Available online at www.sciencedirect.com
SCIENCE @ DIRECT®
Computational Statistics & Data Analysis 48 (2005) 159–205
COMPUTATIONAL STATISTICS & DATA ANALYSIS
www.elsevier.com/locate/csda

# PLS path modeling

Michel Tenenhaus[a], Vincenzo Esposito Vinzi[a,b,*], Yves-Marie Chatelin[c], Carlo Lauro[b]

[a] *HEC School of Management (GREGHEC), Jouy-en-Josas, France*
[b] *Department of Mathematics and Statistics, University of Naples "Federico II", Via Cintia—Complesso di Monte S. Angelo, 80126 Naples, Italy*
[c] *Institut de l'Elevage, Paris, France*

Accepted 10 March 2004

Available online 7 April 2004

**Abstract**

A presentation of the Partial Least Squares approach to Structural Equation Modeling (or PLS Path Modeling) is given together with a discussion of its extensions. This approach is compared with the estimation of Structural Equation Modeling by means of maximum likelihood (SEM-ML). Notwithstanding, this approach still shows some weaknesses. In this respect, some new improvements are proposed. Furthermore, PLS path modeling can be used for analyzing multiple tables so as to be related to more classical data analysis methods used in this field. Finally, a complete treatment of a real example is shown through the available software.


*Keywords:* Structural equation modeling; Partial least squares; PLS approach; Multiple table analysis

## 0. Introduction

There is some confusion in the terminology used in the PLS field. Trying to clarify it is a way to follow the evolution of the ideas in the PLS approach.

Herman Wold first formalized the idea of partial least squares in his paper about principal component analysis (Wold, 1966) where the NILES (=nonlinear iterative least squares) algorithm was introduced. This algorithm (and its extension to canonical correlation analysis and to specific situations with three or more blocks) was latter named

* Corresponding author. Dipartimento di Matematica e Statistica, Università degli Sutdi di Napoli "Federico II", Via Cintia—Complesso Monte Sant'Angelo, 80126 Napoli, Italy. Tel.: +39-081-675112; fax: +39-081-675113.

*E-mail address:* vincenzo.espositovinzi@unina.it (V.E. Vinzi).

0167-9473/$ - see front matter 
doi:10.1016/j.csda.2004.03.005

NIPALS (=nonlinear iterative partial least squares) in the 1973 and 1975 Wold's papers given in reference (Wold, 1973, 1975).

The first presentation of the finalized PLS approach to path models with latent variables (LVs) has been published by Wold, 1979 and then the main references on the PLS algorithm are Wold (1982, 1985).

Herman Wold opposed SEM-ML (Jöreskog, 1970) "hard modeling" (heavy distribution assumptions, several hundreds of cases necessary) to PLS "soft modeling" (very few distribution assumptions, few cases can suffice). These two approaches to structural equation modeling have been compared in Jöreskog and Wold (1982). In the following, these two approaches are compared on an example and it seems that, in fact, LVs estimates by both methods are very correlated if the SEM-ML LVs estimates are modified so that only the manifest variables (MVs) related to an LV are used to estimate the LV itself.

In the chemometrics field, PLS regression (Wold et al., 1983) has a tremendous success and in many publications there is a confusion between the father's (H. Wold) and the son's (S. Wold) work. The term "PLS approach" is somewhat too general and merge now PLS for path models and PLS regression. Following a suggestion by H. Martens, we have decided to name "PLS Path Modeling" the use of PLS for structural equation modeling.

The unique available software has been for many years LVPLS 1.8 developed by (Lohmöller (1987, last available version). Lohmöller has extended the basic PLS algorithm in various directions and published all his research results in 1989. More recently, a new software has been developed by Chin (2001, for the last version, still in beta test however): PLS-Graph 3.0. It contains a Windows user-friendly graphical interface to PLSX, a program for PLS path modeling on units by variables data table available in LVPLS 1.8. Moreover, it proposes a cross-validation of the path model parameters by jack-knife and bootstrap. Both software will be studied in details on a practical example.

A very important review paper on PLS approach to structural equation modeling is Chin (1998). A basic review on PLS path modeling for Marketing is Fornell and Cha (1994). Our paper is more theory oriented and should be seen as a complement to these paper for readers more oriented to Statistics. In particular, we give a detailed study of the relationship between PLS path modeling and multiple table analysis methods.

## 1. The PLS path modeling algorithm

To clarify the presentation of the PLS path modeling algorithm, it is very useful to refer to a practical example. PLS has been applied very extensively in customer satisfaction studies. So we will first present the construction of a customer satisfaction index (CSI).

### 1.1. The ECSI example

The European consumer satisfaction index (ECSI) is an economic indicator that measures customer satisfaction. A model has been derived specifically for the ECSI.

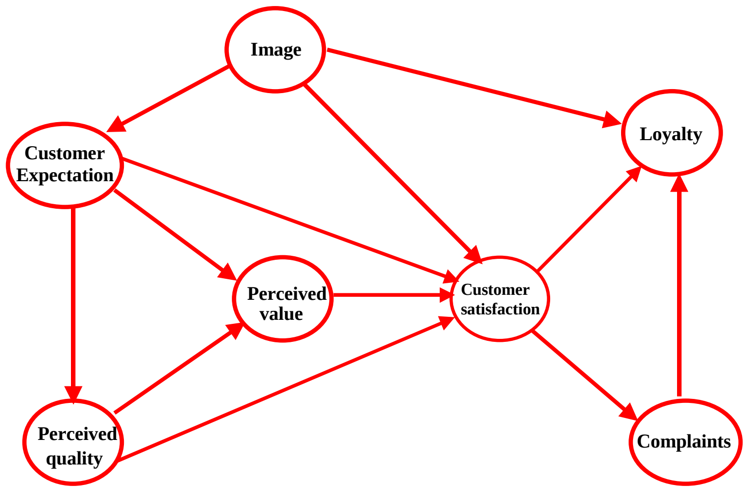


Fig. 1. Causality model describing causes and consequences of customer satisfaction.

In this model, seven interrelated LVs are introduced. It is based on well-established theories and approaches in customer behavior 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 ECSI model is described in Fig. 1. A set of MVs is associated with each of the LVs. This structure is called the ECSI model. The entire model is important for determining the main target variable, being CSI.

In Table 1 the MVs $\mathbf{v}_{jh}$ describing the LVs $\xi_j$ are given for the mobile phone industry. These MVs are normalized as follows: the original items $\mathbf{v}_{jh}$, scaled from 1 to 10, are transformed into new normalized variables $\mathbf{x}_{jh} = (\mathbf{v}_{jh} - 1) \times (100/9)$. The minimum possible value of $\mathbf{x}_{jh}$ is 0 and its maximum possible value is equal to 100. PLS software allows for a specific treatment of missing data that will be described later on in this paper.

### 1.2. The PLS path model

A PLS path model is described by two models: (1) a measurement model relating the MVs to their own LV and (2) a structural model relating some endogenous LVs to other LVs. The measurement model is also called the outer model and the structural model the inner model. Both models are described in this section.

#### 1.2.1. The measurement model

An LV $\xi$ is an unobservable variable (or construct) indirectly described by a block of observable variables $\mathbf{x}_h$ which are called MVs or indicators. There are three ways

Table 1
Measurement instrument for the mobile phone industry

| Latent variables | Manifest variables |
|---|---|
| Image ($\xi_1$) | (a) It can be trusted in what it says and does<br>(b) It is stable and firmly established<br>(c) It has a social contribution for the society<br>(d) It is concerned with customers<br>(e) It is innovative and forward looking |
| Customer expectations of the overall quality ($\xi_2$) | (a) Expectations for the overall quality of “your mobile phone provider” at the moment you became customer of this provider<br>(b) Expectations for “your mobile phone provider” to provide products and services to meet your personal need<br>(c) How often did you expect that things could go wrong at “your mobile phone provider” |
| Perceived quality ($\xi_3$) | (a) Overall perceived quality<br>(b) Technical quality of the network<br>(c) Customer service and personal advice offered<br>(d) Quality of the services you use<br>(e) Range of services and products offered<br>(f) Reliability and accuracy of the products and services provided<br>(g) Clarity and transparency of information provided |
| Perceived value ($\xi_4$) | (a) Given the quality of the products and services offered by “your mobile phone provider” how would you rate the fees and prices that you pay for them?<br>(b) Given the fees and prices that you pay for “your mobile phone provider” how would you rate the quality of the products and services offered by “your mobile phone provider”? |
| Customer satisfaction ($\xi_5$) | (a) Overall satisfaction<br>(b) Fulfillment of expectations<br>(c) How well do you think “your mobile phone provider” compares with your ideal mobile phone provider? |
| Customer complaints ($\xi_6$) | (a) You complained about “your mobile phone provider” last year. How well, or poorly, was your most recent complaint handled<br>or<br>(b) You did not complain about “your mobile phone provider” last year. Imagine you have to complain to “your mobile phone provider” because of a bad quality of service or product. To what extent do you think that “your mobile phone provider” will care about your complaint? |
| Customer loyalty ($\xi_7$) | (a) If you would need to choose a new mobile phone provider how likely is it that you would choose “your provider” again?<br>(b) Let us now suppose that other mobile phone providers decide to lower their fees and prices, but “your mobile phone provider” stays at the same level as today. At which level of difference (in %) would you choose another mobile phone provider?<br>(c) If a friend or colleague asks you for advice, how likely is it that you would recommend “your mobile phone provider”? |

All the items are scaled from 1 to 10. Scale 1 expresses a very negative point of view on the product while scale 10 a very positive opinion.

to relate the MVs to their LVs, respectively, called the reflective way, the formative one, and the MIMIC (multiple effect indicators for multiple causes) way.

In the ECSI example there are seven LVs $\xi_1, \ldots, \xi_7$.

*1.2.1.1. The reflective way.* In this model each MV reflects its LV.

Each MV is related to its LV by a simple regression:

$$\mathbf{x}_h = \pi_{h0} + \pi_h \xi + \boldsymbol{\varepsilon}_h, \tag{1}$$

where $\xi$ has mean $m$ and standard deviation 1. It is a reflective scheme: each MV $\mathbf{x}_h$ reflects its LV $\xi$. The only hypothesis made on model (1) is called by H. Wold the *predictor specification* condition

$$E(\mathbf{x}_h|\xi) = \pi_{h0} + \pi_h \xi. \tag{2}$$

This hypothesis implies that the residual $\boldsymbol{\varepsilon}_h$ has a zero mean and is uncorrelated with the LV $\xi$.

*Check for unidimensionality*: In the reflective way, the block of MVs is unidimensional in the meaning of factor analysis. In PLS path modeling a priori knowledge is incorporated in the algorithm. We suppose in this section that all the MVs of one block are positively correlated. It is not really a lack of generality. The MVs can always be built in this way, at least at the theoretical level.

On practical data this condition has to be checked. Three tools are available to check the unidimensionality of a block: use of principal component analysis of the block MVs, Cronbach's alpha and Dillon–Goldstein's $\rho$.

(a) *Principal component analysis of a block*: A block is essentially unidimensional if the first eigenvalue of the correlation matrix of the block MVs is larger than 1 and the second one smaller than 1, or at least very far from the first one. The first principal component can be built in such a way that it is positively correlated with all (or at least a majority of) the MVs. There is a problem with MV negatively correlated with the first principal component and we suggest that these MVs are inadequate to measure the LV and consequently be removed from the measurement model.

(b) *Cronbach's alpha*: Cronbach's alpha can be used to quantify unidimensionality of a block of $p$ variables $\mathbf{x}_h$ when they are all positively correlated. Cronbach has proposed the following procedure for *standardized* variables:

First, the variance of $\sum_{h=1}^{p} \mathbf{x}_h$ is developed as

$$\mathrm{Var}\left(\sum_{h=1}^{p} \mathbf{x}_h\right) = p + \sum_{h \neq h'} \mathrm{cor}(\mathbf{x}_h, \mathbf{x}_{h'}). \tag{3}$$

The larger the $\sum_{h \neq h'} \mathrm{Cor}(\mathbf{x}_h, \mathbf{x}_{h'})$, the more the block is unidimensional. So the following ratio is calculated:

$$\alpha' = \frac{\sum_{h \neq h'} \mathrm{cor}(\mathbf{x}_h, \mathbf{x}_{h'})}{p + \sum_{h \neq h'} \mathrm{cor}(\mathbf{x}_h, \mathbf{x}_{h'})}. \tag{4}$$

The maximum value of $\alpha'$ is equal to $(p-1)/p$ when all the $\mathrm{cor}(\mathbf{x}_h, \mathbf{x}_{h'})$ are equal to 1.

The Cronbach's alpha is then obtained by dividing $\alpha'$ by its maximum value:

$$\alpha = \frac{\sum_{h \neq h'} \operatorname{cor}(\mathbf{x}_h, \mathbf{x}_{h'})}{p + \sum_{h \neq h'} \operatorname{cor}(\mathbf{x}_h, \mathbf{x}_{h'})} \times \frac{p}{p-1}. \tag{5}$$

The Cronbach's alpha is also defined for original variables as

$$\alpha = \frac{\sum_{h \neq h'} \operatorname{cov}(\mathbf{x}_h, \mathbf{x}_{h'})}{\operatorname{var}(\sum_h \mathbf{x}_h)} \times \frac{p}{p-1}. \tag{6}$$

A block is considered as unidimensional when the Cronbach's alpha is larger than 0.7.

(c) *Dillon–Goldstein's* $\rho$: The sign of the correlation between each MV $\mathbf{x}_h$ and its LV $\xi$ is known by construction of the item and is supposed here to be positive. In Eq. (1) this hypothesis means that all the loadings $\pi_h$ are positive. A block is unidimensional if all these loadings are large.

Let us develop the variance of $\sum_{h=1}^{p} \mathbf{x}_h$ by using Eq. (1) and supposing that the residual terms $\boldsymbol{\varepsilon}_h$ are independent:

$$\begin{aligned} \operatorname{Var}\left(\sum_{h=1}^{p} \mathbf{x}_h\right) &= \operatorname{Var}\left(\sum_{h=1}^{p} (\pi_{h0} + \pi_h \xi + \boldsymbol{\varepsilon}_h)\right) \\ &= \left(\sum_{h=1}^{p} \pi_h\right)^2 \operatorname{Var}(\xi) + \sum_{h=1}^{p} \operatorname{Var}(\boldsymbol{\varepsilon}_h). \end{aligned} \tag{7}$$

A block is as much unidimensional as $(\sum_{h=1}^{p_i} \pi_h)^2$ is large. The Goldstein–Dillon's $\rho$ is then defined by

$$\rho = \frac{(\sum_{h=1}^{p} \pi_h)^2 \operatorname{Var}(\xi)}{(\sum_{h=1}^{p} \pi_h)^2 \operatorname{Var}(\xi) + \sum_{h=1}^{p} \operatorname{Var}(\boldsymbol{\varepsilon}_h)}. \tag{8}$$

Let us now suppose that all the MVs $\mathbf{x}_h$ and the LV $\xi$ are standardized. An approximation of the LV $\xi$ is obtained by standardization of the first principal component $\mathbf{t}_1$ of the block MVs. Then $\pi_h$ is estimated by $\operatorname{cor}(\mathbf{x}_h, \mathbf{t}_1)$ and, using Eq. (1), $\operatorname{Var}(\boldsymbol{\varepsilon}_h)$ is estimated by $1 - \operatorname{cor}^2(\mathbf{x}_h, \mathbf{t}_1)$. So we get an estimate of the Dillon–Goldstein's $\rho$:

$$\hat{\rho} = \frac{[\sum_{h=1}^{p} \operatorname{cor}(\mathbf{x}_h, \mathbf{t}_1)]^2}{[\sum_{h=1}^{p} \operatorname{cor}(\mathbf{x}_h, \mathbf{t}_1)]^2 + \sum_{h=1}^{p} [1 - \operatorname{cor}^2(\mathbf{x}_h, \mathbf{t}_1)]}. \tag{9}$$

A block is considered as unidimensional when the Dillon–Goldstein's $\hat{\rho}$ is larger than 0.7. This statistic is considered to be a better indicator of the unidimensionality of a block than the Cronbach's alpha (Chin, 1998, p. 320).

PLS path modeling is a mixture of a priori knowledge and data analysis. In the reflective way, the a priori knowledge concerns the unidimensionality of the block and the signs of the loadings. The data have to fit this model. If they do not, they can be modified by removing some MVs that are far from the model. Another solution is to change the model and use the formative way that will now be described.

*1.2.1.2. The formative way* In the formative way, it is supposed that the LV $\xi$ is generated by its own MVs. The LV $\xi$ is a linear function of its MVs plus a residual term

$$\boldsymbol{\xi} = \sum_h \varpi_h \mathbf{x}_h + \boldsymbol{\delta}. \tag{10}$$

In the formative model, the block of MVs can be multidimensional.

The predictor specification condition is supposed to hold for (10):

$$E(\boldsymbol{\xi}|\mathbf{x}_1, \ldots, \mathbf{x}_{p_j}) = \sum_h \varpi_h \mathbf{x}_h. \tag{11}$$

This hypothesis implies that the residual vector $\boldsymbol{\delta}$ has a zero mean and is uncorrelated with the MVs $\mathbf{x}_h$.

*Expected sign*: The MVs $\mathbf{x}_h$ are observed variables which are summarized by the LV $\boldsymbol{\xi}$. This summary makes more sense if the sign of each weight $\varpi_h$ is the one of the correlation between $\mathbf{x}_h$ and $\boldsymbol{\xi}$. There are no sign constraints on weights and loadings in the classical PLS algorithm, but unexpected signs of the weights and/or the loadings show problems in the data and some action must be taken. For example, MVs related with estimated parameters with wrong signs can be removed from the data. On another hand, as it will be shown later in the paper, sign constraints can be easily added to the PLS algorithm.

*1.2.1.3. The MIMIC way* The MIMIC way is a mixture of the reflective and formative ways.

The measurement model for a block is the following:

$$\mathbf{x}_h = \pi_{h0} + \pi_h \boldsymbol{\xi} + \boldsymbol{\varepsilon}_h \quad \text{for } h = 1 \text{ to } p_1, \tag{12}$$

where the LV is defined by

$$\boldsymbol{\xi} = \sum_{h=p_1+1}^{p} \varpi_h \mathbf{x}_h + \boldsymbol{\delta}. \tag{13}$$

The $p_1$ first MVs follow a reflective way and the $(p - p_1)$ last ones a formative way. The predictor specification hypotheses hold for (12) and (13) and lead to the same consequences as before on the residuals.

*1.2.1.4. Normalization of the LVs* The normalization of the LVs chosen by Wold (1985)—$\boldsymbol{\xi}$ has a standard deviation equal to 1—has been adopted by Lohmöller in his software. This normalization is arbitrary. Fornell (1992) has proposed another normalization very useful in customer satisfaction studies, but both Wold and Fornell LVs are co-linear. The Fornell's methodology is described in details in Bayol et al. (2000).

### *1.2.2. The structural model*

The causality model described in Fig. 1 leads to linear equations relating the LVs between them (the structural or inner model):

$$\boldsymbol{\xi}_j = \beta_{j0} + \sum_i \beta_{ji} \boldsymbol{\xi}_i + \mathbf{v}_j. \tag{14}$$

The predictor specification hypothesis is applied to (14).

An LV, which never appears as a dependent variable, is called an exogenous variable. Otherwise, it is called an endogenous variable.

The causality model must be a causal chain. That means that there is no loop in the causality model. This kind of model is called recursive, from the Latin *Recursio*, which means *I can return*.

Let us write the six structural equations corresponding to Fig. 1:

(a) customer expectation $= \beta_{20} + \beta_{21}$image $+ \mathbf{v}_2$,

(b) perceived quality $= \beta_{30} + \beta_{32}$customer expectation $+ \mathbf{v}_3$,

(c) perceived value $= \beta_{40} + \beta_{42}$customer expectation $+ \beta_{43}$perceived quality $+ \mathbf{v}_4$,

(d) CSI $= \beta_{50} + \beta_{51}$image $+ \beta_{52}$customer expectation,
$+\beta_{53}$perceived quality $+ \beta_{54}$perceived value $+ \mathbf{v}_5$,

(e) customer complaint $= \beta_{60} + \beta_{65}$CSI $+ \mathbf{v}_6$,

(f) customer loyalty $= \beta_{70} + \beta_{71}$image $+ \beta_{75}$CSI $+ \beta_{76}$customer complaint $+ \mathbf{v}_7$.

The usual hypotheses on the residuals (implied by the predictor specification condition) are made.

A structural model can be summarized by a 0/1 square matrix with dimensions equal to the number of LVs. Rows and columns represent the LVs. A cell $(i, j)$ is filled with a 1 if LV $j$ explains LV $i$, and 0 otherwise. Lohmöller calls this matrix the inner design matrix.

In the ECSI example the model matrix is written as

| | Image | Expectation | P. quality | P. value | Satisfaction | Complaint | Loyalty |
|---|---|---|---|---|---|---|---|
| Image | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Expectation | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| Per. quality | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| Per. value | 0 | 1 | 1 | 0 | 0 | 0 | 0 |
| Satisfaction | 1 | 1 | 1 | 1 | 0 | 0 | 0 |
| Complaint | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| Loyalty | 1 | 0 | 0 | 0 | 1 | 1 | 0 |

For a causal chain, the model matrix can always be written as a lower diagonal matrix with a diagonal of 0s. The Lohmöller's LVPLS software expects this type of model matrix.

### *1.3. The partial least-squares algorithm*

PLS path modeling has been mainly developed by Herman Wold (two main references are Wold, 1982, 1985), by Lohmöller (1987, 1989) for the computational aspects (the LVPLS software) and for theoretical developments, and by Chin (1998, 2001) and Chin and Newsted (1999) for a new software with graphical interface (PLS-Graph) and improved validation techniques. The PLS-Graph software is based on the Lohmöller's program PLSX for units×variables data and consequently allows for the same options.

We remind in this paper the various steps and various options of the original Wold's PLS algorithm. Then, we describe the new options available in Lohmöller's program PLSX and in Chin's software PLS-Graph.

#### 1.3.1. The Wold's algorithm

##### 1.3.1.1. LVs estimation

The LVs $\xi_j$ are estimated according to the following procedure.

###### 1.3.1.1.1. Outer estimate $\mathbf{y}_j$ of the standardized LV $(\xi_j - m_j)$

The standardized LVs (mean $=0$ and standard deviation $=1$) are estimated as linear combinations of their centered MVs

$$\mathbf{y}_j \propto \pm \left[\sum w_{jh}(\mathbf{x}_{jh} - \bar{x}_{jh})\right], \tag{15}$$

where the symbol "$\propto$" means that the left variable represents the standardized right variable and the "$\pm$" sign shows the sign ambiguity. This ambiguity is solved by choosing the sign making $\mathbf{y}_j$ positively correlated to a majority of $\mathbf{x}_{jh}$.

The standardized LV is finally written as

$$\mathbf{y}_j = \sum \tilde{w}_{jh}(\mathbf{x}_{jh} - \bar{x}_{jh}). \tag{16}$$

The coefficients $w_{jh}$ and $\tilde{w}_{jh}$ are both called the outer weights.

The mean $m_j$ is estimated by

$$\hat{m}_j = \sum \tilde{w}_{jh}\bar{x}_{jh} \tag{17}$$

and the LV $\xi_j$ by

$$\hat{\xi}_j = \sum \tilde{w}_{jh}\mathbf{x}_{jh} = \mathbf{y}_j + \hat{m}_j. \tag{18}$$

When all MVs are observed on the same measurement scale, it is nice to express (Fornell, 1992) LVs estimates in the original scale as

$$\hat{\xi}_j^* = \frac{\sum \tilde{w}_{jh}\mathbf{x}_{jh}}{\sum \tilde{w}_{jh}}. \tag{19}$$

Eq. (19) is feasible when all outer weights are positive. Finally, most often in real applications, LVs estimates are required on a 0–100 scale so as to have a reference scale to compare individual scores. From Eq. (19), for the $i$th observed case, this is easily obtained by the following transformation:

$$\hat{\xi}_{ij}^{0-100} = 100 \times \frac{(\hat{\xi}_{ij}^* - x_{\min})}{(x_{\max} - x_{\min})}, \tag{20}$$

where $x_{\min}$ and $x_{\max}$ are, respectively, the minimum and the maximum value of the measurement scale common to all MVs.

*1.3.1.1.2. Inner estimate* $\mathbf{z}_j$ *of the standardized LV* $(\xi_j - m_j)$ The inner estimate $\mathbf{z}_j$ of the standardized LV $(\xi_j - m_j)$ is defined by

$$\mathbf{z}_j \propto \sum_{j' : \xi_{j'} \text{ is connected with } \xi_j} e_{jj'} \mathbf{y}_{j'}, \tag{21}$$

where the inner weights $e_{jj'}$ are equal to the signs of the correlations between $\mathbf{y}_j$ and the $\mathbf{y}_{j'}$'s connected with $\mathbf{y}_j$. Two LVs are connected if there exists a link between the two variables: an arrow goes from one variable to the other in the arrow diagram describing the causality model. This choice of inner weights is called the *centroid scheme*.

This choice shows a drawback in case the correlation is approximately zero as its sign may change for very small fluctuations. But it does not seem to be a problem in practical applications.

In the original algorithm, the inner estimate is the right term of (21) and there is no standardization. We prefer to standardize because it does not change anything for the final inner estimate of the LVs and it simplifies the writing of some equations.

*1.3.1.1.3. Estimation modes for the weights* $w_{jh}$ There are two ways to estimate the weights $w_{jh}$: modes A and B.

*Mode A*: In mode A, the weight $w_{jh}$ is the regression coefficient of $\mathbf{z}_j$ in the simple regression of $\mathbf{x}_{jh}$ on the inner estimate $\mathbf{z}_j$:

$$w_{jh} = \mathrm{cov}(\mathbf{x}_{jh}, \mathbf{z}_j) \tag{22}$$

as $\mathbf{z}_j$ is standardized.

*Mode B*: In mode B, the vector $\mathbf{w}_j$ of weights $w_{jh}$ is the regression coefficient vector in the multiple regression of $\mathbf{z}_j$ on the manifest centered variables $(\mathbf{x}_{jh} - \bar{x}_{jh})$ related to the same LV $\xi_j$:

$$\mathbf{w}_j = (\mathbf{X}_j'\mathbf{X}_j)^{-1}\mathbf{X}_j'\mathbf{z}_j, \tag{23}$$

where $\mathbf{X}_j$ is the matrix with columns defined by the centered MVs $\mathbf{x}_{jh} - \bar{x}_{jh}$ related to the $j$th LV $\xi_j$.

Mode A is appropriate for a block with a reflective measurement model and mode B for a formative one. Mode A is often used for an endogenous LV and mode B for an exogenous one. Modes A and B can be used simultaneously when the measurement model is the MIMIC one. Mode A is used for the reflective part of the model and mode B for the formative part.

In practical situations, mode B is not so easy to use because there is often strong multicollinearity inside each block. When this is the case, PLS regression may be used instead of OLS multiple regression. As a matter of fact, it may be noticed that mode A consists in taking the first component from a PLS regression, while mode B takes all PLS regression components (and thus coincides with OLS multiple regression). Therefore, running a PLS regression and retaining a certain number of significant components may be meant as a new intermediate mode between modes A and B.

*Sign constraints*: In case we wish to incorporate a priori knowledge in the model, sign constraints may be added to the model itself:

$$\mathrm{sign}(w_{jh}) = \mathrm{sign}(\mathrm{cor}(\mathbf{x}_{jh}, \xi_j)). \tag{24}$$

For mode A the sign constraint means that Eq. (22) is replaced by

$$\begin{aligned} w_{jh} &= \mathrm{cov}(\mathbf{x}_{jh}, \mathbf{z}_j) \quad \text{if } \mathrm{sign}[\mathrm{cov}(\mathbf{x}_{jh}, \mathbf{z}_j)] = \mathrm{sign}[\mathrm{cor}(\mathbf{x}_{jh}, \xi_j)], \\ &= 0 \quad \text{otherwise.} \end{aligned} \tag{25}$$

For mode B, the OLS regression of $\mathbf{z}_j$ on $\mathbf{X}_j$ is replaced by the regression of $\mathbf{z}_j$ on $\mathbf{X}_j$ subject to constraints (24). If some regression coefficients in the OLS regression do not respect the sign constraints (24), then the multiple regression with sign constraints on the regression coefficients will lead to some regression coefficients equal to zero and the other regression coefficients with a good sign. The optimal solution is an OLS regression on variables selected by the constrained regression (see Tenenhaus, 1988).

In fact, in both modes A and B, the MVs not selected by the constrained regressions are just removed from the model.

*1.3.1.1.4. The PLS algorithm for estimating the weights* The starting step of the PLS algorithm consists in beginning with an arbitrary vector of weights $w_{jh}$. These weights are then standardized in order to obtain LVs with unitary variance.

A good choice for the initial weight values is to take $w_{jh} = \mathrm{sign}(\mathrm{cor}(\mathbf{x}_{jh}, \xi_h))$ or, more simply, $w_{jh} = \mathrm{sign}(\mathrm{cor}(\mathbf{x}_{jh}, \xi_h))$ for $h = 1$ and 0 otherwise.

Then steps (15), (21) and (22) or (23), depending on the selected mode, are iterated until convergence (guaranteed only for the two-block case, but practically always encountered in practice even with more than two blocks).

After the last step, calculations (16), (17) and (18) give final results for the inner weights $\tilde{w}_{jh}$, the standardized LV $\mathbf{y}_j = \sum \tilde{w}_{jh}(\mathbf{x}_{jh} - \bar{x}_{jh})$, the estimated mean $\hat{m}_j = \sum \tilde{w}_{jh}\bar{x}_{jh}$ of the LV $\xi_j$, and the final estimate $\hat{\xi}_j = \sum \tilde{w}_{jh}\mathbf{x}_{jh} = \mathbf{y}_j + \hat{m}_j$ of $\xi_j$. The latter estimate can be rescaled according to transformations (19) and (20).

The LV estimates are sensitive to the scaling of the MVs in mode A, but not in mode B. In this last case the outer LV estimate is the projection of the inner LV estimate on the *space* generated by its MVs.

*1.3.1.2. Estimation of the structural equations* The structural equations (14) are estimated by individual OLS multiple regressions where the LVs $\xi_j$ are replaced by their estimates $\hat{\xi}_j$. As usual, the use of OLS multiple regressions may be disturbed by the presence of strong multicollinearity between the estimated LVs. In such a case, PLS regression may be applied instead.

### *1.3.2. The new features of the Lohmöller's program PLSX*

In his software LVPLS 1.8, Lohmöller proposes new options for the PLS algorithm. These options are going to be described in this section.

*1.3.2.1. MVs standardization* PLSX proposes four options for the standardization of the MVs depending upon three conditions that eventually hold in the data:

*Condition 1*: The scales of the MVs are comparable. For instance, in the ECSI example the item values (between 0 and 100) are comparable. On the other hand, for instance, *weight* in tons and *speed* in km/h would not be comparable.

*Condition 2*: The means of the MVs are interpretable. For instance, if the difference between two MVs is not interpretable, the location parameters are meaningless.

*Condition 3*: The variances of the MVs reflect their importance.

If condition 1 does not hold, then the MVs have to be standardized (mean 0 and variance 1).

If condition 1 holds, it is useful to get the results based on the raw data. But the calculation of the model parameters depends upon the validity of the other conditions:

- *Condition* 2 *and* 3 *do not hold*: The MVs are standardized (mean 0 variance 1) for the parameter estimation phase. Then the MVs are rescaled to their original means and variances for the final expression of the weights and loadings.
- *Condition* 2 *holds*, *but not condition* 3: The MVs are not centered, but are standardized to unitary variance for the parameter estimation phase. Then the MVs are rescaled to their original variances for the final expression of the weights and loadings.
- *Conditions* 2 *and* 3 *hold*: Use the original variables.

Lohmöller introduced a standardization parameter (called METRIC) to select one of these four options:

| Variable scales are comparable | Means are interpretable | Variance is related to variable importance | Mean | Variance | Rescaling | METRIC |
|---|---|---|---|---|---|---|
| No | | | 0 | 1 | No | 1 |
| Yes | No | No | 0 | 1 | Yes | 2 |
| Yes | Yes | No | Original | 1 | Yes | 3 |
| Yes | Yes | Yes | Original | Original | | 4 |

In the ECSI model, PLS is applied to the raw MVs $\mathbf{x}_{jh}$. We use METRIC = 4.

*1.3.2.2. LVs estimation*

*1.3.2.2.1. Inner estimate* $\mathbf{z}_j$ *of the standardized LV* $(\xi_j - m_j)$ The inner estimate $\mathbf{z}_j$ of the standardized LV $(\xi_j - m_j)$ has been defined by Eq. (21). Herman Wold uses the so-called centroid scheme. Lohmöller has implemented two other schemes: the factorial scheme and the path weighting (or structural) scheme. These two new schemes are defined as follows.

*Factorial scheme*: The inner weights $e_{ji}$ are equal to the correlation between $\mathbf{y}_i$ and $\mathbf{y}_j$. This is an answer to the drawbacks of the centroid scheme described above.

*Path weighting scheme* (or *Structural scheme*): The LVs connected to $\xi_j$ are divided into two groups: the predecessors of $\xi_j$, which are LVs explaining $\xi_j$, and the followers, which are LVs explained by $\xi_j$.

For a predecessor $\xi_{j'}$ of the LV $\xi_j$, the inner weight $e_{jj'}$ is equal to the regression coefficient of $\mathbf{y}_{j'}$ in the multiple regression of $\mathbf{y}_j$ on all the $\mathbf{y}_{j'}$'s related to the predecessors of $\xi_j$. If $\xi_{j'}$ is a successor of $\xi_j$ then the inner weight $e_{jj'}$ is equal to the correlation between $\mathbf{y}_{j'}$ and $\mathbf{y}_j$.

These new schemes do not significantly influence the results but are very important for theoretical reasons. As it will be shown later in the paper, they allow to relate PLS path modeling to usual multiple table analysis methods.

*1.3.2.2.2. Estimation modes for the outer weights $w_{jh}$* Lohmöller has added a new mode C for the calculation of the outer weights.

*Mode C*: In mode C the weights are all equal in absolute value and reflect the signs of the correlations between the MVs and their LVs:

$$w_{jh} = \operatorname{sign}(\operatorname{cor}(\mathbf{x}_{jh}, \mathbf{z}_j)). \tag{26}$$

These weights are then normalized so that the resulting LV has unitary variance. Mode C actually refers to a formative way of linking MVs to their LVs and represents a specific case of mode B whose comprehension is very intuitive to practitioners.

*1.3.2.2.3. The starting values for estimating the outer weights* In PLSX the initial weights are obtained by assigning, per each block, 1 to all MVs but the last one, which is instead, assigned a $-1$. This choice is the main reason for eventual negative estimates of outer weights especially in the case when there are very few MVs in a block (two MVs being the worst case). Nevertheless it allows some sign control for the vector of outer weights related to an LV by acting on the order of the MV. If all weights are negative, the sign can usually be reversed by putting in last position the MV with the smallest weight in absolute value. If a negative sign is required for an MV, this variable should be located in last position.

In any case this choice of Lohmöller leads to sign problems in cross-validation and should not be the standard choice. Therefore, per each block, we propose to take the elements of the first eigenvector from a principal component analysis of the block with a majority of positive signs. In case of a tie between positive and negative signs, the variable with the strongest correlation in absolute value takes the positive sign and the remaining signs are assigned accordingly.

*1.3.2.3. Missing data treatment* PLSX provides a specific treatment for missing data:

1. When some cells are missing in the data, they must be coded with some value chosen by the user (for example 999.00). In the PLSX instructions the presence of missing data implies the number of subjects $N$ to be coded as "$-(N+1)$" and the first line contains the missing code for each MV (for example 999.00). Please notice that in the output of the program, the user's missing value code is always replaced by $-9.999$.
2. Means and standard deviations of the MVs are computed on all the available data.

3. All the MVs are centered.
4. If a unit has missing values on a whole block $j$, the value of the LV estimate $\mathbf{y}_j$ is missing for this unit.
5. If a unit $i$ has some missing values on a block $j$ (but not all), then the outer estimate $y_{ji}$ is defined by

$$y_{ji} = \sum_{jh : x_{jhi} \text{ exists}} \tilde{w}_{jh}(x_{jhi} - \bar{x}_{jh}).$$

   That means that each missing data of variable $\mathbf{x}_{jh}$ is replaced by the mean $\bar{x}_{jh}$.
6. If a unit $i$ has some missing values on its LVs, then the inner estimate $z_{ji}$ is defined by

$$z_{ji} = \sum_{\substack{k : \xi_k \text{ is connected with } \xi_j \\ \text{and } y_{ki} \text{ exists}}} e_{jk} y_{ki}.$$

   That means that each missing data of variable $\mathbf{y}_k$ is replaced by its mean 0.
7. The weights $w_{jh}$ are computed using all the available data on the basis of the following procedures:

   *For mode A*: The outer weight $w_{jh}$ is the regression coefficient of $\mathbf{z}_j$ in the regression of $(\mathbf{x}_{jh} - \bar{x}_{jh})$ on $\mathbf{z}_j$ calculated on the available data.

   *For mode B*: When there are no missing data, the outer weight vector $\mathbf{w}_j$ is equal to

$$\mathbf{w}_j = (\mathbf{X}_j \mathbf{X}_j')^{-1} \mathbf{X}_j' \mathbf{z}_j.$$

   The outer weight vector $\mathbf{w}_j$ is also equal to

$$\mathbf{w}_j = [\mathrm{Var}(\mathbf{X}_j)]^{-1} \mathrm{Cov}(\mathbf{X}_j, \mathbf{z}_j),$$

   where $\mathrm{Var}(\mathbf{X}_j)$ is the covariance matrix of $\mathbf{X}_j$ and $\mathrm{Cov}(\mathbf{X}_j, \mathbf{z}_j)$ the column vector of the covariances between the variables $\mathbf{x}_{jh}$ and $\mathbf{z}_j$.
   When there are missing data, each element of $\mathrm{Var}(\mathbf{X}_j)$ and $\mathrm{Cov}(\mathbf{X}_j, \mathbf{z}_j)$ is computed using all the pairwise available data and $\mathbf{w}_j$ is computed using the previous formula. This pairwise deletion procedure shows the drawback of possibly computing covariances on different sample sizes and/or different statistical units. However, in the case of few missing values, it seems to be very robust. This justifies why the blindfolding procedure, that will be presented in the next section, yields very small standard deviations for parameters.
8. The path coefficients are the regression coefficients in the multiple regressions relating some LVs to some others. When there are some missing values, the procedure described in point 7 (mode B) is also used to estimate path coefficients.

### *1.4. Model validation*

A path model can be validated at three levels: (1) the quality of the measurement model, (2) the quality of the structural model, and (3) each structural regression equation.

#### 1.4.1. Communality and redundancy

The communality index measures the quality of the measurement model for each block. It is defined, for block $j$, as

$$\text{communality}_j = \frac{1}{p_j}\sum_{h=1}^{p_j} \operatorname{cor}^2(\mathbf{x}_{jh}, \mathbf{y}_j). \tag{27}$$

The average communality is the average of all the $\operatorname{cor}^2(\mathbf{x}_{jh}, \mathbf{y}_j)$:

$$\overline{\text{communality}} = \frac{1}{p}\sum_{j=1}^{J} p_j \text{communality}_j, \tag{28}$$

where $p$ is total number of MVs in all blocks.

The redundancy index measures the quality of the structural model for each endogenous block, taking into account the measurement model. It is defined, for an endogenous block $j$, as

$$\text{redundancy}_j = \text{communality}_j \times R^2(\mathbf{y}_j, \{\mathbf{y}_{j'}\text{'s explaining } \mathbf{y}_j\}). \tag{29}$$

The average redundancy for all endogenous blocks can also be computed.

A global criterion of goodness-of-fit (GoF) can be proposed (Amato et al., 2004) as the geometric mean of the average communality and the average $R^2$:

$$\text{GoF} = \sqrt{\overline{\text{communality}} \times \overline{R^2}}. \tag{30}$$

As a matter of fact, differently from SEM-ML, PLS path modeling does not optimize any global scalar function so that it naturally lacks of an index that can provide the user with a global validation of the model (as it is instead the case with $\chi^2$ and related measures in SEM-ML). The GoF represents an operational solution to this problem as it may be meant as an index for validating the PLS model globally.

#### 1.4.2. Modified communality and redundancy for cross-validation

It is useful to slightly modify the communality index and express the redundancy index in terms of the LVs prediction so as to set the basis for validating their values by means of the cross-validation procedure shown in the next section.

The communality index for block $j$ can be written as

$$\text{communality}_j = \frac{1}{p_j}\sum_{h=1}^{p_j} \operatorname{cor}^2(\mathbf{x}_{jh}, \mathbf{y}_j) = \frac{1}{p_j}\sum_{h=1}^{p_j}\left(1 - \frac{\sum_{i=1}^{n}(x_{jhi} - \bar{x}_{jh} - \hat{\pi}_{jh} y_{ji})^2}{\sum_{i=1}^{n}(x_{jhi} - \bar{x}_{jh})^2}\right),$$

where $\hat{\pi}_{jh}$ is the regression coefficient of $\mathbf{x}_{jh}$ on $\mathbf{y}_j$. If we assume that the variances of the variables $\mathbf{x}_{jh}$ are close to each other, we get

$$\text{communality}_j \approx 1 - \frac{\sum_{h=1}^{p_j}\sum_{i=1}^{n}(x_{jhi} - \bar{x}_{jh} - \hat{\pi}_{jh} y_{ji})^2}{\sum_{h=1}^{p_j}\sum_{i=1}^{n}(x_{jhi} - \bar{x}_{jh})^2}. \tag{31}$$

In case the variances of the variables $\mathbf{x}_{jh}$ are quite different from each other, then we work on standardized variables.

The redundancy index for variable $\mathbf{x}_{jh}$ is defined as

$$\text{redundancy}_{jh} = \operatorname{cor}^2(\mathbf{x}_{jh}, \mathbf{y}_j) \times R^2(\mathbf{y}_j, \{\mathbf{y}_{j'}\text{'s explaining } \mathbf{y}_j\}),$$

that can be also expressed as

$$\begin{aligned}\text{redundancy}_{jh} &= \frac{\hat{\pi}_{jh}^2 \times \text{var}(\mathbf{y}_j)}{\text{var}(\mathbf{x}_{jh})} \times \frac{\text{var}[\text{Pred}(\mathbf{y}_j)]}{\text{var}(\mathbf{y}_j)} \\ &= \frac{\hat{\pi}_{jh}^2 \times \text{var}[\text{Pred}(\mathbf{y}_j)]}{\text{var}(\mathbf{x}_{jh})} \\ &= \frac{\text{var}[\hat{\pi}_{jh} \times \text{Pred}(\mathbf{y}_j)]}{\text{var}(\mathbf{x}_{jh})},\end{aligned}$$

where $\text{Pred}(\mathbf{y}_j)$ is the prediction of the LV $\mathbf{y}_j$ from the structural model, that is

$$\text{Pred}(\mathbf{y}_j) = \sum_{j' : \xi_{j'} \text{ explaining } \xi_j} \hat{\beta}_{j'} \mathbf{y}_{j'}.$$

If we assume that the regression coefficient of $\text{Pred}(\mathbf{y}_j)$ in the regression of $(\mathbf{x}_{jh} - \bar{x}_{jh})$ on $\text{Pred}(\mathbf{y}_j)$ is close to $\hat{\pi}_{jh}$, then we get

$$\text{redundancy}_{jh} \approx 1 - \frac{\sum_{i=1}^{n} [x_{jhi} - \bar{x}_{jh} - \hat{\pi}_{jh} \times \text{Pred}(y_{ji})]^2}{\sum_{i=1}^{n} (x_{jhi} - \bar{x}_{jh})^2} \tag{32}$$

and then, if we assume that the variances of the variables $\mathbf{x}_{jh}$ are close to each other:

$$\text{redundancy}_{j} \approx 1 - \frac{\sum_{h=1}^{p_j} \sum_{i=1}^{n} [x_{jhi} - \bar{x}_{jh} - \hat{\pi}_{jh} \times \text{Pred}(y_{ji})]^2}{\sum_{h=1}^{p_j} \sum_{i=1}^{n} (x_{jhi} - \bar{x}_{jh})^2}. \tag{33}$$

### *1.4.3. The blindfolding approach: cross-validated communality and redundancy*

The cv-communality (cv stands for cross-validated) index measures the quality of the measurement model for each block. It is a kind of cross-validated $R^2$ between the block MVs and their own LV calculated by a blindfolding procedure available in PLSX.

The quality of each structural equation is measured by the cv-redundancy index (i.e. Stone–Geisser's $Q^2$). It is a kind of cross-validated $R^2$ between the MVs of an endogenous LV and all the MVs associated with the LVs explaining the endogenous LV, using the estimated structural model.

Following Wold (1982, p. 30), the cross-validation test of Stone and Geisser fits soft modeling like hand in glove. In PLS path modeling statistics on each block and on each structural regression are available.

The significance levels of the regression coefficients can be computed using the usual Student's $t$ statistic or using cross-validation methods like jack-knife or bootstrap available in PLS-Graph.

In the Lohmöller program, the blindfolding approach proposed by Herman Wold is used and it is worth describing.

(1) The data matrix is divided into $G$ groups. The value $G = 7$ is recommended by Herman Wold. We give in the following table an example taken from the LVPLS

software documentation. The first group is related to letter a, the second one to letter b, and so on.

| $\mathbf{x}_1$ | $\mathbf{x}_2$ | $\mathbf{x}_3$ | $\mathbf{x}_4$ | $\mathbf{x}_5$ |
|---|---|---|---|---|
| a | f | d | b | g |
| b | g | e | c | a |
| c | a | f | d | b |
| d | b | g | e | c |
| e | c | a | f | d |
| f | d | b | g | e |
| g | e | c | a | f |
| a | f | d | b | g |
| b | g | e | c | a |
| c | a | f | d | b |
| d | b | g | e | c |
| e | c | a | f | d |

(2) Each group of cells is removed at its turn from the data. So a group of cells appears to be missing (for example all cells with letter a).
(3) A PLS model is run $G$ times by excluding each time one of the groups.
(4) One way to evaluate the quality of the model consists in measuring its capacity to predict MVs using other LVs. Two indices are used: communality and redundancy.
(5) In the communality option, we get prediction for the values of the centered MVs not included in the analysis, using the LV estimate, by the following formula:

$$\text{Pred}(x_{jhi} - \bar{x}_{jh}) = \hat{\pi}_{jh(-i)} y_{j(-i)},$$

where $\hat{\pi}_{jh(-i)}$ and $y_{j(-i)}$ are computed on data where the $i$th value of variable $\mathbf{x}_{jh}$ is missing.

In PLSX, the following terms are computed:

- Sum of squares of observations for one MV: $\text{SSO}_{jh} = \sum_i (x_{jhi} - \bar{x}_{jh})^2$.
- Sum of squared prediction errors for one MV: $\text{SSE}_{jh} = \sum_i (x_{jhi} - \bar{x}_{jh} - \hat{\pi}_{jh(-i)} y_{j(-i)})^2$.
- Sum of squares of observations for block $j$: $\text{SSO}_j = \sum_h \text{SSO}_{jh}$.
- Sum of squared prediction errors for block $j$: $\text{SSE}_j = \sum_h \text{SSE}_{jh}$.
- Cv-communality measure for block $j$: $H_j^2 = 1 - \text{SSE}_j/\text{SSO}_j$.

The index $H_j^2$ is the cross-validated communality index of the expression shown in (31). The mean of the cv-communality indices can be used to measure the global quality of the measurement model if they are positive for all blocks, taking into account the measurement model.

(6) In the redundancy option, we get a prediction for the values of the centered MVs not used in the analysis by using the following formula:

$$\text{Pred}(x_{jhi} - \bar{x}_{jh}) = \hat{\pi}_{jh(-i)} \text{Pred}(y_{j(-i)}),$$

where $\hat{\pi}_{jh(-i)}$ is the same as in the previous paragraph and $\mathrm{Pred}(y_{j(-i)})$ is the prediction for the $i$th observation of the *endogenous* LV $\mathbf{y}_j$ using the regression model computed on data where the $i$th value of variable $\mathbf{x}_{jh}$ is missing.

In PLSX, the following terms are also computed:

- Sum of squared prediction errors for one MV:
$$\mathrm{SSE}'_{jh} = \sum_i (x_{jhi} - \bar{x}_{jh} - \hat{\pi}_{jh(-i)}\mathrm{Pred}(y_{j(-i)}))^2.$$
- Sum of squared prediction errors for block $j$:
$$\mathrm{SSE}'_j = \sum_h \mathrm{SSE}'_{jh}.$$
- Cv-redundancy measure for an endogenous block $j$:
$$F_j^2 = 1 - \frac{\mathrm{SSE}'_j}{\mathrm{SSO}_j}.$$

The index $F_j^2$ is the cross-validated redundancy index of the expression shown in (33). The mean of the various cv-redundancy indices related to the endogenous blocks can be used to measure the global quality of the structural model if they are positive for all endogenous blocks, taking into account the measurement model.

(7) In PLSX, jack-knife, means and standard deviations of model parameters (weights, loadings, path coefficients, correlations between LVs) are computed by using the result of the $G$ blindfolding analyses. Means and standard deviations of a parameter are computed on the sample of the $G$ parameter estimates issued from the $G$ blindfolding analyses. However, this blindfolding procedure gives very small standard deviations thus leading systematically to significant parameters (see Section 1.5.5).

### *1.4.4. Resampling in PLS-Graph*

PLS-Graph really gives some added value with respect to PLSX in the way of assessing the significance of PLS parameters. As a matter of fact, besides the classical blindfolding procedure, PLS-Graph provides with jack-knife and bootstrap resampling options.

*Jack-knife*: The jack-knife procedure builds resamples by deleting a certain number of units from the original sample (with size $N$). The default consists in deleting 1 unit at a time so that each jack-knife sub-sample is made of $N - 1$ units. Increasing the number of deleted units leads to a potential loss in robustness of the $t$-statistic because of a smaller number of sub-samples. The complete statistical procedure used in PLS-Graph is described in Chin (1998, pp. 318–320).

*Bootstrap*: The bootstrap samples, instead, are built by resampling with replacement from the original sample. The procedure produces samples consisting of the same number of units as in the original sample. The number of resamples has to be specified. The default is 100 but a higher number (such as 200) may lead to more reasonable standard error estimates.

*Preprocessing options for both resampling procedures*: In PLS, LVs are defined up to the sign. It means that $\mathbf{y}_j = \sum \tilde{w}_{jh}(\mathbf{x}_{jh} - \bar{x}_{jh})$ and $-\mathbf{y}_j$ are both equivalent solutions. In order to remove this indeterminacy, Wold (1985) suggests retaining the solution where the correlations between the MVs $\mathbf{x}_{jh}$ and the LV $\mathbf{y}_j$ show a majority of positive signs. Unfortunately, PLSX does not consider this suggestion. Consequently, when estimating the PLS outer weights from the resamples, arbitrary sign changes may occur. This implies that also the loadings and the path coefficients estimated on the resamples may show arbitrary differences with respect to the signs of their estimates obtained on the original sample. If the sign changes are not properly taken into account, the standard error of estimates increases dramatically without any real meaning. Therefore, there is a need to make the parameters comparable from a resample to another.

In PLS-Graph the following various options are available:

*Standard*: Resampling statistics are computed without compensating for any sign change. This option may be very conservative as it may yield very high standard errors and, consequently, low $t$-ratios. Therefore, we do not recommend it.

*Individual sign changes*: The signs in each resample are made consistent with the signs in the original sample without ensuring a global coherence. The sign of each individual outer weight in the resample is made equal to the sign of the corresponding $\tilde{w}_{jh}$. This option is not recommend in general because of the lack of global coherence. Nevertheless, it seems to be a good procedure in the case where all signs in the same block are equal, at the original sample level.

*Construct level changes* (*default*): In the case of mode B, the use of outer weights to compare the LVs estimates in the original sample and in the resamples may be misleading in presence of strong multicollinearity between the related MVs. Loadings connecting each LV directly to its own MVs are more appropriate.

The vector of loadings for each LV in each resample is compared to the corresponding vector of loadings in the original sample. Let us denote by $L^S_{jh}$ the estimated loading of the $h$th MV on the $j$th LV from the original sample and by $L^R_{jh}$ the estimated loading of the $h$th MV on the $j$th LV from one resample. The signs of the weights, and consequently the signs of the loadings, are reversed if

$$\left|\sum_h (L^S_{jh} - L^R_{jh})\right| > \left|\sum_h (L^S_{jh} + L^R_{jh})\right|.$$

The options individual sign change and construct level change provide with the same results when the changes of signs within the same block occur for all items.

### *1.5. Use of LVPLS 1.8 and PLS-Graph 3.0: the case of the ECSI model for a mobile phone provider*

We now illustrate the use of Lohmöller's and Chin's programs to compute the parameters of the ECSI model. The data represent the answers given by 250 consumers of a mobile phone provider in a specific European country to the questionnaire defined in Table 1.

Table 2
Check for block unidimensionality

| Block | First eigenvalue | Second eigenvalue | Cronbach's $\alpha$ | Dillon–Goldstein's $\rho$ |
|---|---|---|---|---|
| Image | 2.394 | 0.913 | 0.7228 | 0.819 |
| Customer expectation | 1.444 | 0.903 | 0.4519 | 0.732 |
| Perceived quality | 4.040 | 0.771 | 0.8770 | 0.905 |
| Perceived value | 1.700 | 0.300 | 0.8236 | 0.919 |
| Customer satisfaction | 2.082 | 0.518 | 0.7792 | 0.872 |
| Complaint | — | — | — | — |
| Loyalty | 1.561 | 0.983 | 0.4724 | 0.729 |

#### *1.5.1. Options selection*

Following Fornell's methodology, the following options have been selected:

(1) MVs are not centered nor standardized (METRIC = 4).
(2) Mode A is selected for the external estimation of all LVs.
(3) Centroid scheme is selected for the internal estimation.
(4) *PLSX*: blindfolding (communality and redundancy) on the endogenous LVs customer satisfaction.
(5) *PLS-Graph*: jack-knife and bootstrap.

#### *1.5.2. Check for unidimensionality of each block*

The measurement model chosen for each block is the reflective one. We give in Table 2 the statistics for checking the unidimensionality of each block.

Except for the Cronbach's alpha of blocks customer expectation and loyalty, all these statistics lead to an acceptation of the unidimensionality of all blocks. Statistics for block complaint are not included as this block consists of a single MV.

#### *1.5.3. The ECSI model for a mobile phone provider*

Using PLSX we get (Table 3) the outer weights $\tilde{w}_{jh}$ and the correlations between the MVs and their latent variables. The causality model of Fig. 2 summarizes the various structural regressions of the ECSI model. The path coefficients are the standardized regression coefficients. The $R^2$'s are also shown. The significance levels shown next to the path coefficients in parentheses are coming from PLS-Graph bootstrap with individual sign change option. The significant arrows are in bold.

In Table 4, we check that each MV is more correlated to its own latent variable than to the other LVs. To make this table easier to read, correlations below 0.5 are not shown. It may be noticed that the MV Loyalty 2 does not correctly describe its LV ($\mathrm{cor}(\text{Loyalty 2}, \text{Loyalty}) = 0.273$). This variable should be removed from the model. In fact, it is difficult to give a meaningful answer to this item.

The $R^2$ for customer satisfaction is 0.672 and it is very satisfactory taken into account the complexity of the model.

The value of multiple $R^2$, in the case of standardized variables, may be decomposed in terms of the multiple regression coefficients and correlations between the dependent

Table 3
Outer weights and correlations between MV and LV

| Block | Outer weight $\tilde{w}_{jh}$ | Correlation with LV |
|---|---|---|
| *Image* | | |
| Image 1 | 0.0145 | 0.717 |
| Image 2 | 0.0126 | 0.566 |
| Image 3 | 0.0136 | 0.658 |
| Image 4 | 0.0176 | 0.792 |
| Image 5 | 0.0144 | 0.698 |
| *Customer expectation* | | |
| Cust_exp 1 | 0.0231 | 0.687 |
| Cust_exp 2 | 0.0224 | 0.644 |
| Cust_exp 3 | 0.0253 | 0.726 |
| *Perceived quality* | | |
| Per_qual 1 | 0.0098 | 0.778 |
| Per_qual 2 | 0.0085 | 0.651 |
| Per_qual 3 | 0.0118 | 0.801 |
| Per_qual 4 | 0.0094 | 0.760 |
| Per_qual 5 | 0.0084 | 0.731 |
| Per_qual 6 | 0.0095 | 0.766 |
| Per_qual 7 | 0.0129 | 0.803 |
| *Perceived value* | | |
| Per_val 1 | 0.0239 | 0.993 |
| Per_val 2 | 0.0247 | 0.911 |
| *Customer satisfaction* | | |
| Cust_sat 1 | 0.0158 | 0.711 |
| Cust_sat 2 | 0.0231 | 0.872 |
| Cust_sat 3 | 0.0264 | 0.884 |
| *Complaint* | | |
| Complaint | 0.0397 | 1 |
| *Loyalty* | | |
| Loyalty 1 | 0.0185 | 0.854 |
| Loyalty 2 | 0.0061 | 0.273 |
| Loyalty 3 | 0.0225 | 0.869 |

variable and the explanatory ones as follows:

$$R^2 = \sum_j \hat{\beta}_j \text{cor}(y, x_j).$$

This decomposition allows understanding the contribution of each explanatory variable to the prediction of the dependent one and it makes sense only when the regression coefficients and the related correlations have the same sign.

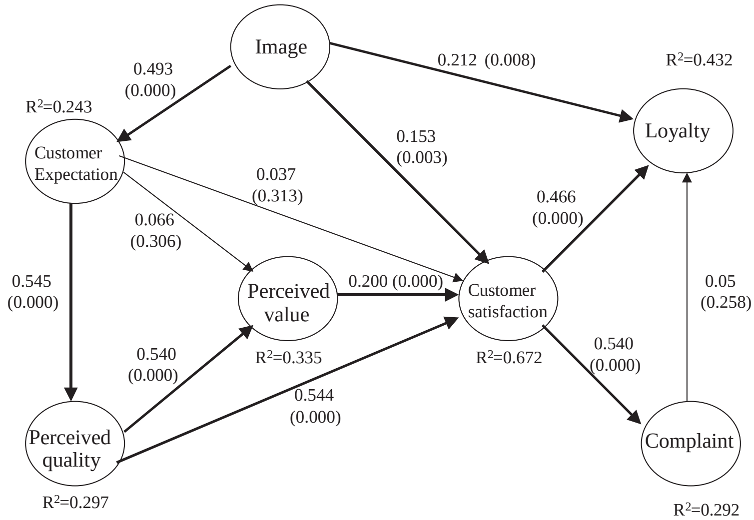


Fig. 2. ECSI causality model for a mobile telephone provider ($p$-values in parentheses come from PLS-Graph bootstrap with individual sign change option).

For our example, Table 5 shows that perceived quality is the most important variable in the prediction of customer satisfaction, contributing to 64.07% of the $R^2$. On the contrary, customer expectation does not contribute at all (about 3%) and in fact it is not significant.

Of course we have to be careful for the interpretation of non-significant path coefficients as they can come from a multicollinearity problem. This suggests to use PLS regression (Wold et al., 1983; Martens and Næs, 1989; Tenenhaus, 1998; Martens and Martens, 2001; Eriksson et al., 2001) instead of OLS multiple regression.

### *1.5.4. The results related to communality and redundancy*

The indices for redundancy, communality and explained variability ($R^2$) are given in Table 6. Redundancy and $R^2$ may not be computed, of course, for exogenous LVs (Image, in this case). It is worth reminding that the average communality is computed as a weighted average of the different communalities with the weights being the number of MVs per each block. However, 1-MV blocks should not be used for the computation of the average communality as they automatically lead to communalities equal to 1.

According to the results in Table 6, the GoF index defined in (30) turns out to be

$$\text{GoF} = \sqrt{0.5702 \times 0.3784} = 0.4645.$$

This index may be also computed within a SEM-ML framework and thus becomes a way of comparing the model validity in terms of goodness of fit between PLS and SEM-ML when they are both applicable.

The cv-communality $H_j^2$ measures the capacity of the path model to predict the MVs directly from their own LV by cross-validation. It uses only the measurement model.

Table 4
Correlations between MVs and LVs

| | Image | Customer expectation | Perceived quality | Perceived value | Customer satisfaction | Complaint | Loyalty |
|---|---|---|---|---|---|---|---|
| Image 1 | 0.717 | | 0.571 | | 0.539 | | |
| Image 2 | 0.565 | | | | | | |
| Image 3 | 0.657 | | | | | | |
| Image 4 | 0.791 | | 0.571 | | 0.543 | | |
| Image 5 | 0.698 | | 0.544 | | 0.500 | | |
| Cust_exp 1 | | 0.689 | | | | | |
| Cust_exp 2 | | 0.644 | | | | | |
| Cust_exp 3 | | 0.724 | | | | | |
| Per_qual 1 | 0.622 | 0.537 | 0.778 | | 0.661 | | |
| Per_qual 2 | — | | 0.651 | | | | |
| Per_qual 3 | 0.621 | | 0.801 | | 0.651 | | |
| Per_qual 4 | — | | 0.760 | | 0.587 | | |
| Per_qual 5 | 0.599 | | 0.732 | | 0.516 | | |
| Per_qual 6 | 0.551 | | 0.766 | | 0.539 | | |
| Per_qual 7 | 0.596 | | 0.803 | 0.547 | 0.707 | | |
| Per_val 1 | | | | 0.933 | | | |
| Per_val 2 | 0.541 | | 0.594 | 0.911 | 0.631 | | 0.524 |
| Cust_sat 1 | 0.558 | | 0.638 | | 0.711 | | |
| Cust_sat 2 | 0.524 | | 0.672 | | 0.872 | | |
| Cust_sat 3 | 0.613 | | 0.684 | 0.588 | 0.884 | 0.547 | 0.610 |
| Complaint | | | 0.537 | | 0.540 | 1 | |
| Loyalty 1 | | | | | | | 0.854 |
| Loyalty 2 | | | | | | | |
| Loyalty 3 | 0.528 | | 0.537 | | 0.659 | | 0.869 |

Table 5
The explanation of customer satisfaction

| Explanatory variables for customer satisfaction | $\hat{\beta}_j$ | Correlation | Contribution to $R^2$ (%) |
|---|---|---|---|
| Image | 0.153 | 0.671 | 15.28 |
| Expectation | 0.037 | 0.481 | 2.67 |
| Perceived value | 0.200 | 0.604 | 17.98 |
| Perceived quality | 0.544 | 0.791 | 64.07 |

The prediction of a MV of an endogenous block is carried out using only the MVs of this block. The communality measures here the quality of the measurement model (mode A for all blocks).

The cv-redundancy $F_j^2$ measures, instead, the capacity of the path model to predict the *endogenous* MVs indirectly from a prediction of their own LV using the related

Table 6
Communality and redundancy

| Block | $R^2$ | Communality | Redundancy |
|---|---|---|---|
| Image | | 0.4760 | |
| Customer expectation | 0.2431 | 0.4711 | 0.1145 |
| Perceived quality | 0.2971 | 0.5737 | 0.1705 |
| Perceived value | 0.3351 | 0.8495 | 0.2846 |
| Customer satisfaction | 0.6717 | 0.6825 | 0.4585 |
| Complaint | 0.2916 | 1.0000 | 0.2916 |
| Loyalty | 0.4318 | 0.5200 | 0.2216 |
| Average | 0.3784 | 0.5702 | 0.2569 |

Table 7
Blindfolding results: cv-communality and cv-redundancy

| Block | Cv-communality $H^2$ | Cv-redundancy $F^2$ |
|---|---|---|
| Image | 0.1977 | |
| Customer expectation | −0.0153 | −0.0218 |
| Perceived quality | 0.4012 | 0.0516 |
| Perceived value | 0.4516 | 0.1211 |
| Customer satisfaction | 0.3877 | 0.4459 |
| Complaint | | 0.0785 |
| Loyalty | 0.1501 | 0.1163 |

structural relation, by cross-validation. The prediction of an MV $\mathbf{x}_{jh}$ of an endogenous block $j$ is carried out using all the MVs of the blocks $j'$ related with the explanatory LVs $\xi_{j'}$ of the LV $\xi_j$. The redundancy measures the quality of the structural model, taking into account the measurement model. This index is recommended by H. Wold for measuring the quality of the path model.

On the ECSI example blindfolding has been carried out using $G = 30$ blocks. The results are given in Table 7.

We may notice that in the ECSI model only the block customer satisfaction has an acceptable cv-redundancy index $F^2$. Due to blindfolding, the cv-communality and the cv-redundancy measures may be negative, as it is here the case for customer expectation. A negative value implies that the corresponding LV is badly estimated.

### *1.5.5. The results of resampling procedures on customer satisfaction*

In this section, the results of jack-knife and bootstrap procedures on the LV customer satisfaction of the ECSI model are given for comparisons with blindfolding and classical regression tests.

*Outer weights*: Table 8 summarizes the $t$-ratios yielded in the different procedures for the outer weights.

As it can be clearly seen, blindfolding produces highly significant results because the estimation procedure is very robust to the presence of missing data thus giving

Table 8
Outer weights validation ($t$-ratios) for customer satisfaction

| Item | Blindfolding | Bootstrap standard | Bootstrap ind. and const. | Jack-knife (adjusted) |
|---|---|---|---|---|
| Cust_sat1 | 54.00 | 9.51 | 9.51 | 6.02 |
| Cust_sat2 | 79.00 | 17.31 | 17.31 | 15.33 |
| Cust_sat3 | 67.75 | 15.53 | 15.53 | 7.73 |

Table 9
Path coefficients validation ($t$-ratios) for customer satisfaction

| Explanatory LV | Blindfolding | Bootstrap standard | Bootstrap ind. and const. | Jack-knife (adjusted) | Multiple regression |
|---|---|---|---|---|---|
| Image | 11.63 | 3.01 | 3.01 | 3.28 | 2.81 |
| Expectation | 3.08 | 0.72 | 1.01 | 1.19 | 0.81 |
| Perceived quality | 38.77 | 8.83 | 8.83 | 4.61 | 9.12 |
| Perceived value | 17.97 | 1.10 | 3.74 | 1.69 | 4.44 |

almost always the same weight for each blindfolded sample. The three options for the bootstrap procedure have given the same results because the weights have been positive for all the resamples. The jackknife statistics have been adjusted in order to allow for the correlation between the samples.

*Path coefficients*: Table 9 summarizes the $t$-ratios yielded in the different procedures for the path coefficients of explanatory LVs on *customer satisfaction*.

This table confirms that blindfolding is not an acceptable procedure. Bootstrap with a standard option does not detect the significance of perceived value due to the many sign changes (104 out of 200 samples!) of its loadings. These changes cause a high standard deviation and, consequently, a low $t$-ratio. Bootstrap with the default construct level changes option seems to be the most valid procedure among the existing ones.

### *1.5.6. The results of resampling procedures on path coefficients of the model*

We have run the bootstrap resampling procedure with the construct level changes option and with different numbers of resamples (namely: 100, 200, 300, 500, 1000). Results are very stable with respect to the number of resamples. However, for some path coefficients, the means of resamples are very far from the estimates on the original sample and the standard deviations are too high. This leads to non-significant path coefficients. Hereinafter, we show the results for some problematic path coefficients in the case of 200 resamples.

For all path coefficients in Table 10, the means of resamples are very far from the original sample estimates with a high standard deviation. The last two path coefficients even show a change in the signs. All these problems lead to non-significant $t$-ratios. Unfortunately, even the construct level changes option does not seem to be enough to control the sign indeterminacy.

Table 10
Path coefficients (construct level changes option, 200 resamples)

| Path coefficients between | Original sample estimate | Mean of resamples | Standard deviation | Bootstrap with constr. option $t$-ratio |
|---|---|---|---|---|
| Perceived quality ⇒ perceived value | 0.540 | 0.3334 | 0.4307 | 1.2538 |
| Perceived quality ⇒ satisfaction | 0.544 | 0.3339 | 0.4266 | 1.2752 |
| Satisfaction ⇒ loyalty | 0.466 | −0.0295 | 0.4773 | 0.9763 |
| Satisfaction ⇒ complaint | 0.540 | −0.0472 | 0.5352 | 1.0089 |

This example is actually a case where all signs within the same block are equal. Consequently, we have also tried the individual sign change option, which yields proper $p$-values as shown in Fig. 2.

Due to the fact that none of the mentioned procedures is uniformly better than the others, at present we keep recommending the usual $t$-test from OLS multiple regression.

However, in order to solve the problem of sign changes when drawing different resamples, we propose to keep, in each resample, the signs of the starting values for the estimation of outer weights equal to the signs as determined in Section 1.3.2.2.3, i.e. on the basis of the elements in the first eigenvector of principal component analysis run on the original sample. Furthermore, we propose to test the significance of coefficients by means of bootstrap confidence intervals built by using the percentiles of the empirical bootstrap distributions yielded by the resamples. This procedure seems to be coherent with the non-parametric nature of the bootstrap validation.

### *1.6. Comparison between PLS and ML approaches in structural equation modeling: application to the ECSI example*

We now compare the PLS path modeling of H. Wold and the maximum likelihood (ML) approach to Structural Equation Modeling of K. Jöreskog. The differences between both approaches are known. Among others, PLS favors the measurement model (the outer model) and ML the structural model (the inner model). In the following, we want to focus on the LVs estimation. The usual procedure used in SEM-ML for estimating the LVs leads to rather different results from PLS. So it is admitted that PLS and SEM-ML give different results.

In SEM-ML, the LVs are estimated using multiple regression from the "theoretical" LVs on the *whole* set of MVs. In PLS, the LVs are estimated using only their own MVs.

When mode A is used in the PLS algorithm, the outer weights $\tilde{w}_{jh}$ are proportional to the covariances between the MVs $\mathbf{x}_{jh}$ and the inner LV estimate $\mathbf{z}_j$.

The outer standardized LV are computed as

$$\mathbf{y}_j = \sum_h \tilde{w}_{jh}\mathbf{x}_{jh} \propto \sum_h \mathrm{cov}(\mathbf{x}_{jh}, \mathbf{z}_j)\mathbf{x}_{jh}. \tag{34}$$

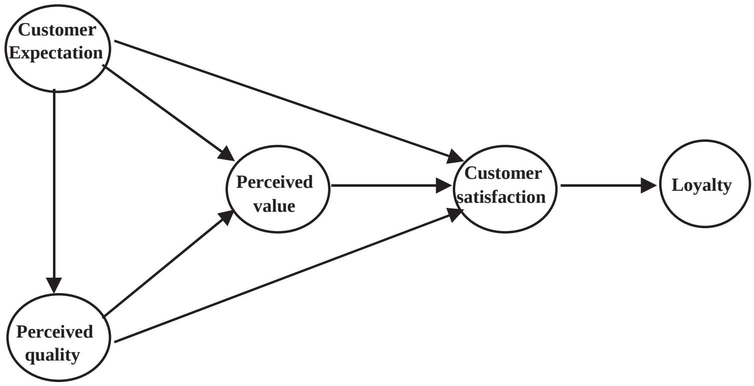


Fig. 3. Fornell's model describing causes and consequences of customer satisfaction.

If we use SEM-ML, we get, as a standard output, the covariances between the MVs and the LVs. We can replace in Eq. (34) the weight $w_{jh} = \text{cov}(\mathbf{x}_{jh}, \mathbf{z}_j)$ by the ML estimate of $\text{cov}(\mathbf{x}_{jh}, \boldsymbol{\xi}_j)$, which is equal to $\pi_{jh} \times \text{Var}(\boldsymbol{\xi}_j)$ (see Eq. (1)). Using ML, we can then obtain the loading estimates $\hat{\pi}_{jh}$ and get a new "PLS/ML" estimate of the LV $\boldsymbol{\xi}_j$:

$$\mathbf{y}_j^{\text{PLS/ML}} = \sum_h \tilde{w}_{jh}^{\text{PLS/ML}} \mathbf{x}_{jh} \propto \sum_h \hat{\pi}_{jh} \mathbf{x}_{jh}. \tag{35}$$

We have found on the ECSI example that both LV estimates $\mathbf{y}_j$ and $\mathbf{y}_j^{\text{PLS/ML}}$ are highly correlated. This confirms a general remark of Noonan and Wold (1982) on the fact that the final outer LV estimates depend very little on the selected scheme of calculation of the inner LV estimates.

To be able to compare ML and PLS we had to come back to the original Fornell's model shown in Fig. 3 because ML did not converge using the model described in Fig. 1.

### 1.6.1. PLS estimation of the Fornell's model for a mobile phone provider

The outer weights are given in column 1 of Table 11 and the model estimates in Fig. 4.

### 1.6.2. ML estimation of the Fornell's model for a mobile phone provider

To estimate the simplified ECSI model with ML, it is necessary to fix some model parameters to solve the indetermination problem. In SEM-ML, the MVs $\mathbf{x}_{jh}$ are centered, but not standardized.

The measurement model (1) is now written as

$$\mathbf{x}_{jh} = \pi_{jh} \boldsymbol{\xi}_j + \boldsymbol{\varepsilon}_{jh}, \tag{36}$$

where the LV $\boldsymbol{\xi}_j$ is centered. For each LV the first loading $\pi_{j1}$ is put to 1. All the other parameters are estimated by maximum likelihood under a hypothesis of multinormality.

Table 11
PLS and ML weights and loadings for Fornell's model (Fig. 3)

| | PLS weight $\tilde{w}_{hj}$ | ML loading $\hat{\pi}_{jh}$ | PLS/ML weight $\tilde{w}_{hj}^{\mathrm{PLS/ML}}$ |
|---|---|---|---|
| Cust_exp 1 | 0.0237 | 1.000 | 0.02428 |
| Cust_exp 2 | 0.0206 | 0.925 | 0.02246 |
| Cust_exp 3 | 0.0262 | 0.998 | 0.02423 |
| Per_quality 1 | 0.0098 | 1.000 | 0.00936 |
| Per_quality 2 | 0.0085 | 0.987 | 0.00924 |
| Per_quality 3 | 0.0118 | 1.241 | 0.01162 |
| Per_quality 4 | 0.0094 | 1.060 | 0.00992 |
| Per_quality 5 | 0.0084 | 0.911 | 0.00853 |
| Per_quality 6 | 0.0095 | 1.045 | 0.00978 |
| Per_quality 7 | 0.0129 | 1.265 | 0.01184 |
| Per_value 1 | 0.0239 | 1.000 | 0.02344 |
| Per_value 2 | 0.0247 | 1.073 | 0.02520 |
| Cust_sat 1 | 0.0157 | 1.000 | 0.01557 |
| Cust_sat 2 | 0.0240 | 1.547 | 0.02409 |
| Cust_sat 3 | 0.0256 | 1.631 | 0.02540 |
| Cus_loyalty 1 | 0.0188 | 1.000 | 0.01929 |
| Cus_loyalty 2 | 0.0050 | 0.202 | 0.00390 |
| Cus_loyalty 3 | 0.0226 | 1.154 | 0.02226 |

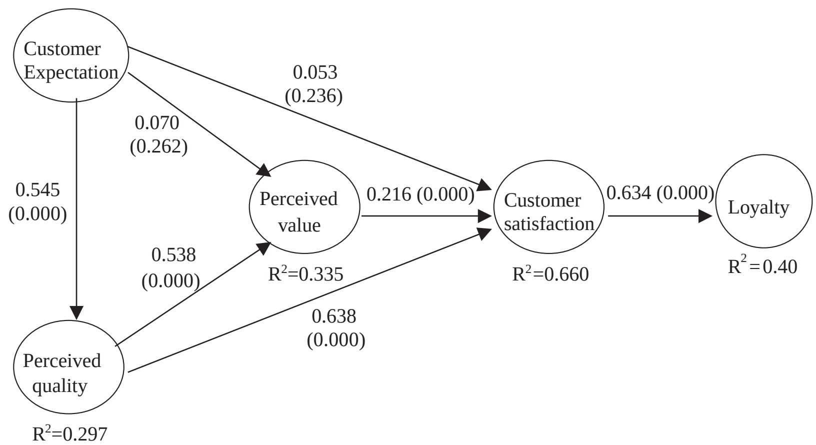


Fig. 4. Fornell's model estimated by PLS (standardized LV).

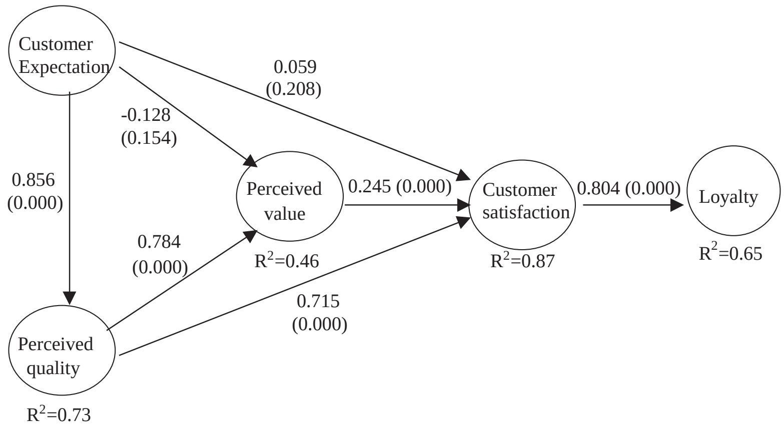


Fig. 5. Fornell's model estimated by ML (standardized LV).

We have used Amos 4.0 (Arbuckle and Wothke, 1999) to run SEM-ML on these data. The SEN-ML estimated loadings $\hat{\pi}_{jh}$ are given in Table 11. The model estimated by ML is then given in Fig. 5.

### 1.6.3. Comparison between PLS and PLS/ML LV estimates for the Fornell's model applied to a mobile phone provider

Using Eq. (35) we get the PLS/ML weights $\tilde{\mathbf{w}}_j^{\text{PLS/ML}}$ leading to the standardized LV estimates $\mathbf{y}_j^{\text{PLS/ML}}$. These loadings $\tilde{\mathbf{w}}_j^{\text{PLS/ML}}$ are given in Table 11.

A graphical comparison between the PLS and the PLS/ML weights is given in Fig. 6.

It appears that the PLS and the PLS/ML weights are very close to each other and will lead to highly correlated LV estimates. We give in Fig. 7 a graphical comparison between the LV estimates $\mathbf{y}_j$ and $\mathbf{y}_j^{\text{PLS/ML}}$. The smallest correlation between $\mathbf{y}_j$ and $\mathbf{y}_j^{\text{PLS/ML}}$ is equal to 0.997.

### 1.6.4. Comparison between PLS and ML estimation of the Fornell's model for a mobile phone provider

The differences between PLS and ML estimates of a causal model come from the order in which model parameters and LVs are calculated, and from the constraints on these ones.

With PLS, LV estimates are first computed subject to the constraint that they must belong to their MV space. Model parameters are then computed using OLS multiple regression.

With SEM-ML, model parameters are first computed and few constraints are imposed on the LVs. The LV estimation does not play any role in the model estimation.

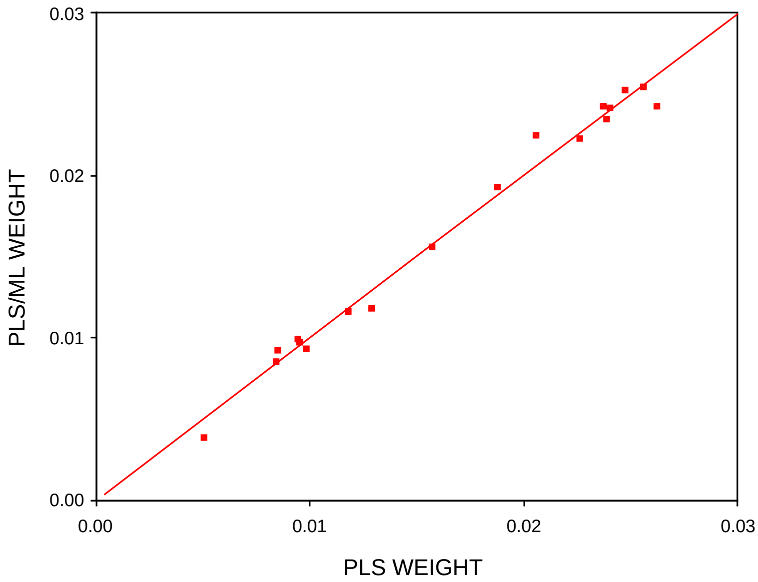


Fig. 6. Comparison between the PLS and the PLS/ML weights.

Consequently, it may be expected that the structural equations are more significant with ML than with PLS (the $R^2$'s are larger) and that the correlations between the MVs and their LVs are stronger with PLS than with ML. Table 12 confirms these results on the ECSI example.

Table 12 contains all needed information for computing the GoF index (introduced in (30)) for both the PLS and the ML models. The GoF values are reported in Table 13. The first two columns are computed for the two models (PLS and ML) whose correlations between each MV and its own LV, and whose $R^2$'s yielded by the structural model are given in Table 12.

The last column of Table 13 reports, instead, the GoF value when considering, in SEM-ML, the estimates of LVs based on all MVs included in the model.

The GoF referred to ML is in both cases better then the one related to PLS. This result has not to be taken as a defeat for PLS because of the following reasons:

1. ML, differently from PLS, did not converge on the full model described in Fig. 1. Therefore, we are forced to compare PLS and ML on the simplified model described in Fig. 3.
2. The GoF index for ML when theoretical LVs are considered is actually computed on unobservable variables. This makes the interpretation meaningless.
3. As a matter of fact, ML with estimated LVs seems the best performing situation in terms of the GoF index. In this case, ML LVs are estimated using a multiple regression of the "theoretical" LV on the whole set **X** of MVs:

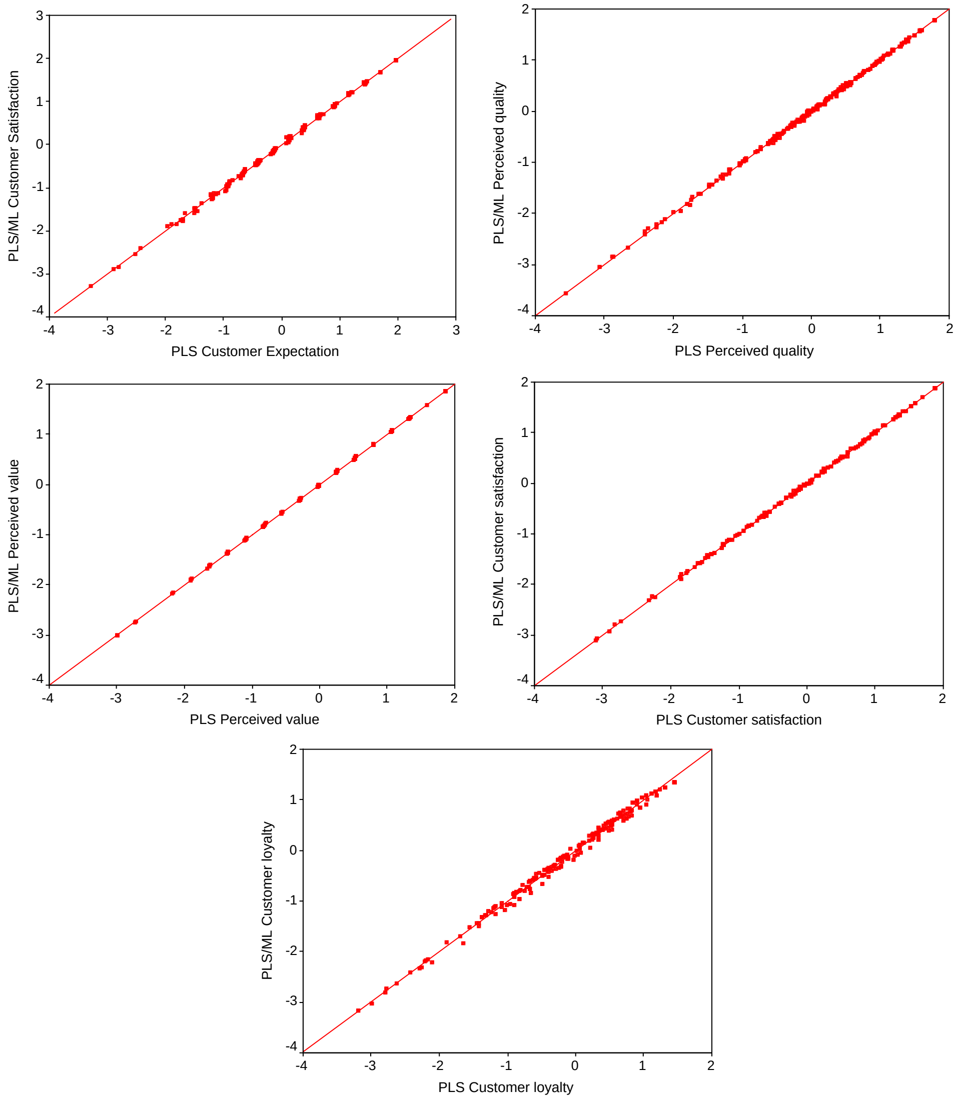


Fig. 7. Comparison between the LVs $\mathbf{y}_j$ and $\mathbf{y}_j^{\mathrm{PLS/ML}}$.

$$\mathbf{y}_j^{\mathrm{ML}} = \mathbf{X}\hat{\boldsymbol{\beta}}_j \text{ with}$$

$$\hat{\boldsymbol{\beta}}_j = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\xi_j = \mathrm{Var}(\mathbf{X})^{-1}\mathrm{Cov}(\mathbf{X}, \xi_j), \tag{37}$$

where $\mathrm{Var}(\mathbf{X})$ and $\mathrm{Cov}(\mathbf{X}, \xi_j)$ are estimated in by maximum likelihood. The regression coefficient vectors $\hat{\boldsymbol{\beta}}_j$ are given in Table 14. This procedure has maybe some optimal mathematical properties, but clearly it is not an acceptable method from a user point of view. Of course, the largest weights are mainly for the MVs related

Table 12
Comparison between PLS and ML

| Measurement model | | |
|---|---|---|
| | Cor with PLS LV | Cor with ML LV |
| Cust_exp 1 | 0.687 | 0.550 |
| Cust_exp 2 | 0.644 | 0.460 |
| Cust_exp 3 | 0.726 | 0.423 |
| Per_quality 1 | 0.778 | 0.775 |
| Per_quality 2 | 0.651 | 0.575 |
| Per_quality 3 | 0.801 | 0.750 |
| Per_quality 4 | 0.760 | 0.707 |
| Per_quality 5 | 0.732 | 0.691 |
| Per_quality 6 | 0.766 | 0.707 |
| Per_quality 7 | 0.803 | 0.756 |
| Per_value 1 | 0.933 | 0.742 |
| Per_value 2 | 0.911 | 0.943 |
| Cust_sat 1 | 0.711 | 0.696 |
| Cust_sat 2 | 0.872 | 0.726 |
| Cust_sat 3 | 0.885 | 0.801 |
| Cus_loyalty 1 | 0.855 | 0.626 |
| Cus_loyalty 2 | 0.273 | 0.118 |
| Cus_loyalty 3 | 0.869 | 0.864 |
| Structural model | | |
| | $R^2$ (PLS) | $R^2$ (ML) |
| Perceived quality | 0.30 | 0.73 |
| Perceived value | 0.33 | 0.46 |
| Customer satisfaction | 0.66 | 0.87 |
| Loyalty | 0.40 | 0.65 |

Table 13
GoF comparison between PLS and ML

| | PLS | ML (theoretical LV) | ML (estimated LV on all MV) |
|---|---|---|---|
| GoF index | 0.5021 | 0.5650 | 0.6589 |

with the estimated LV. But it does not really make sense to estimate a LV using the MVs of the other LVs. The LVs estimated by PLS and by this standard SEM-ML procedure are of course much less correlated than using the PLS/ML procedure. This result is visualized in Fig. 8.

Table 14
ML weights for standard LV estimation

| | Customer expectation | Perceived quality | Perceived value | Customer satisfaction | Customer loyalty |
|---|---|---|---|---|---|
| Cust_exp 1 | 0.11102 | 0.03242 | −0.00083 | 0.01321 | 0.00906 |
| Cust_exp 2 | 0.07433 | 0.02170 | −0.00055 | 0.00884 | 0.00606 |
| Cust_exp 3 | 0.05577 | 0.01628 | −0.00041 | 0.00663 | 0.00455 |
| Per_quality 1 | 0.07362 | 0.12987 | 0.01235 | 0.04578 | 0.03140 |
| Per_quality 2 | 0.02450 | 0.04323 | 0.00411 | 0.01524 | 0.01045 |
| Per_quality 3 | 0.05078 | 0.08959 | 0.00852 | 0.03158 | 0.02166 |
| Per_quality 4 | 0.04627 | 0.08163 | 0.00776 | 0.02877 | 0.01973 |
| Per_quality 5 | 0.04902 | 0.08648 | 0.00822 | 0.03048 | 0.02091 |
| Per_quality 6 | 0.04670 | 0.08238 | 0.00783 | 0.02904 | 0.01992 |
| Per_quality 7 | 0.05152 | 0.09089 | 0.00864 | 0.03204 | 0.02197 |
| Per_value 1 | −0.00071 | 0.00466 | 0.10833 | 0.00992 | 0.00680 |
| Per_value 2 | −0.00440 | 0.02873 | 0.66864 | 0.06122 | 0.04199 |
| Cust_sat 1 | 0.03085 | 0.04709 | 0.02707 | 0.09322 | 0.06393 |
| Cust_sat 2 | 0.02775 | 0.04236 | 0.02435 | 0.08386 | 0.05751 |
| Cust_sat 3 | 0.00360 | 0.05506 | 0.03165 | 0.10898 | 0.07474 |
| Cus_loyalty 1 | 0.00387 | 0.00591 | 0.00340 | 0.01170 | 0.10838 |
| Cus_loyalty 2 | 0.00042 | 0.00064 | 0.00037 | 0.00127 | 0.01183 |
| Cus_loyalty 3 | 0.01533 | 0.02340 | 0.01345 | 0.04632 | 0.42895 |

## 2. The NIPALS algorithm

The roots of the PLS algorithm are in the nonlinear iterative least-squares estimation (NILES), which later became nonlinear iterative partial least squares (NIPALS), algorithm for principal component analysis (Wold, 1966). We now remind the original algorithm of H. Wold and show how it can be included in the PLS framework described in this paper. The interests of the NIPALS algorithm are double as it shows: how PLS handles missing data and how to extend the PLS approach to more than one dimension. We denote by $\mathbf{X}$ the $(n \times p)$ units by variables data table. The variables are supposed to be centered.

*The original algorithm*: The first principal component $\mathbf{t}_1$ is obtained by the following procedure.

Let $\mathbf{w}_1$ be an arbitrary $p$-dimension weight vector. The vector $\mathbf{w}_1$ is supposed to be normalized: $\|\mathbf{w}_1\| = 1$.

The algorithm is based on three steps.

*Step* 1:

$$\mathbf{t}_1 = \frac{\mathbf{X}\mathbf{w}_1}{\mathbf{w}_1'\mathbf{w}_1}.$$

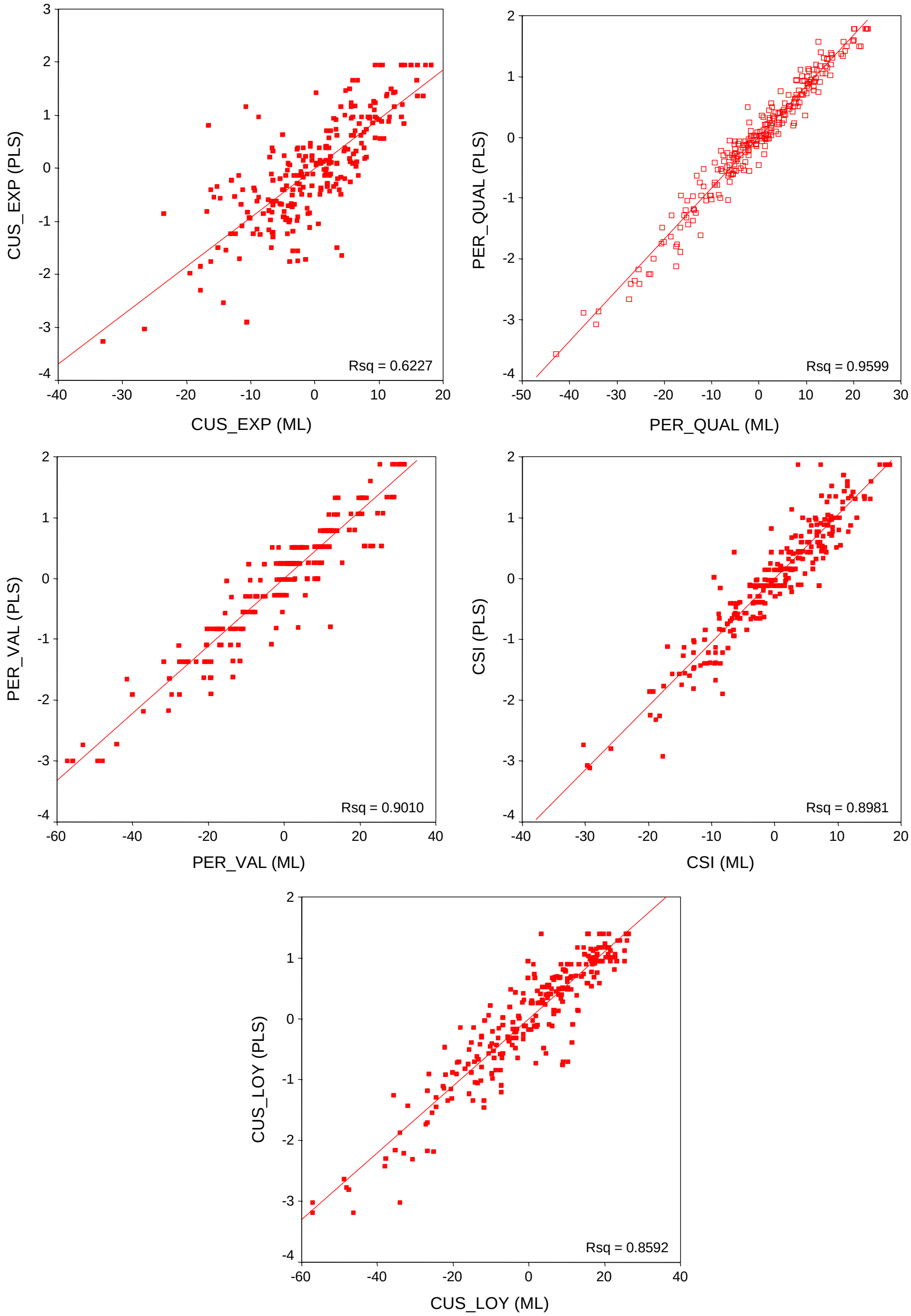


Fig. 8. Comparison between the LV estimates by PLS and ML (standard formula).

*Step* 2:

$$\mathbf{w}_1 = \frac{\mathbf{X}'\mathbf{t}_1}{\mathbf{t}_1'\mathbf{t}_1}.$$

*Step* 3:

$$\mathbf{w}_1 = \frac{\mathbf{w}_1}{\|\mathbf{w}_1\|}.$$

These three steps are iterated until convergence.

It can be noticed that each value $t_{1i}$ of $\mathbf{t}_1$ is the slope of the least-squares line without intercept going through the cloud of points $(\mathbf{w}_1, \boldsymbol{x}_i)$, where $\boldsymbol{x}_i$ is the transposed $i$th row of $\mathbf{X}$.

Similarly, each value $w_{1j}$ is the slope of the least-squares line without intercept going through the cloud of points $(\mathbf{t}_1, \mathbf{x}_j)$, where $\mathbf{x}_j$ is the $j$th column of $\mathbf{X}$.

When there are some missing data, these slopes can still be computed on all the available data. When there are no missing data this algorithm leads to the first normalized eigenvector $\mathbf{w}_1$ of the covariance matrix $\mathrm{Var}(\mathbf{X})$ and to the first principal component $\mathbf{t}_1$ of the data table $\mathbf{X}$.

The research of the second principal component relies on an important result used in multivariate analysis: the set of linear combinations of $\mathbf{x}_j$ orthogonal to $\mathbf{t}_1$ is equal to the set of linear combinations of the residuals $\mathbf{x}_{1j}$ in the simple regressions without intercept of the $\mathbf{x}_j$'s on $\mathbf{t}_1$.

So, the second eigenvector $\mathbf{w}_2$ of $\mathrm{Var}(\mathbf{X})$ and the second principal component $\mathbf{t}_2$ are obtained by replacing in the previous procedure $\mathbf{X}$ by $\mathbf{X}_1 = \mathbf{X} - \mathbf{t}_1\mathbf{w}_1'$, $\mathbf{w}_1$ by $\mathbf{w}_2$, and $\mathbf{t}_1$ by $\mathbf{t}_2$.

The other eigenvectors and principal components are obtained in the same way.

*A modified algorithm*: The original algorithm can be slightly modified to go into the PLS framework described in part 1 of this paper. It is simply a question of normalization: the $\mathbf{t}_h$'s are now standardized.

The modified algorithm is the following. Let $\mathbf{w}_1$ be an arbitrary $p$-dimension weight vector.

The algorithm is now based on two steps.

*Step* 1:

$$\mathbf{t}_1 \propto \mathbf{X}\mathbf{w}_1.$$

*Step* 2:

$$\mathbf{w}_1 = \frac{\mathbf{X}'\mathbf{t}_1}{n} = \mathrm{Cov}(\mathbf{X}, \mathbf{t}_1).$$

These two steps are iterated until convergence.

The reader can recognize that step 2 is exactly mode A of the PLS approach when only one block of data is available.

The other dimensions are obtained by working on the residuals of $\mathbf{X}$ on the previous standardized principal components.

## 3. The PLS approach for two sets of variables

We show in this section how using PLS path modeling allows to find again the main data analysis methods to relate two sets of variables. We give in Table 15 the methods corresponding to the various choices of modes A or B for the LVs $\mathbf{y}_1$ or $\mathbf{y}_2$.

To show these results, it is enough to write the stationary conditions of PLS path modeling. For simplification purposes, but without any lack of generality, we suppose that the LVs $\xi_1$ and $\xi_2$ are positively correlated. Consequently $\mathbf{z}_1 = \mathbf{y}_2$ and $\mathbf{z}_2 = \mathbf{y}_1$.

Denoting by $\mathbf{X}_1$ and $\mathbf{X}_2$ the data matrices associated with the two sets of variables, PLS path modeling leads to the following stationary conditions:

$$\mathbf{y}_j \propto \mathbf{X}_j \mathbf{w}_j \quad \text{for } j = 1, 2, \tag{38}$$

with

$$\mathbf{w}_1 = \frac{1}{n} \mathbf{X}_1' \mathbf{y}_2 \quad \text{for mode A}, \tag{39}$$

$$\mathbf{w}_2 = \frac{1}{n} \mathbf{X}_2' \mathbf{y}_1 \quad \text{for mode A}, \tag{40}$$

$$\mathbf{w}_1 = (\mathbf{X}_1' \mathbf{X}_1)^{-1} \mathbf{X}_1' \mathbf{y}_2 \quad \text{for mode B}, \tag{41}$$

$$\mathbf{w}_2 = (\mathbf{X}_2' \mathbf{X}_2)^{-1} \mathbf{X}_2' \mathbf{y}_1 \quad \text{for mode B}. \tag{42}$$

*Canonical correlation analysis*: Using mode B for blocks $\mathbf{X}_1$ and $\mathbf{X}_2$, the stationary conditions are written as

$$\mathbf{y}_1 \propto \mathbf{X}_1 (\mathbf{X}_1' \mathbf{X}_1)^{-1} \mathbf{X}_1' \mathbf{y}_2, \tag{43}$$

$$\mathbf{y}_2 \propto \mathbf{X}_2 (\mathbf{X}_2' \mathbf{X}_2)^{-1} \mathbf{X}_2' \mathbf{y}_1. \tag{44}$$

So the PLS algorithm does converge to the first canonical components of canonical correlation analysis of $\mathbf{X}_1$ and $\mathbf{X}_2$.

*Inter-battery factor analysis*: Stationary conditions are written as

$$\mathbf{y}_1 \propto \mathbf{X}_1 \mathbf{X}_1' \mathbf{y}_2, \tag{45}$$

$$\mathbf{y}_2 \propto \mathbf{X}_2 \mathbf{X}_2' \mathbf{y}_1, \tag{46}$$

when mode A is applied for blocks $\mathbf{X}_1$ and $\mathbf{X}_2$.

Table 15
Equivalence between the first step of various methods and the PLS algorithm applied to two blocks of variables $\mathbf{X}_1$ and $\mathbf{X}_2$

| | Canonical correlation analysis | Inter-battery factor analysis | Redundancy analysis of $\mathbf{X}_2$ with respect to $\mathbf{X}_1$ |
|---|---|---|---|
| Computation mode for $\mathbf{y}_1$ | B | A | B |
| Computation mode for $\mathbf{y}_2$ | B | A | A |

The PLS algorithm does converge to the first components of the inter-battery factor analysis of two data tables $\mathbf{X}_1$ and $\mathbf{X}_2$ proposed by Tucker (1958).

It may be noticed that the specific case where $\mathbf{X}_1 = \mathbf{X}_2 = \mathbf{X}$ allows to get the first standardized principal component $\mathbf{t}_1$ of $\mathbf{X}$ since $\mathbf{y}_1$, eigenvector of $\mathbf{XX}'$ associated with the largest eigenvalue, is also the eigenvector of $(\mathbf{XX}')^2$ associated with the largest eigenvalue.

*Redundancy analysis of* $\mathbf{X}_2$ with respect to $\mathbf{X}_1$: Using mode B for block $\mathbf{X}_1$ and mode A for block $\mathbf{X}_2$, the stationary conditions become:

$$\mathbf{y}_1 \propto \mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{y}_2, \tag{47}$$

$$\mathbf{y}_2 \propto \mathbf{X}_2\mathbf{X}_2'\mathbf{y}_1. \tag{48}$$

The PLS algorithm converges to the first component $\mathbf{y}_1$ of the redundancy analysis of $\mathbf{X}_2$ with respect to $\mathbf{X}_1$ (Israels, 1984): $\mathbf{y}_1 = \mathbf{X}_1\mathbf{w}_1$ is eigenvector of the matrix $\mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{X}_2\mathbf{X}_2'$ associated with the largest eigenvalue. We find again that $\mathbf{y}_1$ is the first standardized principal component of the $\mathbf{X}_2$ variables projected onto the $\mathbf{X}_1$-variable space.

*Looking for higher dimensions*: By replacing the data tables $\mathbf{X}_1$ and/or $\mathbf{X}_2$ by their residuals $\mathbf{X}_{1j}$ in the regressions of $\mathbf{X}_j$ on $\mathbf{y}_j$ in Eqs. (43)–(48) we get higher-dimension components. This calculation of residuals is called the deflation operation.

For canonical correlation analysis and inter-battery factor analysis, we get the second components by deflating both $\mathbf{X}_1$ and $\mathbf{X}_2$.

In redundancy analysis of $\mathbf{X}_2$ with respect to $\mathbf{X}_1$, the second dimension is obtained by keeping $\mathbf{X}_2$ and deflating $\mathbf{X}_1$.

If this same deflation system is kept in inter-battery factor analysis, we get the second $\mathbf{X}_1$ component in the PLS regression of $\mathbf{X}_2$ on $\mathbf{X}_1$.

The other dimensions are obtained in the same way by replacing one and/or both data tables by their residuals on the previous components.

Finally, wet get (Table 16) the complete equivalence between PLS path modeling of two data tables and four classical multivariate analysis methods. In this table, the use of the deflation operation for the research of higher-dimension components is mentioned.

Table 16
Equivalence between the PLS algorithm applied to two blocks of variables $\mathbf{X}_1$ and $\mathbf{X}_2$ and various method

| | Canonical correlation analysis | Inter-battery factor analysis | PLS regression of $\mathbf{X}_2$ on $\mathbf{X}_1$ | Redundancy analysis of $\mathbf{X}_2$ with respect to $\mathbf{X}_1$ |
|---|---|---|---|---|
| Computation mode for $\mathbf{X}_1$ | B (deflation) | A (deflation) | A (deflation) | B (deflation) |
| Computation mode for $\mathbf{X}_2$ | B (deflation) | A (deflation) | A (no deflation) | A (no deflation) |

## 4. The PLS approach for $J$ sets of variables

In this section, we show that the various options of PLS path modeling (modes A or B for outer estimation; centroid, factorial or path weighting schemes for inner estimation) allow to find again many methods for multiple tables analysis: generalized canonical analysis (the Horst's (1961) one and the Carroll's (1968) one), Multiple Factor Analysis (Escofier and Pagès, 1994), Lohmöller's (1989) split principal component analysis, Horst's (1965) maximum variance algorithm.

The links between PLS and these methods have been studied on practical examples in Guinot et al. (2001) and in Pagès and Tenenhaus (2001).

Let us consider a situation where $J$ blocks of variables $\mathbf{X}_1, \ldots, \mathbf{X}_J$ are observed on the same set of statistical units. In this presentation, we first suppose that each block is essentially unidimensional and can be summarized by one LV $\xi_j$. This assumption will be then relaxed later on.

For estimating these LVs $\xi_j$, Wold (1982) has proposed the hierarchical model defined as follows:

- A new block $\mathbf{X}$ is constructed by merging the $J$ blocks $\mathbf{X}_1, \ldots, \mathbf{X}_J$ into a super block.
- The super block $\mathbf{X}$ is summarized by one LV $\xi$.
- A path model connects each exogenous LV $\xi_j$ to the endogenous LV $\xi$.

An arrow scheme describing a hierarchical model for $J$ blocks of variables is shown in Fig. 9.

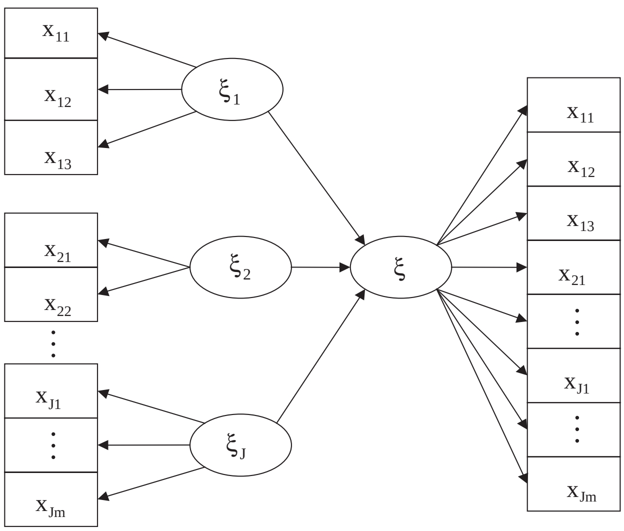


Fig. 9. A hierarchical model for a PLS analysis of $J$ blocks of variables.

### 4.1. PLS approach and Horst's generalized canonical correlation analysis

Lohmöller (1989) has shown that the stationary equations of Horst's generalized canonical analysis (SUMCOR criterion) can be found again by estimating the LVs of the path model described in Fig. 9 when mode B for the outer estimation and centroid scheme for the inner estimation are selected.

#### 4.1.1. The Horst's approach (SUMCOR criterion)

We begin this section by reminding the Horst's approach using the SUMCOR criterion, as described by Saporta (1975).

One searches for standardized variables $\mathbf{v}_j = \mathbf{X}_j\mathbf{a}_j$ maximizing the criterion

$$\sum_j \sum_k \operatorname{cor}(\mathbf{X}_j\mathbf{a}_j, \mathbf{X}_k\mathbf{a}_k). \tag{49}$$

So one searches for the maximum of the criterion

$$\sum_j \sum_k \mathbf{a}'_j \left(\frac{1}{n}\mathbf{X}'_j\mathbf{X}_k\right)\mathbf{a}_k, \tag{50}$$

subject to the constraints

$$\mathbf{a}'_j \left(\frac{1}{n}\mathbf{X}'_j\mathbf{X}_j\right)\mathbf{a}_j = 1 \quad \text{for each } j. \tag{51}$$

Use of Lagrange multipliers leads to the following stationary conditions:

$$\sum_k \mathbf{X}'_j\mathbf{X}_k\mathbf{a}_k - \lambda_j\mathbf{X}'_j\mathbf{X}_j\mathbf{a}_j = 0 \quad \text{for each } j. \tag{52}$$

Denoting by $\mathbf{P}_j$ the projection operator on the $\mathbf{X}_j$-space ($\mathbf{P}_j = \mathbf{X}_j(\mathbf{X}'_j\mathbf{X}_j)^{-1}\mathbf{X}'_j$) Eq. (52) becomes

$$\mathbf{P}_j\left(\sum_k \mathbf{X}_k\mathbf{a}_k\right) = \lambda_j\mathbf{X}_j\mathbf{a}_j \quad \text{for each } j. \tag{53}$$

Moreover, we deduce from Eq. (52) and constraint (51), the equality

$$\sum_j \sum_k \mathbf{a}'_j \left(\frac{1}{n}\mathbf{X}'_j\mathbf{X}_k\right)\mathbf{a}_k = \sum_j \lambda_j. \tag{54}$$

So we must select the solution of the stationary equations (53) that maximizes $\sum_j \lambda_j$.

Let us denote by $\mathbf{v}$ the sum of all the variables $\mathbf{v}_j = \mathbf{X}_j\mathbf{a}_j$. Then, the variance of $\mathbf{v}$ is also equal to $\sum_j \lambda_j$. Let us show this result

$$\begin{aligned}
\frac{1}{n}\mathbf{v}'\mathbf{v} &= \frac{1}{n}\left(\sum_j \mathbf{X}_j\mathbf{a}_j\right)'\left(\sum_k \mathbf{X}_k\mathbf{a}_k\right)\\
&= \sum_j \sum_k \mathbf{a}'_j\left(\frac{1}{n}\mathbf{X}'_j\mathbf{X}_k\right)\mathbf{a}_k\\
&= \sum_j \lambda_j.
\end{aligned}$$

We also deduce from (51) and (53) that $\lambda_j = \mathrm{cov}(\mathbf{v}_j, \mathbf{v})$:

$$\begin{aligned}\lambda_j &= \frac{1}{n}\,\mathbf{v}_j'\mathbf{P}_j\mathbf{v}\\ &= \frac{1}{n}\,\mathbf{v}_j'\mathbf{v}\\ &= \mathrm{cov}(\mathbf{v}_j, \mathbf{v})\end{aligned}$$

and, consequently,

$$\mathrm{cor}(\mathbf{v}_j, \mathbf{v}) = \frac{\lambda_j}{\sqrt{\sum_j \lambda_j}}.$$

We deduce that the Horst's method consists also in maximizing the criterion

$$\sum_j \mathrm{cor}(\mathbf{v}_j, \mathbf{v}), \tag{55}$$

subject to the constraints:

$$\text{Variables } \mathbf{v}_j = \mathbf{X}_j\mathbf{a}_j \text{ are standardized}, \tag{56}$$

$$\mathbf{v} = \sum_j \mathbf{v}_j. \tag{57}$$

The maximum value of (55) is thus equal to $\sqrt{\sum_j \lambda_j}$.

*4.1.2. Equivalence between the Horst's approach and the PLS analysis of the hierarchical model associated with J blocks of variables*

Let us now show that the use of mode B and of the centroid scheme on the hierarchical model associated with $J$ blocks of variables yields the Horst's approach.

In multiple table analysis, one searches for common dimensions among the $J$ tables $\mathbf{X}_j$. Therefore, without lack of generality, we can suppose that all the correlations between the LVs $\boldsymbol{\xi}_j$ and $\boldsymbol{\xi}$ are positive.

The outer estimates $\mathbf{y}_j$ and $\mathbf{y}$ of the LV $\boldsymbol{\xi}_j$ and $\boldsymbol{\xi}$ are written as

$$\mathbf{y}_j \propto \mathbf{X}_j\mathbf{w}_j$$

and

$$\mathbf{y} \propto \mathbf{X}\mathbf{w}.$$

Using the centroid scheme, the inner estimates $\mathbf{z}_j$ and $\mathbf{z}$ of the LVs $\boldsymbol{\xi}_j$ and $\boldsymbol{\xi}$ are written as

$$\mathbf{z}_j \propto \mathbf{y}$$

and

$$\mathbf{z} \propto \sum_j \mathbf{y}_j.$$

The stationary equations of PLS path modeling under mode B can now be written as

$$\mathbf{y}_j \propto \mathbf{P}_j(\mathbf{y}) \tag{58}$$

and

$$\mathbf{y} \propto \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\left(\sum_j \mathbf{y}_j\right). \tag{59}$$

As $\sum_j \mathbf{y}_j$ belongs to the $\mathbf{X}$-space we deduce that

$$\mathbf{y} \propto \sum_j \mathbf{y}_j$$

and, consequently, using Eq. (58), we get the equations

$$\mathbf{X}_j\mathbf{w}_j \propto \mathbf{P}_j\left(\sum_j \mathbf{X}_j\mathbf{w}_j\right), \tag{60}$$

similar to the stationary conditions (53). The announced equivalence is thus obtained.

### 4.2. PLS approach and Carroll's generalized canonical correlation analysis

It has been established in Tenenhaus (1999) that the choice of mode B and factorial scheme on the hierarchical model of Fig. 9 yields Carroll's generalized canonical analysis.

#### 4.2.1. The Carroll's approach

Let us first remind Carroll's generalized canonical correlation analysis. One searches for a standardized variable $\mathbf{v} = \mathbf{Xa}$ maximizing the criterion

$$\sum_j R^2(\mathbf{v}, \mathbf{X}_j). \tag{61}$$

As criterion (61) may be written as

$$\frac{1}{n}\mathbf{v}'\sum_j \mathbf{P}_j\mathbf{v}, \tag{62}$$

we deduce that $\mathbf{v}$ is the eigenvector of the matrix $\sum_j \mathbf{P}_j$ associated with the largest eigenvalue, which, divided by $n$, is then the maximum value of (61).

#### 4.2.2. Equivalence between the Carroll's approach and the PLS analysis of the hierarchical model associated with $J$ blocks of variables

Let us show now that a PLS analysis of the model described in Fig. 9, using mode B and a factorial scheme, yields the Carroll's approach.

The standardized outer estimations $\mathbf{y}_j$ and $\mathbf{y}$ of the LV $\xi_j$ and $\xi$ are written as

$$\mathbf{y}_j \propto \mathbf{X}_j\mathbf{w}_j$$

and

$$\mathbf{y} \propto \mathbf{Xw}.$$

In the factorial scheme, the standardized inner estimates $\mathbf{z}_j$ and $\mathbf{z}$ are computed using the correlation $r_j$ between the LVs $\mathbf{y}$ and $\mathbf{y}_j$. So we get

$$\mathbf{z}_j = \mathrm{sign}(r_j)^* \mathbf{y}$$

and

$$\mathbf{z} \propto \sum_j r_j \mathbf{y}_j.$$

The PLS stationary equations under mode B can now be written as

$$\mathbf{y}_j \propto \mathrm{sign}(r_j)^* \mathbf{P}_j \mathbf{y}$$

and

$$\mathbf{y} \propto \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \left( \sum_j r_j \mathbf{y}_j \right).$$

As the variable $\mathbf{y}$ is standardized, the standard deviation of variable $\mathbf{P}_j\mathbf{y}$ is equal to the absolute value of $r_j$ and, consequently,

$$\mathbf{y}_j = \frac{1}{r_j} \mathbf{P}_j \mathbf{y}. \tag{63}$$

As $\sum_j r_j \mathbf{y}_j$ belongs to the $\mathbf{X}$-space, we deduce

$$\mathbf{y} \propto \sum_j r_j \mathbf{y}_j. \tag{64}$$

Consequently, from (63) and (64) we get the equation

$$\mathbf{y} \propto \left( \sum_j \mathbf{P}_j \right) \mathbf{y}, \tag{65}$$

which shows that $\mathbf{y}$ is eigenvector of $\sum_j \mathbf{P}_j$.

In the PLS approach, the iterative process will be based upon Eq. (65) and will converge towards the eigenvector of $\sum_j \mathbf{P}_j$ associated with the largest eigenvalue. The announced equivalence is thus obtained.

## 4.3. PLS approach and principal component analysis of $\mathbf{X}$

### 4.3.1. Lohmöller's split principal component analysis

Lohmöller (1989) has studied the use of mode A and of the path weighting scheme for estimating the LVs of the structural model described in Fig. 9. He has shown that a solution of the stationary equations related to this model is obtained for the first standardized principal component $\mathbf{y}$ of $\mathbf{X}$ and for variables $\mathbf{y}_j$ which are the standardized fragment of $\mathbf{y}$ related to block $\mathbf{X}_j$. In practice, he has noted that the PLS algorithm converges toward the *first* principal component. Lohmöller has called "split principal component analysis" this calculation of the principal components and of their fragments. In this section we go over his presentation again.

The standardized outer estimates $\mathbf{y}_j$ and $\mathbf{y}$ of the LV $\xi_j$ and $\xi$ are written as

$$\mathbf{y}_j \propto \mathbf{X}_j\mathbf{w}_j$$

and

$$\mathbf{y} \propto \mathbf{X}\mathbf{w}.$$

In the path weighting scheme, the standardized inner estimates $\mathbf{z}_j$ and $\mathbf{z}$ are computed using the correlation $r_j$ between $\mathbf{y}_j$ and $\mathbf{y}$ and the regression coefficient $b_j$ of $\mathbf{y}_j$ in the multiple regression of $\mathbf{y}$ on $\mathbf{y}_1, \ldots, \mathbf{y}_J$. Here we can a priori decide that the correlations $r_j$ are positive and check this result a posteriori. So we get

$$\mathbf{z}_j = \text{sign}(r_j)^*\mathbf{y} = \mathbf{y} \tag{66}$$

and

$$\mathbf{z} \propto \sum_j b_j\mathbf{y}_j. \tag{67}$$

The stationary equations of the PLS approach under mode A can now be written as

$$\mathbf{y}_j \propto \mathbf{X}_j\mathbf{X}_j'\mathbf{y} \tag{68}$$

and

$$\mathbf{y} \propto \mathbf{X}\mathbf{X}'\left(\sum_j b_j\mathbf{y}_j\right). \tag{69}$$

As announced, it appears that the correlation between $\mathbf{y}_j$ and $\mathbf{y}$ is actually positive.

We now show that a solution of the stationary equations (68) and (69) is obtained for a principal component of $\mathbf{X}$.

Let $\mathbf{y} = \mathbf{X}\mathbf{w} = \sum_j \mathbf{X}_j\mathbf{w}_j$ be a standardized principal component. The vector $\mathbf{X}_j\mathbf{w}_j$ is the fragment of $\mathbf{y}$ on the block $\mathbf{X}_j$. The vector $\mathbf{y}$ can be written as $\sum_j b_j(\mathbf{X}_j\mathbf{w}_j/b_j)$ where $b_j$ is the standard deviation of the fragment $\mathbf{X}_j\mathbf{w}_j$. Then putting $\mathbf{y}_j = (1/b_j)\mathbf{X}_j\mathbf{w}_j$, we get $\mathbf{y} = \sum_j b_j\mathbf{y}_j$. The coefficients $b_j$ are also the regression coefficients of $\mathbf{y}_j$ in the regression of $\mathbf{y}$ on $\mathbf{y}_1, \ldots, \mathbf{y}_J$.

As a standardized principal component, $\mathbf{y}$ verifies the expression in (69). Using the fact that the vector $\mathbf{w}$ is proportional to the covariance between $\mathbf{X}$ and $\mathbf{y}$, we can also verify the expression in (68).

As already noted by Lohmöller, we have also verified that, in practice, this algorithm converges toward the *first* principal component of $X$.

#### 4.3.2. Horst's maximum variance algorithm

In the Horst's maximum variance algorithm (Horst, 1965) for generalized canonical correlation models, each block $\mathbf{X}_j$ is transformed in a new block $\mathbf{X}_j^*$ of standardized and uncorrelated variables

$$\mathbf{X}_j^* = \left(\frac{1}{n}\mathbf{X}_j'\mathbf{X}_j\right)^{-1/2}\mathbf{X}_j. \tag{70}$$

Split principal component analysis is then applied to the blocks $\mathbf{X}_j^*$. In this method, the internal structure of each block does not influence the construction of the LVs. These variables only reflect the common structures between the blocks. In practical situation, from a computational point of view, it will be much easier to replace each block $\mathbf{X}_j$ by its standardized principal components associated with a not too small eigenvalue rather than $\mathbf{X}_j^*$. This last method is a second order principal component analysis.

#### 4.3.3. Escofier and Pagès' multiple factor analysis

In multiple factor analysis (Escofier and Pagès, 1994), each block $\mathbf{X}_j$ is transformed into a new block defined by

$$\tilde{\mathbf{X}}_j = \frac{1}{\sqrt{\lambda_{j1}}}\mathbf{X}_j, \tag{71}$$

where $\lambda_{j1}$ is the first eigenvalue of the principal component analysis of block $\mathbf{X}_j$. The new blocks become comparable in the sense that the first eigenvalues of the principal component analyses of the various blocks $\tilde{\mathbf{X}}_j$ are all equal to one. Split principal component analysis is then applied to the blocks $\tilde{\mathbf{X}}_j$. In this method, the internal variance of each block does not influence the construction of the LVs.

### 4.4. PLS generalized canonical correlation analysis

PLS generalized canonical correlation analysis has been introduced in Guinot et al. (2001). Generalized canonical correlation analyses of Horst and Carroll are obtained by using mode B and the centroid and factorial schemes, respectively. PLS Horst and Carroll's generalized correlation analyses are obtained by using mode A and the centroid and factorial schemes, respectively. These two methods lead to an iterative use of PLS regressions. We shall now present these two methods in greater detail.

#### 4.4.1. PLS Horst's generalized canonical correlation analysis

The stationary equations corresponding to the PLS treatment of the arrow diagram in Fig. 9 by choosing mode A and the centroid scheme are reproduced here:

$$\mathbf{y}_j \propto \mathbf{X}_j\mathbf{X}_j\mathbf{y} \tag{72}$$

and

$$\mathbf{y} \propto \mathbf{X}\mathbf{X}'\left(\sum_j \mathbf{y}_j\right). \tag{73}$$

The correlation between $\mathbf{y}_j$ and $\mathbf{y}$ being always positive, it is not necessary to introduce it in Eqs. (72) and (73). A solution of these stationary equations is thus obtained by a sequence of 1-component PLS regressions of $\mathbf{y}$ on each $\mathbf{X}_j$, then of $\sum_j \mathbf{y}_j$ on $\mathbf{X}$. This procedure is iterated until convergence (not proven, but almost always verified in practice). This method, consequently, has all the advantages of PLS regression: there can be more variables than observations and there may be a small amount of data that are missing completely at random.

#### 4.4.2. PLS Carroll's generalized canonical correlation analysis

The stationary equations corresponding to the PLS treatment of the arrow diagram in Fig. 9, when choosing mode A and the factorial scheme, are written as

$$\mathbf{y}_j \propto \mathbf{X}_j\mathbf{X}_j'\mathbf{y}, \tag{74}$$

$$\mathbf{y} \propto \mathbf{X}\mathbf{X}'\left(\sum_j \mathrm{cor}(\mathbf{y}, \mathbf{y}_j) \times \mathbf{y}_j\right). \tag{75}$$

The correlation between $\mathbf{y}_j$ and $\mathbf{y}$ being always positive, it is not necessary to introduce it in Eq. (74). A solution of the stationary equations (74) and (75) is thus obtained by a sequence of 1-component PLS regressions of $\mathbf{y}$ on each $\mathbf{X}_j$, then of $\sum_j \mathrm{cor}(\mathbf{y}_j, \mathbf{y}) \times \mathbf{y}_j$ on $\mathbf{X}$. This procedure is iterated until convergence (same remark as above). This method has all the advantages of PLS regression mentioned above.

### 4.5. A synthetic view of PLS path modeling and multiple table analysis

All the methods described in this part can be organized with respect to the choice of the outer estimation mode (A or B) and of the inner estimation scheme (centroid, factorial or path weighting). Table 17 summarizes these results.

In the methods described in Table 17, the higher-dimension components are obtained by re-running the PLS model after deflation of the **X**-block.

It is also possible to obtain higher dimension orthogonal components on some $\mathbf{X}_j$-blocks (or on all of them). The hierarchical PLS model is re-run on the selected deflated $\mathbf{X}_j$-blocks.

The orthogonality control for higher dimension components is a tremendous advantage of the PLS approach (see Tenenhaus (2004) for more details and an example of application).

Table 17
PLS path modeling and multiple table analysis

| Mode of calculation for the outer estimation | Scheme of calculation for the inner estimation | | |
|---|---|---|---|
| | Centroid | Factorial | Path weighting scheme |
| A | PLS Horst's generalized canonical correlation analysis | PLS Carroll's generalized canonical correlation analysis | Lohmöller's split principal component analysis<br>Horst's maximum variance algorithm<br>Escofier and Pagès multiple factor analysis |
| B | Horst's generalized canonical correlation analysis (SUMCOR criterion) | Carroll's generalized canonical correlation analysis | |

Finally, PLS path modeling may be meant as a general framework for the analysis of multiple tables. It is demonstrated that this approach recovers usual data analysis methods in this context but it also allows for new methods to be developed when choosing different mixtures of estimation modes and schemes in the two steps of the algorithm (internal and external estimation of the LVs) as well as different orthogonality constraints. Therefore, we can state that PLS path modeling provides a very flexible environment for the study of a multi-block structure of observed variables by means of structural relationships between LVs. Such a general and flexible framework also enriches the data analysis methods with non-parametric validation procedures (such as bootstrap, jackknife and blindfolding) for the estimated parameters and fit indices for the different blocks that are more classical in a modeling approach than in data analysis.

## Acknowledgements

This paper has been supported by the ESIS (European Satisfaction Index System) IST Project within the Vth Framework Programme (IST-2000-31071) and the Fondation HEC (Paris, France).

## References

Amato, S., Esposito Vinzi, V., Tenenhaus, M., 2004. A global goodness-of-fit index for PLS structural equation modeling. Oral Communication to PLS Club, HEC School of Management, France, March 24.

Arbuckle, J.L., Wothke, W., 1999. Amos 4.0 User's Guide. Smallwaters Corporation, Chicago.

Bayol, M.P., De La Foye, A., Tellier, C., Tenenhaus, M., 2000. Use of PLS path modeling to estimate the European consumer satisfaction index (ECSI) model. Statist. Appl. Italian J. Appl. Statist. 12 (3), 361–375.

Carroll, J.D., 1968. A generalization of canonical correlation analysis to three or more sets of variables. Proceedings of the 76th Convention of the American Psychological Association, pp. 227–228.

Chin, W.W., 1998. The partial least squares approach for structural equation modeling. In: Marcoulides, G.A. (Ed.), Modern Methods for Business Research. Lawrence Erlbaum Associates, London, pp. 295–336.

Chin, W.W., 2001. PLS-Graph User's Guide. C.T. Bauer College of Business, University of Houston, USA.

Chin, W.W., Newsted, P.R., 1999. Structural equation modeling analysis with small samples using partial least squares. In: Hoyle, R. (Ed.), Statistical Strategies for Small Sample Research. Sage Publications, Beverly Hills, CA, pp. 307–341.

Eriksson, L., Johansson, E., Kettaneh-Wold, N., Wold, S., 2001. Multi- and Megavariate Data Analysis: Principles and Applications. Umetrics Academy, Umeå.

Escofier, B., Pagès, J., 1994. Multiple factor analysis (AFMULT package). Comput. Statist. Data Anal. 18, 121–140.

Fornell, C., 1992. A national customer satisfaction barometer: the Swedish experience. J. Marketing 56, 6–21.

Fornell, C., Cha, J., 1994. Partial least squares. In: Bagozzi, R.P. (Ed.), Advanced Methods of Marketing Research. Basil Blackwell, Cambridge, MA, pp. 52–78.

Guinot, C., Latreille, J., Tenenhaus, M., 2001. PLS path modelling and multiple table analysis. Application to the cosmetic habits of women in Ile-de-France. Chemometrics Intell. Lab. Systems 58, 247–259.

Horst, P., 1961. Relations among $M$ sets of variables. Psychometrika 26, 126–149.

Horst, P., 1965. Factor Analysis of Data Matrices. Holt, Rinehart and Winston, New York.

Israels, A.Z., 1984. Redundancy analysis for qualitative variables. Psychometrika 49, 331–346.

Jöreskog, K.G., 1970. A general method for analysis of covariance structure. Biometrika 57, 239–251.

Jöreskog, K.G., Wold, H., 1982. The ML and PLS techniques for modeling with latent variables: historical and comparative aspects. In: Jöreskog, K.G., Wold, H. (Eds.), Systems Under Indirect Observation, Part 1. North-Holland, Amsterdam, pp. 263–270.

Lohmöller, J.-B., 1987. LVPLS Program Manual, Version 1.8. Zentralarchiv für Empirische Sozialforschung, Universität Zu Köln, Köln.

Lohmöller, J.-B., 1989. Latent Variables Path Modeling with Partial Least Squares. Physica-Verlag, Heildelberg.

Martens, H., Martens, M., 2001. Multivariate Analysis of Quality: An Introduction. Wiley, Chichester.

Martens, H., Næs, T., 1989. Multivariate Calibration. Wiley, New York.

Noonan, R., Wold, H., 1982. PLS path modeling with indirectly observed variables: a comparison of alternative estimates for the latent variable. In: Jöreskog, K.G., Wold, H. (Eds.), Systems under Indirect Observation. North-Holland, Amsterdam, pp. 75–94.

Pagès, J., Tenenhaus, M., 2001. Multiple factor analysis combined with PLS path modelling. Application to the analysis of relationships between physicochemical variables, sensory profiles and hedonic judgements. Chemometrics Intell. Lab. Systems 58, 261–273.

Saporta, G., 1975. Liaison entre plusieurs ensembles de variables et codage des données qualitatives. Thèse de Troisième Cycle, Université Pierre et Marie Curie, Paris.

Tenenhaus, M., 1988. Canonical analysis of two convex polyhedral cones. Psychometrika 53 (4), 503–524.

Tenenhaus, M., 1998. La Régression PLS. Éditions Technip, Paris.

Tenenhaus, M., 1999. L'approche PLS. Rev. Statist. Appl. 47 (2), 5–40.

Tenenhaus, M., 2004. PLS Regression and PLS Path Modeling for Multiple Table Analysis. COMPSTAT 2004 Proceedings. Physica-Verlag, Springer, Wurzburg, Berlin.

Tucker, L.R., 1958. An inter-battery method of factor analysis. Psychometrika 23 (2), 111–136.

Wold, H., 1966. Estimation of principal components and related models by iterative least squares. In: Krishnaiah, P.R. (Ed.), Multivariate Analysis. Academic Press, New York, pp. 391–420.

Wold, H., 1973. Non-linear iterative partial least squares (NIPALS) modelling. Some current developments. In: Krishnaiah, P.R. (Ed.), Multivariate Analysis, Vol. III. Academic Press, New York, pp. 383–407.

Wold, H., 1975. Soft modelling by latent variables: the non-linear iterative partial least squares (NIPALS) approach. In: Gani, J. (Ed.), Perspectives in Probability and Statistics: Papers, in Honour of M.S. Bartlett on the Occasion of his Sixty-fifth Birthday. Applied Probability Trust, Academic, London, pp. 117–142.

Wold, H., 1979. Model Construction and Evaluation when Theoretical Knowledge is Scarce: an Example of the Use of Partial Least Squares. Cahier 79.06 du Département d'économétrie, Faculté des Sciences Économiques et Sociales, Université De Genève, Genève.

Wold, H., 1982. Soft modeling: the basic design and some extensions. In: Jöreskog, K.G., Wold, H. (Eds.), Systems under Indirect Observation, Part 2. North-Holland, Amsterdam, pp. 1–54.

Wold, H., 1985. Partial least squares. In: Kotz, S., Johnson, N.L. (Eds.), Encyclopedia of Statistical Sciences, Vol. 6. Wiley, New York, pp. 581–591.

Wold, S., Martens, H., Wold, H., 1983. The multivariate calibration problem in chemistry solved by the PLS method. In: Ruhe, A., Kågström, B. (Eds.), Proceedings of the Conference on Matrix Pencils. Lectures Notes in Mathematics. Springer, Heidelberg.