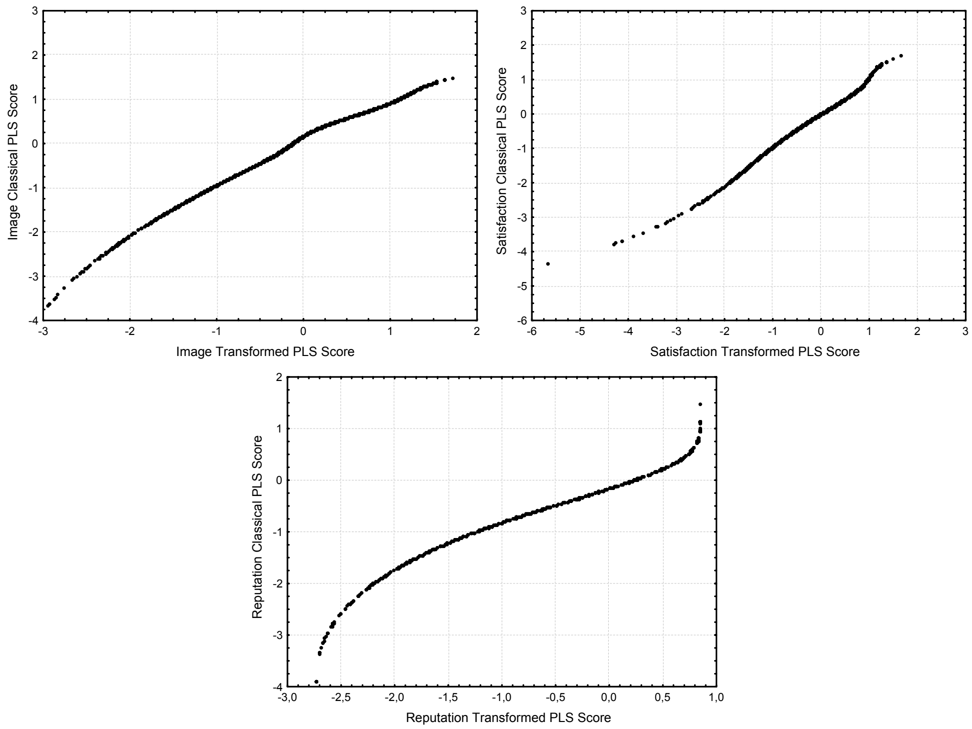

FIGURE 2. Latent variable image, satisfaction and reputation transformations

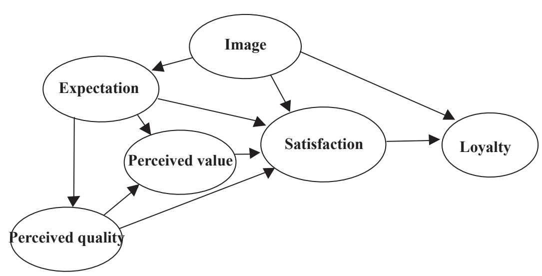

FIGURE 3. Simulated structural model

Method	Nonlinear PLS-PM	PLS-PM
$R^2_{loyalty}$	0.940	0.937
$H^2_{loyalty}$	0.958	0.958
$F^2_{loyalty}$	0.901	0.898
Struct. coef. (Image/Sat.)	0.399/0.582	0.392/0.587
GoF	0.945	0.946

TABLE 2. Comparison between nonlinear PLS-PM and classical PLS-PM on simulated data

these transformations does not strongly modify the outer score estimates. Using more complex transformations could be of interest; however researcher should be cautious about the final interpretation of the global model. If non monotone functions are used, conclusions on the impact of the studied latent variable will not be possible in the structural equation model.

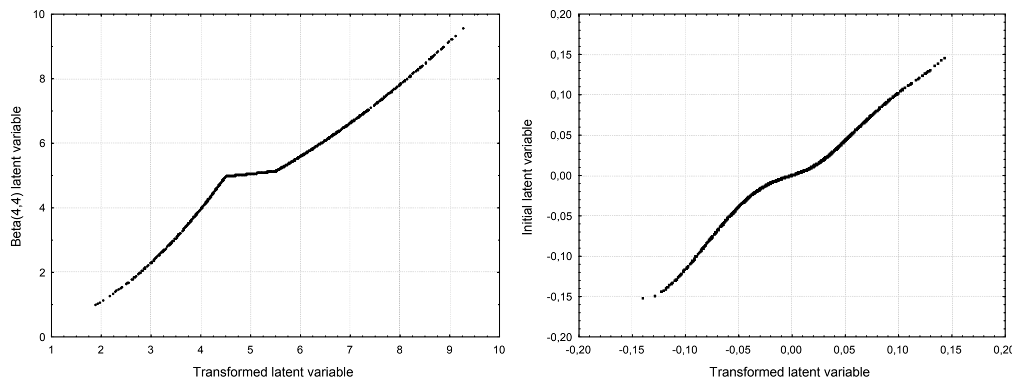

FIGURE 4. Latent variable satisfaction simulated transformation and transformation estimation

- PLS path modeling favours the outer model when estimating the latent variables. To prove this assumption, we run a principal component analysis (PCA) on each block of manifest variables and compare the first principal component to the latent variable score. Figure 5 illustrates how close the estimated are with both real and simulated data. We can see that both with real and simulated data, first principal component and PLS score are very close. This is due to the fact that PLS scores are obtained as linear combination of the latent variable associated manifest variables. It is computed in the space of its manifest variables and is thus highly related to them.

A major drawback of PLS path modeling is highlighted, the outer model explains most of the latent variable score and changes in the inner model do not greatly influence the latent variables estimates. However this conclusion should be tempered by the first reason and further research is necessary.

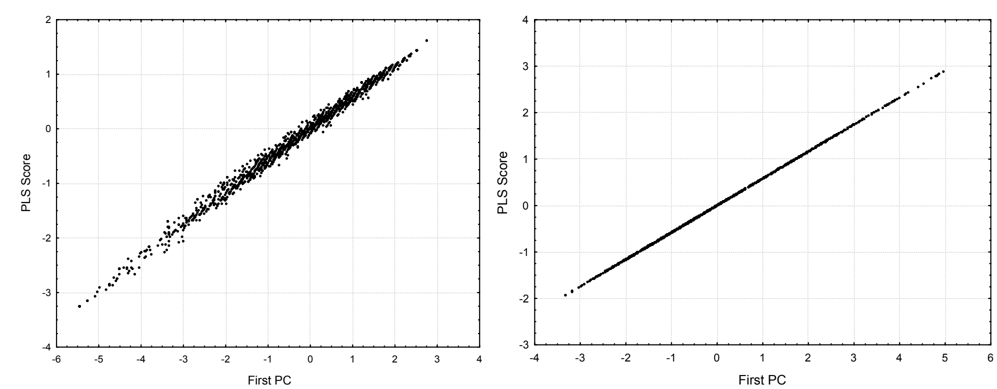

FIGURE 5. Comparison of PLS score and first principal component for real and simulated data

4 Concluding remarks

We introduced a new way to include nonlinear relationships between construct in the framework of PLS path modeling. Using optimal scaling techniques to linearize OLS regression allows to obtain both a visualisation of the transformation and a simple interpretation.

Practical applications have shown that our algorithm converges when PLS path modeling converges.
This approach is based on the improvement of a target latent variable predictive quality; a study on the effect of this change on the other latent variables of the model should be performed. These interactions could not be estimated analytically, a Monte Carlo simulation process should be applied. We would not present it in this paper and it should be included in further research. It appears in basic applications on simple real life and simulated data that this modification only slightly affects the global quality of the model and has not a significant effect on the latent variables estimates. We have shown how in our cases PLS estimates are directly related to the first principal component of the PCA on the studied block of variables.

This study has not only a research interest but also a practical interest. In the marketing domain, it is important to understand behaviour of concepts; the visualisation of the transformation helps the practitioner to understand how relations between global concepts behave. In that field, data is widely nonlinear and our approach apprehends well the transformations and favours the predictive quality of the target latent variable.

Finally more research is needed in order to transform the data following a global optimization criterion like in the case of PLS mode B. The transformation of the outer model should also be studied.

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 *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.

De Leeuw, J. (2005). Multivariate Analysis with Optimal Scaling. *Department of Statistics, UCLA. Department of Statistics Papers.*

Durand, J.-F. (2001). Local Polynomial Additive Regression Through PLS and Splines: PLSS. *Chemometrics and Intelligent Laboratory Systems*, **58** 235–246.

Krämer, N. (2005). Nonlinear Partial Least Squares Path Models. In *3rd World conference of the IASC, Cyprus.*

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

Wold, H. (1982) Soft Modeling: The Basic Design and Some Extensions. In *System under indirect observation, vol. 2*, North-Holland, Amsterdam, 1–54.

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

Young, F.W., De Leeuw, J. and Takane, Y. (1976) Regression with Qualitative and Quantitative Variables: An Alternating Least Squares Method with Optimal Scaling Features. *Psychometrika*, **41**(4) 505–529.