Available online at www.sciencedirect.com
ScienceDirect
Computational Statistics & Data Analysis 51 (2007) 3666–3678
COMPUTATIONAL STATISTICS & DATA ANALYSIS
www.elsevier.com/locate/csda

A modified PLS path modeling algorithm handling reflective categorical variables and a new model building strategy

Emmanuel Jakobowicz[a, b, *], Christian Derquenne[a]

[a] *EDF R&D, 1 Avenue du Général de Gaulle, 92121 Clamart Cedex, France*
[b] *CEDRIC-CNAM, 292 rue Saint Martin, 75141 Paris Cedex 03, France*

Received 29 March 2006; received in revised form 5 December 2006; accepted 6 December 2006
Available online 3 January 2007

Abstract

Partial least squares (PLS) path modeling has found increased applications in customer satisfaction analysis thanks to its ability to handle complex models. A modified PLS path modeling algorithm together with a model building strategy are introduced and applied to customer satisfaction analysis at the French energy supplier *Electricité de France*. The modified PLS algorithm handles all kinds of scales (categorical or nominal variables) and is well suited when nominal or binary variables are involved. PLS path modeling and structural equation modeling are confirmatory approaches and thus need an initial conceptual model. A two-step model building strategy is presented; the first step is based on Bayesian networks structure learning to build the measurement model and the second step is based on partial correlation and hypothesis tests to build the structural model. Applications to customer satisfaction data are presented.

Keywords: PLS path modeling; Partial least squares; Structural equation models; Model building strategy; Marketing application

1. Introduction

Partial least squares (PLS) path modeling is becoming a major tool in customer satisfaction studies. When dealing with structural equation models, two important points should be analyzed in more depth. When building questionnaires for customer satisfaction studies many variables may be of interest but have to be abandoned because they are not continuous or Likert scale like variables. Another issue is the model building strategy; when many variables are involved, expert knowledge can be too weak to build a robust and well suited model. Furthermore when variables are of mixed type the usual approach can lead to poor results. We thus present: First, an adapted version of PLS path modeling for different types of data (ordinal or nominal) called partial maximum likelihood (PML) (note that this approach is not related to Cox, 1975's partial likelihood), it is especially advantageous in the case of nominal or binary variables (Derquenne, 2005, 2006; Jakobowicz et al., 2005). Then, a two-step model building strategy is introduced. The first step uses discrete Bayesian networks; it is very useful when using categorical variables and many constraints can be

* Corresponding author. EDF R&D, 1 Avenue du Général de Gaulle, 92121 Clamart Cedex, France. Tel.: +33 0 660272156; fax: +33 0 147653133.
E-mail address: emmanuel.jakobowicz@free.fr (E. Jakobowicz).

0167-9473/$ - see front matter
doi:10.1016/j.csda.2006.12.004

added to the measurement model obtained. The second step is based on statistical testing to build the structural model using the first component of the principal components analysis on each block of the model (Derquenne and Hallais, 2004; Jakobowicz et al., 2005).

The article is organized as follows: In Section 2, a basic introduction to the PLS path modeling estimation method is provided. The modified PLS path modeling algorithm is presented in Section 3. In Section 4, the model building strategy is described. Applications to customer satisfaction surveys follow in Section 6. Remarks on the contributions of these approaches to structural equation modeling with latent variables and customer satisfaction analysis, together with future researches conclude this paper.

2. PLS path modeling

For an extensive review on PLS path modeling, the reader is referred to Tenenhaus et al. (2005). We focus on two points: the treatment of non-continuous data and model building strategies.

However, we first introduce some basic concepts on model specification and structural equation modeling in customer satisfaction analysis. In customer satisfaction analysis, concepts are complex and each notion is defined by many facets. In our models, answers to customer satisfaction survey questions will be included as manifest variables ($\mathbf{x}_{jh}$), which are grouped in blocks related to a concept like image, satisfaction or loyalty associated to a latent variable (ξ_j). The relations between manifest and latent variables form the outer or measurement model and relations between latent variables form the inner or structural model. The usual methods like LISREL (see Bollen, 1989 for a review) were unlikely to be adapted to complex marketing models, as shown by Tenenhaus et al. (2005). We decided to focus on PLS path modeling, which was introduced by Wold (1982) and adapted to customer satisfaction index analysis by Fornell (1992).

Throughout this article, we decided to focus on mode A for the outer estimation of the latent variables and the centroid scheme for the inner estimation. It is a deliberate choice and our approaches could be extended to the formative case and other inner schemes that are available in PLS path modeling, at least for the PML approach. Furthermore, we use mode A because we consider that in customer satisfaction analysis indicators are reflective (Johnson et al., 2001).

The PLS path modeling algorithm can be divided into three main steps:

(1) *A quantification step* where initial latent variables are calculated. This step is done using:

$$\mathbf{y}_j^t = \sum_{i=1}^{p_j} w_{ji}^t \mathbf{x}_{ji}, \tag{1}$$

where $\mathbf{y}_j^t$ is the outer estimate of the latent variable ξ_j at step t, $\mathbf{x}_{ji}$ is the manifest variable i associated with the latent variables ξ_j and p_j is the number of manifest variables in block j. During this step, the initial outer weight w_{ji}^0 can be chosen randomly. Usually it is set at 1 for the first manifest variable of each block and at 0 for the others (Lohmöller, 1989).

(2) *An iterative algorithm* where latent variables are estimated iteratively with respect to the inner model and the outer model.

- *Inner estimation*: Throughout this article we focus on the centroid scheme for inner estimate $\mathbf{z}_j^t$ of the latent variable ξ_j:

$$\mathbf{z}_j^t = \sum_{\xi_{j'} \in J} \operatorname{sign}(\operatorname{cor}(\mathbf{y}_j^t, \mathbf{y}_{j'}^t))\mathbf{y}_{j'}^t, \tag{2}$$

where J is the set of all latent variables connected to ξ_j.

- *Outer estimation*: Before performing the outer estimation $\mathbf{y}_j^t$ of latent variable ξ_j, outer weights are updated using mode A (reflective indicators) estimation:

$$w_{jh}^t = cov(\mathbf{x}_{jh}, \mathbf{z}_j^t). \tag{3}$$