Università degli Studi di Napoli
Federico II

Partial Least Squares Methods for Non-Metric Data

Giorgio Russolillo

Doctoral Thesis
in Statistics

XXII Ciclo

Partial Least Squares Methods for Non-Metric Data

Napoli, November 30$^{\text{th}}$ 2009

a mia madre,
esempio di umiltà
e di coraggio

Acknowledgments

Il mio primo ringraziamento non può che andare al mio Maestro, il Prof. Carlo Lauro. Tutto ciò che c'è di buono in questo lavoro è il frutto dei suoi insegnamenti; eventuali errori sono invece da attribuire al sottoscritto, ed alla sua ottusa testardaggine che gli impone di fare sempre tutto da solo. In questi tre anni il Prof. Lauro mi ha dato la possibilità di crescere, guidandomi con passione ed affetto. Inoltre, ha fatto in modo che potessi conoscere altri illustri professori, che si sono dimostrati disponibili ad aiutarmi con consigli preziosi.

Tra questi c'è il Prof. Michel Tenenhaus. C'est bizarre comment un petit mot peut changer les événements, n'est-ce pas, monsieur? Il y a quatre ans vous me dîtes: "Pourquoi pas?" Et cette thèse a commencé à naître.

Un'altra persona che si è sempre dimostrata disponibile nei miei confronti è il Prof. Vincenzo Esposito Vinzi. Ci tengo a dirgli che ho apprezzato molto i suoi preziosi consigli, ma ancora di più il fatto che

mi abbia onorato della sua amicizia.

A proposito di eventi.. Per una serie di eventi fortuiti la mia brillante carriera nel settore alimentare è stata bruscamente interrotta per lasciar posto a quella di statistico. In particolare, il fato ha voluto che incontrassi una persona di un'intelligenza francamente imbarazzante, seconda solo al suo altruismo. Mi sento in debito con lui, perchè ha fatto molto per me. Per ora posso solo contraccambiare con queste righe, oltre che con la mia amicizia. Grazie, Elia.

Infine, un ringraziamento speciale va alla Dott.ssa Laura Trinchera, statistica per professione, correttrice di bozze per amore. Da qualche tempo me la ritrovo accanto nel letto ogni qual volta apro gli occhi la mattina. Dicesi felicità.. Grazie, Amore mio, senza di te non ce l'avrei mai fatta.

Giorgio

Contents

List of Figures

List of Tables

Introduction

Partial Least Squares (PLS) methods embrace a suite of data analysis techniques based on algorithms belonging to PLS family. These algorithms consist in various extensions of the Nonlinear estimation by Iterative PArtial Least Squares (NIPALS) algorithm, which was proposed by Herman Wold [Wold 1966b] as an alternative algorithm for implementing a Principal Component Analysis (PCA) [Hotelling 1933]. The peculiarity of this algorithm is that it calculates principal components by means of an iterative sequence of simple ordinary least squares (OLS) regressions. This feature allows overcoming computational problems due to missing data or landscape data matrices, i.e. matrix having more columns than rows. Later on, Wold proposed NIPALS to analyze causal relations between several blocks of variables [Wold 1975b]: the PLS approach to Structural Equation Modeling (SEM) [Bollen 1989], later called PLS-Path Modeling (PLS-PM), was born. In the same period Svante Wold, Herman's son, perceived that PLS approach could be used in order to implement a regularized component-based regression. He called this regression technique

PLS-Regression (PLS-R) [Wold, Martens & Wold 1983].

PLS techniques, as all quantitative methods, were born to handle data sets forming metric spaces. This involves that all the variables embedded in the analysis are observed on interval or ratio scales.

An interval scale consists of a set of numerical values, for which it makes sense to calculate differences. In practice, there are not so many variables which are actually measured at interval scale level. A good example of an interval scale is the Fahrenheit scale for temperature. Equal differences on this scale represent equal differences in temperature, but a temperature of 30 degrees is not twice as warm as one of 15 degrees. The distinguishing feature of a ratio scale is the possession of a non-arbitrary zero value. Other examples of variables measured at ratio scale level are most of physical measurements, time, and count variables.

The most important central tendency indexes (mode, median and arithmetic mean) and dispersion indexes (standard deviation, range), as well as the Pearson product-moment correlation coefficient can be calculated on both interval and ratio measurements. Hence, in most of statistical techniques differences between ratio and interval scale is not relevant. In this work, variables measured at ratio or interval scale level will be referred as *numeric* or *metric* variables.

Unfortunately, in many fields where PLS methods are largely applied (*e.g.* genomics, sensorial analysis, consumer analysis, marketing) researchers are interested in analyzing set of variables measured on a non-metric scale, *i.e.* categorical variables.

A categorical variable has a measurement scale consisting of a set of categories. Categorical variables have two primary types of scales. Variables having categories without a natural ordering are called nominal. Example are religious affiliation, mode of transportation to work or favorite type of music. For the nominal variables, the order of listing the categories is irrelevant. Categorical variables having ordered categories are called ordinal. Ordinal variables have ordered categories (or levels), but even in this case distances between categories are unknown. Classical examples of ordinal variable are the education level and social-economic status. However many other variables, which could seem numerical (and are often handled if they were) are instead ordinal. In particular, any variable which is expression of a judgement can not be considered numeric, because judgments are subjective evaluations. Teacher's judgement expressed in terms of grade, for example, is measured at a metric scale level only if it is obtained as a count variable (number of right answers to a set of questions). In the other cases, grade ought to be considered as ordinal data; as a matter of fact, as it is well known in psychometrics, pairs of differences between consecutive points on the scale are non equal.

Variables measured on a nominal scale are sometime referred to as categorical, while here the word *categorical* is referred to both ordinal and nominal measurements. To avoid confusion, in this work ordinal and nominal variables will be referred to as *non-metric* data as well. Moreover, defining nominal and ordinal variables as non-metric variables allows us to pinpoint that from the mathematical-statistical

point of view, ordinal data are much more similar to nominal data than to numerical data, as they do not follow a metric.

Handling categorical variables is still an open issue in all PLS methodologies. This thesis focuses on new methodological proposals to make PLS techniques able to handle jointly metric and non-metric data. In particular, a new generation of PLS based algorithms, called Non-Metric PLS (NM-PLS) algorithms, is proposed.

The core of all PLS algorithms is an iterative process with which parameters are calculated. The main idea in this thesis is that potentiality of these algorithms are not fully exploited. Until now, PLS has been used in order to analyze data sets of any shape and with missing data. The aim of this thesis is to show how to modify NIPALS, PLS-R and PLS-PM algorithms in order to make them able to work as *optimal scaling* techniques. Three new algorithms, *i.e.* the Non-Metric NIPALS algorithm, the Non-Metric PLS Regression algorithm, and the Non-Metric PLS Path Modeling algorithm, are proposed and their proprieties are discussed in detail.

Throughout this thesis, notation and wording will be used respecting the field and the framework in which the PLS techniques are used. Hence, for example, in order to indicate the scores vector resuming a block of variables, word *component* will be used referring to Principal Component Analysis and PLS Regression, while *latent variable* will be used in SEM framework. At the same way, latin notations typical of completely explorative analysis are used in PCA and PLS-Regression context, while greek notations typical of SEM framework are used

when PLS-PM is discussed.

Thesis outline

In chapter one PLS approaches to Principal Component Analysis, Regression analysis, and Structural Equation Modeling are reviewed. In section 1.2 the PLS approach to Principal Component Analysis is discussed. In this section particular attention is payed to NIPALS algorithm, and its links to Power Method. In the following sections, PLS Regression (section 1.3) and PLS Path Modeling (section 1.4) are discussed in details, paying particular attention to both algorithmic aspects and diagnostic tools.

In chapter two measurement scales and scaling methods are introduced. The various scales of measurement are reviewed, and their properties are discussed (section 2.2). The concept of scaling is introduced (section 2.3). The scaling analysis, *i.e.* the transformation of measurements in order to yield a new set of measurements at a different level is investigated. Metrics of new interval scales can be constrained depending on the scale at which a raw variable is measured and on which of its properties the researcher wants to preserve. All of these restrictions are discussed in the final section of the chapter.

In chapter three first we explained why PLS methods can handle only metric data (section 3.1); afterwards, we discuss the use of binary coding in PLS framework (section 3.2). Requirements and fundamental properties of optimal scaling methods are discussed; Alternating Least Squares (ALS) basic principles, as well as its algorithmic flow,

are illustrated (section 3.3.1). To conclude, a new suite of optimal scaling methods, based on PLS algorithms, is proposed (section 3.4). This new class of methods, called Non-Metric PLS (NM-PLS), perform non-metric and non-linear analysis in PLS framework. In particular, NM-PLS methods exploit the features of PLS iteration in order to provide optimal scaling of variables. The different levels of scaling analysis conceived in NM-PLS methods are discussed, as well as their properties (section 3.4.1).

Chapters four, five and six describe in details non-metric PLS approaches to Principal Component Analysis, Regression analysis and Structural Equation Modeling. In each chapter first we review the specific literature concerning the use of categorical variables, then we explain in detail the original proposition. Optimality properties of each proposition are discussed.

In particular, in chapter four first we review the history and the methodology of the multivariate descriptive analysis of categorical variables by non parametric techniques (sections 4.2 and 4.3). Then, the Non-Metric NIPALS algorithm is described (section 4.5 and 4.5.1), and its connections to other non-metric approaches to PCA are discussed.

In chapter five we focus on PLS Regression (PLS-R). First, the main approaches to non-linear and non-metric analysis in PLS-R framework are briefly presented (section 5.2). Then, an adjusted PLS-R algorithm, called Non-Metric PLS Regression (NM-PLSR) algorithm, is proposed (section 5.4). The optimal scaling properties of the Non-

Metric PLS regression algorithm are proven, and its algorithmic flow is showed. To conclude an application of NM-PLSR to consumer preferences analysis is provided in order to show the potentiality of the method and its interpretation rules (section 5.6).

Non-Metric PLS method for Structural Equation Modeling is presented in the chapter six. First the main techniques to handle non-metric data in PLS Path Modeling framework are discussed (section 6.2). Then, the Non-Metric PLS Path Modeling algorithm (MNPLS-PM) is proposed (section 6.4). The algorithm is explained in details and its optimality properties are discussed. To conclude, an application of NMPLS-PM to macro-economic data is presented in order to highlight how it can handle variables observed on a variety of measurement scales, as well non linearity (section 6.7).

The codes for NM-PLSR and NM-PLSPM algorithms in R environment are provided in the appendix.

Chapter 1

Partial Least Squares methods

1.1 Introduction

Partial Least Squared (PLS) methods involve a set of multivariate techniques based on algorithms belonging to the PLS family. The father of these algorithms was Herman Wold, who in 1966 devised the NILES (Non-linear Iterative Least Squares) algorithm [Wold 1966a]. Wold [1966b] proposed NILES as an iterative estimation method for Principal Components Analysis (PCA) [Hotelling 1933]. NILES calculates principal components by means of an iterative sequence of simple Ordinary Least Squares (OLS) regressions. Its usefulness is due to the fact that NILES yields a Singular Value Decomposition (SVD) of a data matrix regardless of the shape of the matrix and the presence of

missing data. NILES was later re-named Non-linear Iterative PArtial Least Squares (NIPALS) by the same author [Wold 1975a].

Later on, H. Wold [1975b] extended NIPALS basic principles to a more general technique that analyzes several blocks of variables linked by a network of relations specified by a path diagram. This technique summarizes each block of observed variables (manifest variables, MV), in a latent variable (LV) and investigates the relations among the LVs. Since this technique avoids restrictive hypothesis, i.e. multivariate normality and large samples, underlying maximum likelihood techniques, it is used to estimate Structural Equation Models (SEM) [Bollen 1989] parameters, as a *Soft Modeling* [Wold 1982] alternative to Jöreskog's Covariance Structure Analysis (CSA) [Jöreskog 1970], commonly known as LISREL (LInear Structural RELations). During the 1980's, Fred Bookstein deepened the study of PLS in Structural Equation Modeling framework [Bookstein 1982] and, in collaboration with Claes Fornell, was the first to apply PLS to Consumer Satisfaction Analysis [Fornell & Bookstein 1982]. Nowadays, the Partial Least Squares approach to SEM is known with the acronym PLS-PM (PLS Path Modeling), and has became a standard tool in that field.

PLS-PM algorithm was slightly modified by Wold's son, Svante, and Harald Martens, in order to obtain a regularized component based regression tool, known as PLS Regression (PLS-R) [Wold et al. 1983, Wold, Ruhe, Wold & Dunn 1984]. PLS1 and PLS2 algorithms implement respectively single and multiple response PLS Regression. Due to its usefulness in handling a large number of multicollinear predic-

tors, PLS-R has become in the following years a standard tool in Chemometrics and Sensometrics [Wold, Sjöström & Eriksson 2001].

For an extensive historical review of PLS methods, refer to Sánchez [2009].

The core of any PLS algorithm is the iterative procedures used to compute model parameters. PLS iterative procedures are exploited to analyze covariance within a block of variables, or cross-covariance among two o more blocks of variables, avoiding problems linked to missing data and landscape shaped matrices.

In the next, PLS approaches to PCA (section 1.2), Regression (section 1.3) and SEM (section 1.4) will be reviewed, paying particular attention to their algorithmic aspects.

1.2 PLS approach to Principal Component Analysis

Let $\boldsymbol{X}$ be an observation $\times$ variable data matrix. Each matrix element x_{ip} is the measurement of the p-th variable ($p = 1 \dots P$) on the i-th observation ($i = 1 \dots N$). We suppose the variables be centered and normalized to unitary variance.

Principal Component Analysis (PCA) [Hotelling 1933] summarizes $\boldsymbol{X}$ in a H-dimensional space ($H << P$) spanned by orthogonal Principal Components (PCs) $\boldsymbol{t}_h$ ($h = 1 \dots H$). The h-th PC $\boldsymbol{t}_h$ is obtained as a linear combination of the P variables with a unit-norm weight vector $\boldsymbol{p}_h$. PCA searches for the set of weights maximizing the variability of

each $\boldsymbol{t}_h$.

Classical PCA algorithm consists in the eigen analysis of the correlation matrix $\boldsymbol{X}^{'}\boldsymbol{X}$. The matrix $\boldsymbol{P}_H = [\boldsymbol{p}_1 \dots \boldsymbol{p}_H]$ of the eigenvectors associated to the H greatest eigenvalues of $\boldsymbol{X}^{'}\boldsymbol{X}$ contains the set of weights we are searching for. The score matrix $\boldsymbol{T}_H = [\boldsymbol{t}_1 \dots \boldsymbol{t}_A]$ is successively calculated as $\boldsymbol{T}_H = \boldsymbol{X}\boldsymbol{P}_H$. It is noteworthy that loading vectors $\boldsymbol{p}_h$ may be interpreted even as the weight vectors used for building the components, as

$$\boldsymbol{t}_h = \boldsymbol{X}_h \boldsymbol{p}_h$$

In fact, in PCA the concepts of weight and loading coincide because of the double orthogonality of components and weights.
Since eigenvectors are orthogonal by construction, $\boldsymbol{X}$ can be obtained as $\boldsymbol{T}\boldsymbol{P}^{'}$, where $\boldsymbol{T} = [\boldsymbol{t}_1 \dots \boldsymbol{t}_A]$, $\boldsymbol{P} = [\boldsymbol{p}_1 \dots \boldsymbol{p}_A]$, and A is the rank of $\boldsymbol{X}$. That's why the matrix $\boldsymbol{P}$ is commonly called loading matrix.

The PLS approach to PCA, that is the Nonlinear Iterative PArtial Least Squares (NIPALS) algorithm [Wold 1966a, Wold 1966b, Wold 1975a], works in a slightly different way. Firstly, It finds the first loading vector as the dominant right singular vector of $\boldsymbol{X}$, and the first score vector as $\boldsymbol{t}_1 = \boldsymbol{X}\boldsymbol{p}_1$. Afterwards, it deflates $\boldsymbol{X}$ by calculating the residuals of the regression of $\boldsymbol{X}$ on $\boldsymbol{t}_1$, and finds the loading vector of the second PC as the dominant right singular vector of the residual matrix. Working on the deflated matrices assures the orthogonality among the components. High order components are obtained analogously. For each component, loadings are computed through an

iterative procedure in which loading and score vectors are calculated iteratively each one as a function of the other.

The NIPALS algorithm pseudo-code is shown in algorithm 1.

Algorithm 1 NIPALS algorithm

Input: $\boldsymbol{E}_0 = \boldsymbol{X}$
Output: $\boldsymbol{P} = [\boldsymbol{p}_1, \ldots, \boldsymbol{p}_H], \boldsymbol{T} = [\boldsymbol{t}_1, \ldots, \boldsymbol{t}_H]$
  for all $h = 1, \ldots, H$ **do**
    Step 0: Initialize $\boldsymbol{t}_h$
    Step 1:
    repeat
      Step 1.1: $\boldsymbol{p}_h = \boldsymbol{E}_{h-1}^{'}\boldsymbol{t}_h/(\boldsymbol{t}_h^{'}\boldsymbol{t}_h)$
      Step 1.2: $\boldsymbol{p}_h = \boldsymbol{p}_h/\|\boldsymbol{p}_h\|$
      Step 1.3: $\boldsymbol{t}_h = \boldsymbol{E}_{h-1}\boldsymbol{p}_h/(\boldsymbol{p}_h^{'}\boldsymbol{p}_h)$
    until convergence of $\boldsymbol{p}_h$
    Step 2: $\boldsymbol{E}_h = \boldsymbol{E}_{h-1} - \boldsymbol{t}_h\boldsymbol{p}_h^{'}$
  end for

The relationships in the iteration step of the algorithm verify the following equations:

$$\boldsymbol{E}_{h-1}^{'}\boldsymbol{E}_{h-1}\boldsymbol{p}_h = \lambda_h\boldsymbol{p}_h \tag{1.1}$$

$$\boldsymbol{E}_{h-1}\boldsymbol{E}_{h-1}^{'}\boldsymbol{p}_h = \lambda_h\boldsymbol{t}_h \tag{1.2}$$

where $\lambda_h = (1/N)\boldsymbol{t}_h^{'}\boldsymbol{t}_h$ is the largest eigenvalue shared by $\boldsymbol{E}_{h-1}^{'}\boldsymbol{E}_{h-1}$ and $\boldsymbol{E}_{h-1}\boldsymbol{E}_{h-1}^{'}$, and $\boldsymbol{E}_1$ is the residual matrix of the regression of $\boldsymbol{X}$ on $\boldsymbol{t}_1$.

Since

$$\frac{1}{N}\boldsymbol{X}'\boldsymbol{X} = \frac{1}{N}(\boldsymbol{X} - \boldsymbol{t}_1\boldsymbol{p}_1')'(\boldsymbol{X} - \boldsymbol{t}_1\boldsymbol{p}_1') = \frac{1}{N}\boldsymbol{X}'\boldsymbol{X} - \lambda_1\boldsymbol{p}_1\boldsymbol{p}_1' \qquad (1.3)$$

for $h = 2$ the eigenvector $\boldsymbol{p}_2$ of $(1/N)\boldsymbol{E}_1'\boldsymbol{E}_1$ associated to the largest eigenvalue equals the eigenvector of $(1/N)\boldsymbol{X}'\boldsymbol{X}$ associated to the second largest eigenvalue.
In general,

$$\frac{1}{N}\boldsymbol{X}'\boldsymbol{X} = \frac{1}{N}\boldsymbol{X}'\boldsymbol{X} - \lambda_1\boldsymbol{p}_1\boldsymbol{p}_1' - \cdots - \lambda_{h-1}\boldsymbol{p}_{h-1}\boldsymbol{p}_{h-1}'. \qquad (1.4)$$

Hence, the eigenvector $\boldsymbol{p}_h$ of $(1/N)\boldsymbol{E}_{h-1}'\boldsymbol{E}_{h-1}$ associated to the largest eigenvalue equals the eigenvector of $(1/N)\boldsymbol{X}'\boldsymbol{X}$ associated to the h-th largest eigenvalue.

The main feature of NIPALS algorithm is that it works towards a suite of scalar products between pairs of vectors, i.e. (normalized) sums of products of element pairs. This feature allows us to easily handle missing data, by summing up in each operation only the available pairs (so-called element wise deletion procedure).
From the geometrical point of view, these scalar products can be interpreted as slopes of OLS regression lines. In particular, each value t_{1i} of $\boldsymbol{t}_1$ is the slope of the least-squares line without intercept going through the cloud of points $(\boldsymbol{p}_1; \boldsymbol{x}_i)$, where $\boldsymbol{x}_i$ is the transposed i-th row of $\boldsymbol{X}$. Similarly, each value p_{1p} is the slope of the least-squares line without intercept going through the cloud of points $(\boldsymbol{t}_1; \boldsymbol{x}_p)$, where $\boldsymbol{x}_p$ is the p-th column of $\boldsymbol{X}$. So, geometrically speaking, element wise deletion

procedure handles missing elements as if they lied on the regression line.

1.2.1 NIPALS and Power Method

The iterative sequence of NIPALS algorithm is very similar to well known Power Method [Frazer, Duncan & Collar 1938]. Power Method embraces a suite of algorithms, which generate a convergent sequence of vectors and compute the largest eigenvalue of a symmetric matrix by an iterative process. These methods use the relation that the eigenvalue of the p-th power of a matrix is the p-th power of the eigenvalue. Quoting Svante Wold, "*the difference is that the Power method applies to symmetrical matrices and is used to find the largest eigen-value and the corresponding eigen-vector of a symmetrical matrix. After "peeling" off the first eigen-vector, one can get the second, and then the third, etc.. The NIPALS method applies directly to the (scaled and centered) data matrix,* $\boldsymbol{X}$*, and hence is an SVD method*".

In effect, classic Power Method yields the largest eigenvalue of a squared matrix, say $\boldsymbol{S}$, by the following sequence:

$$\begin{aligned} &\boldsymbol{v}^{(0)} \\ &\boldsymbol{S}\boldsymbol{v}^{(0)} = \boldsymbol{v}^{(1)} \\ &\boldsymbol{S}\boldsymbol{v}^{(1)} = \boldsymbol{v}^{(2)} \\ &\vdots \end{aligned}$$

NIPALS algorithm, instead, uses a double chain iterative process

[Amato 1977], yielding the dominant singular value and associated right and left singular vectors of a rectangular matrix $\boldsymbol{X}$

$$\begin{aligned} &\boldsymbol{u}^{(0)} \\ &\boldsymbol{X}'\boldsymbol{u}^{(0)} = \boldsymbol{v}^{(0)} \; ; \; \boldsymbol{X}\boldsymbol{v}^{(0)} = \boldsymbol{u}^{(1)} \\ &\boldsymbol{X}'\boldsymbol{u}^{(1)} = \boldsymbol{v}^{(1)} \; ; \; \boldsymbol{X}\boldsymbol{v}^{(1)} = \boldsymbol{u}^{(2)} \\ &\vdots \end{aligned} \tag{1.5}$$

Since in our case $\boldsymbol{S} = \boldsymbol{X}'\boldsymbol{X}$, its eigenvalues equal the right singular vectors of $\boldsymbol{X}$, the two algorithms give the same (normalized) solution for $\boldsymbol{v}$.

In algorithm 2 a modified version of the original NIPALS algorithm is presented, where Steps 1.1 and 1.3 of classic NIPALS algorithm are joined in order to highlight similarities with Power Method. This algorithm consists of three steps. In the first, $\boldsymbol{p}_h$ is obtained as the eigenvector associated to the greatest eigenvalue of matrix $\boldsymbol{X}'\boldsymbol{X}$; in the second step $\boldsymbol{t}_h$ is obtained as a function of $\boldsymbol{p}_h$; finally, in the third step the matrix $\boldsymbol{E}_h$ is deflated.

1.3 PLS Regression

Partial Least Squares Regression (PLS-R) [Wold et al. 1983, Tenenhaus 1998] is a linear regression technique that allows relating a set of predictor variables to one or several response variables. At the same time, PLS-R decomposes the predictor matrix by sequentially extracting

Algorithm 2 The "Power Method type" NIPALS algorithm

Input: $\boldsymbol{E}_0 = \boldsymbol{X}$
Output: $\boldsymbol{P} = [\boldsymbol{p}_1, \ldots, \boldsymbol{p}_H], \boldsymbol{T} = [\boldsymbol{t}_1, \ldots, \boldsymbol{t}_H]$
  for all $h = 1, \ldots, H$ **do**
    Step 0: Initialize $\boldsymbol{p}_h$ such that $\|\boldsymbol{p}_h\| = 1$
    Step 1:
    repeat
      Step 1.1: $\boldsymbol{p}_h = \boldsymbol{E}_{h-1}^{'}\boldsymbol{E}_{h-1}\boldsymbol{p}_h$
      Step 1.2: $\boldsymbol{p}_h = \boldsymbol{p}_h/\|\boldsymbol{p}_h\|$
    until convergence of $\boldsymbol{p}_h$
    Step 2: $\boldsymbol{t}_h = \boldsymbol{E}_{h-1}\boldsymbol{p}_h$
    Step 3: $\boldsymbol{E}_h = \boldsymbol{E}_{h-1} - \boldsymbol{t}_h\boldsymbol{p}_h^{'}$
  end for

orthogonal components which at the same time summarize the explanatory variables and allow modelling and predicting the response variables.

PLS Regression has been shown greatly efficient in applications where data are characterized by many measured variables on few observations. This type of data generate three levels of problems: inferential, computational and descriptive. The inferential problem is due to the fact that large set of variables are always strongly correlated. Multicollinearity raises the variability of regression coefficient estimators, to the detriment of their significance. The computational problem is due to the rank of predictor matrix, leading to a singular correlation matrix. The descriptive problem regards the difficulty in analyzing at the same time relations among dozens or hundreds variables. PLS-R offers a solution for all of these drawbacks.

PLS-R can be included among regularized regression methods, as PLS estimators have be proved to be shrinkage estimators [De Jong 1995] (see section 1.3.6); moreover, PLS-R algorithm does not involve inversion of matrices: it consists of simple scalar products between pairs of vectors (see section 1.3.2). Finally, PLS is a factorial analysis method which resumes redundant information of predictor matrix in few orthogonal components (see section 1.3.1). This makes PLS-R a powerful visualization tool, because components compose a lower dimensional subspace in which information on predictor variables, useful to explain the responses, is resumed. By means of projections on the space spanned by the PLS components, it is possible to visualize non redundant information, eliminating noise.

Due to these features, PLS-R in last twenty years has bees used in a variety of fields. It has become a standard tool in chemometrics for multivariate calibration with chemical composition predicted from many high-speed but non-selective instrument measurements (e.g. NIR reflectance at different wavelengths) [Martens & Naes 1989]; it has been used extensively in quantitative structure-activity relationship (QSAR) research to relate descriptors of molecules and their biological activity [Hasegawa, Miyashita & Funatsu 1997]; in sensory science, it has been used for relating human sensory response to chemical, physical measurements and experimental design descriptors [Schulbach, Rouseff & Sims 2006]. Finally, PLS-R has been proposed in genetics for Quantitative Trait Loci (QTL) analysis, with the aim to predict phenotypic trait data from genetic markers [Bjørnstad, Westad &

Martens 2004].

From the algorithmic point of view, PLS Regression can be seen as an extension of NIPALS algorithm to the analysis of a cross-covariance matrix. Moreover, it can be considered as a slightly modified version of the two blocks PLS Path Modeling algorithm.

1.3.1 The Model

Let $\boldsymbol{x}_1 \dots \boldsymbol{x}_p \dots \boldsymbol{x}_P$ be a set of P predictor variables and $\boldsymbol{y}_1 \dots \boldsymbol{y}_r \dots \boldsymbol{y}_R$ be a set of R response variables measured on N observations. We suppose that all variables are centered.
PLS-R model assumes that there is a common structure underlying the two blocks of variable, and that this structure can be resumed by few latent components $\boldsymbol{t}_h$ $(h = 1 \dots H)$, calculated as a linear combination of the predictor variables. Predictor and response matrices $\boldsymbol{X}$ and $\boldsymbol{Y}$ are decomposed as

$$\begin{aligned} \boldsymbol{X} &= \boldsymbol{T}_H \boldsymbol{P}_H^{'} + \boldsymbol{E}_H \\ \boldsymbol{Y} &= \boldsymbol{T}_H \boldsymbol{C}_H^{'} + \boldsymbol{F}_H \end{aligned} \tag{1.6}$$

where $\boldsymbol{P}_H$ and $\boldsymbol{C}_H$ are the loading matrices, and $\boldsymbol{E}_H$ and $\boldsymbol{F}_H$ the residual matrices representing the part of variability in data due to noise.

Parameters of the model in are calculated by means of PLS Regression algorithm called also PLS2 in the multiple response case and PLS1 in the single response case [Tenenhaus 1998]. Since PLS1 is a

particular case of PLS2 (see section 1.3.2), this distinction is purely formal.

1.3.2 The algorithm

In PLS-R algorithm an iterative loop is used to calculate model parameters for each component. In any iteration $\boldsymbol{Y}$-scores, $\boldsymbol{X}$-weights, $\boldsymbol{X}$-scores and $\boldsymbol{Y}$-weights are sequentially calculated each one as a function of the previous one. All these steps can be interpreted as a sequence of bivariate regressions.

The loop for calculating parameters of the first order model starts choosing an initial value for the first component $\boldsymbol{u}_1$ in the $\boldsymbol{Y}$-space. Different options can be chosen for the initialization of $\boldsymbol{u}_h$: one of the response variables, the first principal component of response matrix among others. This choice, however, poorly affects the quickness of the convergence, which is always verified. In the following step an approximation for w_{1p}, element of the $\boldsymbol{X}$-weight vector $\boldsymbol{w}_1$, is obtained as a regression coefficient of $\boldsymbol{u}_1$ on $\boldsymbol{x}_p$. After having normalized $\boldsymbol{w}_1$, the $\boldsymbol{X}$-score t_{i1} is approximated by the regression coefficient of the i-th row of $\boldsymbol{X}$ on $\boldsymbol{w}_1$. Then the $\boldsymbol{Y}$-weight c_{r1} is computed as regression coefficient of $\boldsymbol{t}_1$ on $\boldsymbol{y}_r$. The loop is closed by approximating $\boldsymbol{X}$-score t_{i1} by the regression coefficient of the i-th row of $\boldsymbol{Y}$ on $\boldsymbol{c}_1$. These steps are repeated until convergence.

Once the convergence is obtained, $\boldsymbol{X}$ and $\boldsymbol{Y}$ are regressed on $\boldsymbol{t}_1$. The residual matrices of these regressions, respectively $\boldsymbol{E}_1$ and $\boldsymbol{F}_1$, are successively used for the computation of the second component $\boldsymbol{t}_2$.

Higher order components are similarly obtained.

The PLS-R algorithm, as presented in Tenenhaus [1998] is shown in algorithm 3.

Algorithm 3 Tenenhaus PLS-R algorithm

Input: $\boldsymbol{E}_0 = \boldsymbol{X}, \boldsymbol{F}_0 = \boldsymbol{Y}$
Output: $\boldsymbol{W}, \boldsymbol{C}, \boldsymbol{T}, \boldsymbol{U}, \boldsymbol{P}$
  for all $h = 1, \ldots, H$ **do**
    Step 0: Initialize $\boldsymbol{u}_h$
    Step 1:
    repeat
      Step 1.1: $\boldsymbol{w}_h = \boldsymbol{E}_{h-1}^{'}\boldsymbol{u}_h / \|\boldsymbol{E}_{h-1}^{'}\boldsymbol{u}_h\|$
      Step 1.2: $\boldsymbol{t}_h = \boldsymbol{E}_{h-1}\boldsymbol{w}_h / (\boldsymbol{w}_h^{'}\boldsymbol{w}_h)$
      Step 1.3: $\boldsymbol{c}_h = \boldsymbol{F}_{h-1}^{'}\boldsymbol{t}_h / (\boldsymbol{t}_h^{'}\boldsymbol{t}_h)$
      Step 1.4: $\boldsymbol{u}_h = \boldsymbol{F}_{h-1}\boldsymbol{c}_h / (\boldsymbol{c}_h^{'}\boldsymbol{c}_h)$
    until convergence of $\boldsymbol{w}_h$
    Step 2: $\boldsymbol{p}_h = \boldsymbol{E}_{h-1}^{'}\boldsymbol{t}_h / (\boldsymbol{t}_h^{'}\boldsymbol{t}_h)$
    Step 3: $\boldsymbol{E}_h = \boldsymbol{E}_{h-1} - \boldsymbol{t}_h\boldsymbol{p}_h^{'}$
    Step 4: $\boldsymbol{F}_h = \boldsymbol{F}_{h-1} - \boldsymbol{t}_h\boldsymbol{c}_h^{'}$
  end for

A slightly different version of PLS-R algorithm exists in literature [Höskuldsson 1988] (see algorithm 4), in which the vector of $\boldsymbol{Y}$-weights $\boldsymbol{c}$ is scaled to unitary norm. In this version, $\boldsymbol{Y}$-residuals are calculated as

$$\boldsymbol{F}_h = \boldsymbol{F}_{h-1} - b_{(\boldsymbol{u}_h|\boldsymbol{t}_h)}\boldsymbol{c}_h'\boldsymbol{t}_h$$

where $b_{(\boldsymbol{u}_h|\boldsymbol{t}_h)}$ is the OLS regression coefficient of $\boldsymbol{u}_h$ on $\boldsymbol{t}_h$.
This regression coefficient measures the so-called *inner relation* between the latent score vectors in the two spaces. The regression of $\boldsymbol{u}_h$

on $\boldsymbol{t}_h$ is an implicit step in PLS-R algorithm presented in Tenenhaus [1998], where $b_{(\boldsymbol{u}_h|\boldsymbol{t}_h)}$ equals the unity. In fact, if $\boldsymbol{c}_h$ is not normalized $\text{cov}(\boldsymbol{u}_h, \boldsymbol{t}_h) = \text{var}(\boldsymbol{t}_h)$.

In Step 3 of algorithm 4 the vector $\boldsymbol{q}_h$ of $\boldsymbol{Y}$-loadings is calculated. The calculation of $\boldsymbol{q}_h$ has no algorithmic relevance, but highlights similarities and dissimilarities in the treatment of $\boldsymbol{X}$ and $\boldsymbol{Y}$ in PLS-R: both are decomposed in own sets of weights ($\boldsymbol{w}$ and $\boldsymbol{c}$), loadings ($\boldsymbol{p}$ and $\boldsymbol{q}$) and components ($\boldsymbol{t}$ and $\boldsymbol{u}$); moreover, $\boldsymbol{X}$ is deflated as a function of own component $\boldsymbol{t}$, while $\boldsymbol{Y}$ is deflated as a function of the prediction of $\boldsymbol{u}$ by means of a linear function of $\boldsymbol{t}$.

Algorithm 4 Höskuldsson PLS-R algorithm

Input: $\boldsymbol{E}_0 = \boldsymbol{X}, \boldsymbol{F}_0 = \boldsymbol{Y}$
Output: $\boldsymbol{W}, \boldsymbol{C}, \boldsymbol{T}, \boldsymbol{U}, \boldsymbol{P}, \boldsymbol{Q}$
  for all $h = 1, \ldots, H$ **do**
    Step 0: Initialize $\boldsymbol{u}_h$
    Step 1:
    repeat
      Step 1.1: $\boldsymbol{w}_h = \boldsymbol{E}_{h-1}^{'}\boldsymbol{u}_h / \|\boldsymbol{E}_{h-1}^{'}\boldsymbol{u}_h\|$
      Step 1.2: $\boldsymbol{t}_h = \boldsymbol{E}_{h-1}\boldsymbol{w}_h / (\boldsymbol{w}_h^{'}\boldsymbol{w}_h)$
      Step 1.3: $\boldsymbol{c}_h = \boldsymbol{F}_{h-1}^{'}\boldsymbol{t}_h / \|\boldsymbol{F}_{h-1}^{'}\boldsymbol{t}_h\|$
      Step 1.4: $\boldsymbol{u}_h = \boldsymbol{F}_{h-1}\boldsymbol{c}_h / (\boldsymbol{c}_h^{'}\boldsymbol{c}_h)$
    until convergence of $\boldsymbol{w}_h$
    Step 2: $\boldsymbol{p}_h = \boldsymbol{E}_{h-1}^{'}\boldsymbol{t}_h / (\boldsymbol{t}_h^{'}\boldsymbol{t}_h)$
    Step 3: $\boldsymbol{q}_h = \boldsymbol{F}_{h-1}^{'}\boldsymbol{u}_h / (\boldsymbol{u}_h^{'}\boldsymbol{u}_h)$
    Step 4: $b_{(\boldsymbol{u}_h|\boldsymbol{t}_h)} = \boldsymbol{u}_h^{'}\boldsymbol{t}_h / (\boldsymbol{t}_h^{'}/\boldsymbol{t}_h)$
    Step 5: $\boldsymbol{E}_h = \boldsymbol{E}_{h-1} - \boldsymbol{t}_h\boldsymbol{p}_h^{'}$
    Step 6: $\boldsymbol{F}_h = \boldsymbol{F}_{h-1} - b_{(\boldsymbol{u}_h|\boldsymbol{t}_h)}\boldsymbol{t}_h\boldsymbol{c}_h^{'}$
  end for

It is possible to show Tenenhaus [1998] that deflation of $\boldsymbol{Y}$ is unnecessary when there are no missing data; parameters of the h-th component can be obtained running the iterative algorithm on $\boldsymbol{Y}$ and $\boldsymbol{E}_{h-1}$. This modification leads to the same vectors $\boldsymbol{w}_h$, $\boldsymbol{p}_h$ and $\boldsymbol{c}_h$ and yields more interpretable $\boldsymbol{Y}$-components, as they are functions of the original response variables.

Moreover, it is worth to notice that all mathematical operations in algorithms 3 and 4 imply just scalar products of pairs of vectors. This is the reason for which PLS-R (as NIPALS too) algorithm easily handle missing data. In fact, computational problems due to missing data can be avoided by means of pairwise deletion, that is calculating the sum of the product between couple of available data.

For a detailed review of the mathematical properties of PLS-R algorithm, refer to Tenenhaus [1998].

The single response case

In the single response case, it does not make sense calculating components in the unidimensional response space. For the h-th component, the weight w_p can be directly calculated as a function of $\boldsymbol{y}$. As a consequence, step 1.4 in algorithms 3 and 4 becomes trivial, as well the iterative loop, which stops itself in a single iteration. So, in the univariate case, PLS Regression algorithm can be oversimplified as shown in algorithm 5.

Though algorithm 5 is a particular case of algorithm 3, with which it shares all the properties, for historical reasons due to its extensive use

Algorithm 5 PLS1 algorithm

Input: $\boldsymbol{E}_0 = \boldsymbol{X}, \boldsymbol{f}_0 = \boldsymbol{y}$
Output: $\boldsymbol{W}, \boldsymbol{c}, \boldsymbol{T}, \boldsymbol{P}$

for all $h = 1, \ldots, H$ **do**
- **Step 1**: $\boldsymbol{w}_h = \boldsymbol{E}_{h-1}^{'} \boldsymbol{f}_{h-1} / \| \boldsymbol{E}_{h-1}^{'} \boldsymbol{f}_{h-1} \|$
- **Step 2**: $\boldsymbol{t}_h = \boldsymbol{E}_{h-1} \boldsymbol{w}_h / (\boldsymbol{w}_h^{'} \boldsymbol{w}_h)$
- **Step 3**: $c_h = \boldsymbol{f}_{h-1}^{'} \boldsymbol{t}_h / (\boldsymbol{t}_h^{'} \boldsymbol{t}_h)$
- **Step 4**: $\boldsymbol{p}_h = \boldsymbol{E}_{h-1}^{'} \boldsymbol{t}_h / (\boldsymbol{t}_h^{'} \boldsymbol{t}_h)$
- **Step 5**: $\boldsymbol{E}_h = \boldsymbol{E}_{h-1} - \boldsymbol{t}_h \boldsymbol{p}_h^{'}$
- **Step 6**: $\boldsymbol{f}_h = \boldsymbol{f}_{h-1} - c_h \boldsymbol{t}_h$

end for

as regularization technique (see section 1.3.6), it is often considered as an algorithm in itself, called PLS1 algorithm.

1.3.3 The algorithm as an extension of NIPALS

The link between PLS-R and NIPALS (and the double-chain algorithm 1.5) becomes clear if we work on the cross-covariance matrix $\boldsymbol{F}_{h-1}^{'} \boldsymbol{E}_{h-1}$ and we join steps 1.1 and 1.4 as well as steps 1.2 and 1.3 in algorithm 3, as shown in algorithm 6.

From this point of view, the only difference between PLS-R and NIPALS algorithms is in that PLS-R has not the double orthogonality property for which weights and loadings coincide; PLS-R loop for the computation of the h-th component, in fact, has as an output the vector $\boldsymbol{w}_h$, while loading vector $\boldsymbol{p}_h$ is calculated after the iteration as the regression coefficient of $\boldsymbol{E}_{h-1}^{'}$ on $\boldsymbol{t}_h$.

Algorithm 6 NIPALS type PLS-R algorithm

Input: $\boldsymbol{E}_0 = \boldsymbol{X}, \boldsymbol{F}_0 = \boldsymbol{Y}$
Output: $\boldsymbol{W}, \boldsymbol{C}, \boldsymbol{T}, \boldsymbol{U}$

for all $h = 1, \ldots, H$ **do**
Step 0: Initialize $\boldsymbol{c}_h$
Step 1:
repeat
Step 1.1: $\boldsymbol{w}_h = (\boldsymbol{F}'_{h-1}\boldsymbol{E}_{h-1})'\boldsymbol{c}_h / \|(\boldsymbol{F}'_{h-1}\boldsymbol{E}_{h-1})'\boldsymbol{c}_h\|$
Step 1.2: $\boldsymbol{c}_h = (\boldsymbol{F}'_{h-1}\boldsymbol{E}_{h-1})\boldsymbol{w}_h / (\boldsymbol{t}'_h\boldsymbol{t}_h)$
until convergence of $\boldsymbol{w}_h$
Step 2: $\boldsymbol{p}_h = \boldsymbol{E}'_{h-1}\boldsymbol{t}_h / (\boldsymbol{t}'_h\boldsymbol{t}_h)$
Step 3: $\boldsymbol{E}_h = \boldsymbol{E}_{h-1} - \boldsymbol{t}_h\boldsymbol{p}'_h$
Step 4: $\boldsymbol{F}_h = \boldsymbol{F}_{h-1} - \boldsymbol{t}_h\boldsymbol{c}'_h$
end for

1.3.4 Choosing the number of components

From the computational point of view, PLS-R algorithm can extract a number of components equal to the rank of $\boldsymbol{X}$. However, PLS Regression model supposes that the common information carried by $\boldsymbol{X}$ and $\boldsymbol{Y}$ matrices can be summarized in few latent components. So, a crucial issue in PLS-R model is the definition of the number H of components to retain.

In PLS Regression the explicative ability of the model (measured in terms of R^2 index) increases as long as the number of the components increases. On the contrary, the predictive ability of the model, intended as the explicative ability of the model referred to units that have not been considered in building the model (validation set), begins

to decrease after a certain number of components. This means that model overfits data, and we have to stop in extracting components.

A cross validation procedure is usually performed in order to evaluate if the h-th component increases the predictive ability of the model. The original sample is partitioned into S subsamples. For S times, a different subsample is retained as validation data and the remaining $(S-1)$ subsamples are used as training data. Each time, for each unit of the validation set, the squared prediction errors $e^2_{(-i)r}$ referred to $\boldsymbol{y}_r$ are calculated. For each h-component model, the PRediction Error Sum of Squares ($PRESS$) index is obtained as

$$PRESS_{rh} = \sum e^2_{(-i)r}.$$

Model over-fitting is investigated by plotting the $PRESS$ index against the number of components. Typically, the $PRESS$ decreases for a certain of components; then, it begins to increase: obviously, one choices the number of components giving the minimum $PRESS$.

In order to measure the marginal contribution of the h-th component to the predictive power of the model the Q^2 index [Ball 1963] is used

$$Q^2_h = 1 - \frac{\sum_{r=1}^{R} PRESS_{rh}}{\sum_{r=1}^{R} RESS_{r(h-1)}} \tag{1.7}$$

where $RESS_{hr}$ is the sum of the squared residuals of $\boldsymbol{y}_r$ in a $h-1$

component model on the whole data-set, with

$$RESS_{0r} = \sum_{i=1}^{N} (y_{ir} - \bar{y}_r)^2.$$

There are not *ad hoc* tests for assessing the significance of this index; in the practice, the h-th component is retained if $Q_h^2 \geq 0.0975$.

1.3.5 The optimizing criterion

Höskuldsson [1988] proved that the PLS iteration verifies the following equations for $\boldsymbol{c}_h$, $\boldsymbol{w}_h$, $\boldsymbol{t}_h$ and $\boldsymbol{u}_h$:

$$\begin{aligned}
(\boldsymbol{E}'_{h-1}\boldsymbol{F}_{h-1}\boldsymbol{F}'_{h-1}\boldsymbol{E}_{h-1})\boldsymbol{w}_h &= \lambda_h \boldsymbol{w}_h \\
(\boldsymbol{F}'_{h-1}\boldsymbol{E}_{h-1}\boldsymbol{E}'_{h-1}\boldsymbol{F}_{h-1})\boldsymbol{c}_h &= \lambda_h \boldsymbol{c}_h. \\
(\boldsymbol{E}_{h-1}\boldsymbol{E}'_{h-1}\boldsymbol{F}_{h-1}\boldsymbol{F}'_{h-1})\boldsymbol{t}_h &= \lambda_h \boldsymbol{t}_h \\
(\boldsymbol{F}_{h-1}\boldsymbol{F}'_{h-1}\boldsymbol{E}_{h-1}\boldsymbol{E}'_{h-1})\boldsymbol{u}_h &= \lambda_h \boldsymbol{u}_h
\end{aligned}$$

Hence, $\boldsymbol{w}_h$ and $\boldsymbol{c}_h$ are the normalized eigenvectors corresponding to the common largest eigenvalue λ_h of these symmetric squared matrices. PLS-R criterion directly descends as the maximization of

$$\mathrm{cov}^2(\boldsymbol{E}_{h-1}\boldsymbol{w}_h, \boldsymbol{F}_{h-1}\boldsymbol{c}_h) \tag{1.8}$$

Höskuldsson's work confirms the interpretation of the PLS iteration as the double-chain algorithm 1.5, as well the interpretation of $\boldsymbol{w}_h$ and $\boldsymbol{c}_h$ as the right and left dominant singular vectors of matrix $\boldsymbol{F}'_{h-1}\boldsymbol{E}_{h-1}$.

1.3.6 The regression equation

PLS Regression provides a classical regression equation, in which the response is estimated as a linear combination of the predictor variables. The following equation can be derived from the last step of algorithm 3:

$$\boldsymbol{Y} = \boldsymbol{t}_1\boldsymbol{c'}_1 + \boldsymbol{t}_2\boldsymbol{c'}_2 +, \ldots, \boldsymbol{t}_H\boldsymbol{c'}_H + \boldsymbol{F}_H = \boldsymbol{T}_H\boldsymbol{C'}_H + \boldsymbol{F}_H. \tag{1.9}$$

This is the regression equation of a H-component PLS-R model, where response variables are expressed as a function of the PLS components.

In PLS-R algorithm each $\boldsymbol{t}_h$ is calculated as a function of $\boldsymbol{E}_{h-1}$

$$\boldsymbol{t}_h = \boldsymbol{E}_{h-1}\boldsymbol{w}_h. \tag{1.10}$$

In a model with H components the matrix $\boldsymbol{T}_H$ of the $\boldsymbol{X}$-score factors is

$$\boldsymbol{T}_H = [\boldsymbol{X}\boldsymbol{w}_1, \boldsymbol{E}_1\boldsymbol{w}_2, \ldots, \boldsymbol{E}_{H-1}\boldsymbol{w}_H] \tag{1.11}$$

$\boldsymbol{T}_H$ can be obtained also as a function of the original $\boldsymbol{X}$ variables

$$\boldsymbol{T}_H = \boldsymbol{X}\boldsymbol{W}_H(\boldsymbol{P'}_H\boldsymbol{W}_H)^{-1} \tag{1.12}$$

Posing $\boldsymbol{R}_H = \boldsymbol{W}_H(\boldsymbol{P'}_H\boldsymbol{W}_H)^{-1}$ we obtain the responses as a linear function of the predictor variables

$$\boldsymbol{Y} = \boldsymbol{T}_H\boldsymbol{C'}_H + \boldsymbol{F}_H = \boldsymbol{X}\boldsymbol{R}_H\boldsymbol{C'}_H + \boldsymbol{Y}_H = \boldsymbol{X}\boldsymbol{B}_H^{\mathrm{PLS}} + \boldsymbol{F}_H \tag{1.13}$$

where $\boldsymbol{B}_H^{\text{PLS}}$ is the matrix of the coefficients of a H-component PLS regression model.

PLS-R coefficient estimators have been proved to be shrinkage estimators by De Jong [1995], which showed that the sequence of PLS coefficient estimators forms a suite of vectors whose length strictly increases with the number of components. The upper bound of this suite is the length of Least Squares estimator. As shown by Frank & Friedman [1993] and Garthwaite [1994], this feature makes PLS-R a valid regularization tool, whose validity is comparable to Ridge Regression [Hoerl & Kennard 1970] and higher than Principal Component Regression [Jolliffe 1982].

As all the shrinkage estimators, also PLS estimators are biased; however, they have been shown to be asymptotically unbiased.
Regression coefficients of an A-component PLS-R univariate model equal the Least Squares regression coefficients. Hence, in PLS-R solution can be interpreted as a trade off between explicative power of regression and stability of its parameters. The more multicollinearity among predictors increases, the more this trade becomes convenient.

1.3.7 Outlier detection

Distances of each observation from the model can be calculated in order to detect outliers. The distance of the i-th unit is calculated in

$\boldsymbol{X}$ and $\boldsymbol{Y}$ spaces, respectively as

$$DModX_i = \sqrt{\frac{\sum_{p=1}^{P} e_{ip}^2}{P-H}} \times \sqrt{\frac{N}{N-H-1}} \tag{1.14}$$

and

$$DModY_i = \sqrt{\frac{\sum_{r=1}^{R} f_{ir}^2}{R-H}} \tag{1.15}$$

where $e_{ip} = (x_{ip} - \hat{x}_{ip})$ and $f_{ip} = (y_{ip} - \hat{y}_{ip})$.

To compare distances of different units from the model, these indexes are normalizated with respect to the respective standard deviations s_X and s_Y

$$DModX_i, N_i = \frac{DModX_i}{s_X} \tag{1.16}$$

and

$$DModY_i N_i = \frac{DModY_i}{s_Y}, \tag{1.17}$$

where

$$s_x = \sqrt{\frac{\sum_{i=1}^{N} \sum_{p=1}^{P} e_{ip}^2}{(N-H-1)(P-H)}} \tag{1.18}$$

and

$$s_y = \sqrt{\frac{\sum_{i=1}^{N}\sum_{r=1}^{r} f_{ir}^2}{(N-H-1)(R-H)}} \tag{1.19}$$

Under the hypothesis that $DModX_i, N_i$ and $DModY_i, N_i$ follow a Fischer-Snedecor distribution, it is possible to calculate a critical threshold for assessing the significance of these distances. Howeveer, since this hypothesis has not a theoretical fundament, this threshold have to be considered purely empirical.

1.3.8 VIP index

The explicative power of a predictor with respect to the whole set of responses is measured by the VIP (Variable Importance in the Projection) index [Wold 2009]. The VIP index is a normalized average of the explicative power of the components (measured in terms of redundancy) weighted by the contributions of $\boldsymbol{x}_p$ to the construction of the components (measured by the squares of the weights w_{hp}). It is calculated as follows:

$$VIP_p = \sqrt{\frac{\sum_{i=1}^{H} Red(\boldsymbol{Y}, \boldsymbol{t}_h) w_{hp}^2}{(1/p) \times \sum_{i=1}^{H} Red(\boldsymbol{Y}, \boldsymbol{t}_h)}}, \tag{1.20}$$

where $Red(\boldsymbol{Y}, \boldsymbol{t}_h) = \sum_{r=1}^{R} \mathrm{cor}^2(\boldsymbol{y}_r, \boldsymbol{t}_h)$.

Since $\sum_p VIP_p^2 = P$, greater than one rule is generally used as a criterion for variable selection. However, Wold [2009] suggests to

consider important a variable with a $VIP > 0.8$.

1.3.9 PLS Regression as a visualization technique

PLS-R allows visualizing information on bidimensional subspaces generated by pairs of components. Some of the plots are common to all the factorial methods for multidimensional data analysis, others are specific PLS-R outputs.

Plots common to the other component based methods are the observation plot and the correlation circle. Mapping the observation on the plot $(\boldsymbol{t}_h, \boldsymbol{t}_{h'})$, it is possible to investigate similarities and differences in unit behaviors by simply looking at their euclidean distances. Moreover, correlations of $\boldsymbol{X}$ and $\boldsymbol{Y}$ variables with the components can be mapped in the correlation circle as points or as arrows joining each point to the origin.

Other graphics are typical PLS-R outputs. For example, relations between component in different space can be investigated by means of the plot $(\boldsymbol{t}_h, \boldsymbol{u}_h)$. Another standard plot in PLS-R framework is obtained by overlapping points $(\boldsymbol{r}_h, \boldsymbol{r}_{h'})$ and $(\boldsymbol{c}_h, \boldsymbol{c}_{h'})$. This representation allows us to interpret the contribution of the predictors in building components $\boldsymbol{t}_h$, as well the component capability in explicating $\boldsymbol{Y}$-variables. This plot, referred to the first two components, has another interesting interpretation. Any regression coefficient $b^{\text{PLS}}_{\boldsymbol{y}_r|\boldsymbol{x}_p}$ of $\boldsymbol{x}_p$ on $\boldsymbol{y}_r$ in the two-component model, in fact, can be read on the plot as the orthogonal projection of the point (r_{p1}, r_{p2}) on the straight

line passing for the origin and the point (c_{r1}, r_{p2}). Hence, the matrix of regression coefficients, which is usually hard to read due to the large number of variables in PLS models, can be visualized and easily interpreted on this plot.

1.3.10 Alternative PLS approaches to cross covariance analysis

PLS loop can be used for the cross-covariance analysis of two block of matrices in several methods, depending on the way we deflate $\boldsymbol{X}$ and $\boldsymbol{Y}$. All of these methods provide the same pair of score vectors $\boldsymbol{t}_1$ and $\boldsymbol{u}_1$ while differ in higher order components.

Since the aim of PLS Regression is prediction, in PLS-R the deflation is obtained by regressing both $\boldsymbol{X}$ and $\boldsymbol{Y}$ on $\boldsymbol{t}_1$. As a matter of fact, this was the genial intuition which allowed Svante Wold and Harald Martens to transform the two block PLS-PM in a powerful tool for regularization of the OLS regression, giving up the orthogonality of $\boldsymbol{Y}$-space score vectors. In Herman Wold's original approach to deflation, instead, each matrix is deflated by means of its own components and loadings in the following way:

$$\boldsymbol{E}_h = \boldsymbol{E}_{h-1} - \boldsymbol{t}_h \boldsymbol{p}_h'$$
$$\boldsymbol{F}_h = \boldsymbol{F}_{h-1} - \boldsymbol{u}_h \boldsymbol{q}_h'$$

that is, by subtracting to $\boldsymbol{E}_h$ and $\boldsymbol{F}_h$ their own best rank-1 approximation in the Least Squares sense. This approach leads to a maximum

number of $min(P, R)$ pair of components such that matrices $\boldsymbol{T}'\boldsymbol{T}$ and $\boldsymbol{U}'\boldsymbol{U}$ are diagonal.

A last way for implementing two blocks PLS is a method known with a number of different names in American literature: PLS-SB [Sampson, Streissguth, Barr & Bookstein 1989, Rosipal & Krämer 2006], PLS-SVD [Wegelin 2000], Intercorrelation Analysis, Canonical Covariance [Tishler, Dvir, Shenhar & Lipovetsky 1996], and so on. We will call it Tucker-PLS. Tucker-PLS works on successive deflations of the cross-covariance matrix $\boldsymbol{X}'\boldsymbol{Y}$. For the h-th component, residual matrix $\boldsymbol{Z}_h$ (with $\boldsymbol{Z}_0 = \boldsymbol{X}'\boldsymbol{Y}$) is calculated as

$$\boldsymbol{Z}_h = \boldsymbol{Z}_{h-1} - \lambda_h \boldsymbol{w}_h \boldsymbol{c}_h'$$

In Tucker-PLS, if we deal with full rank matrices, the maximum number of extractable components is $min(P, R, \text{rank}(\boldsymbol{X}'\boldsymbol{Y}))$. Neither $\boldsymbol{X}$ nor $\boldsymbol{Y}$ factor scores are orthogonal, but the matrix $\boldsymbol{T}'\boldsymbol{U}$ is diagonal. Since this approach leads to the singular value decomposition of $\boldsymbol{X}'\boldsymbol{Y}$, it is the sam that Tucker's Inter-Battery Analysis [Tucker 1958] implemented by PLS loop: from here the name.

Among these methods, PLS-R is the only one prediction oriented. The others can be useful to investigate relations between the two blocks, but they do not suppose cause-effect relations because of their symmetry in decomposition of the matrices.

1.4 PLS approach to Structural Equation Modeling

PLS Path Modeling [Wold 1975a, Wold 1982, Tenenhaus, Esposito Vinzi, Chatelin & Lauro 2005] aims to estimate the relationships among Q blocks $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_q, \ldots, \boldsymbol{X}_Q$ of manifest variables (MVs), which are expression of Q unobservable constructs $\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_q, \ldots, \boldsymbol{\xi}_Q$, that are usually called latent variables (LVs). Specifically, PLS-PM estimates the network of relations among the manifest variables and their own latent variables, and among the latent variables inside the model, through a system of interdependent equations based on simple and multiple regressions. The corresponding conceptual model can be represented through path diagrams (see figure 1.1), adhering to certain common drawing conventions. Specifically, ellipses or circles represent LVs and rectangles or squares refer to the MVs. Arrows showing causations among the variables (either latent or manifest), and the direction of the array defines the direction of the relation, *i.e.* variables receiving the array have to be considered as endogenous variables in the specific relationship.

1.4.1 The predictive Path Model

PLS Path Model consists of two sub models: the Structural (or Inner) Model and the Measurement (or Outer) Model (see figure 1.2).

The structural model specifies the relationships between the LVs; a LV is called exogenous, if it is supposed to depend on other LVs

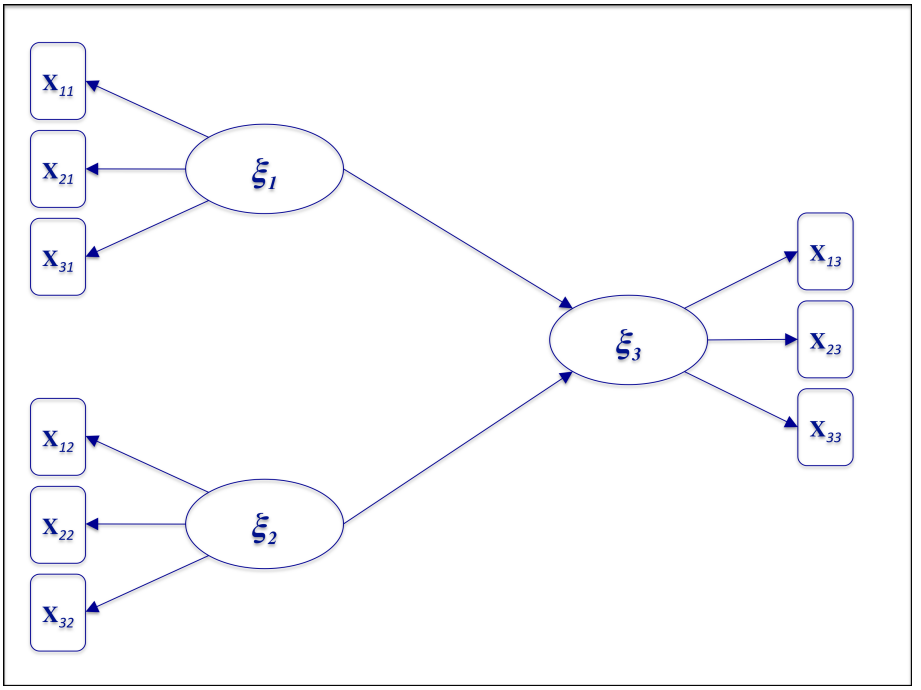

Figure 1.1: *An example of Path Diagram*

and endogenous otherwise. Structural relationships can be taken in account by means of a lower triangular matrix $\boldsymbol{L}$ of order Q. The element $l_{qq'}$ is filled with a 1 if $\boldsymbol{\xi}_q$ depends on $\boldsymbol{\xi}_{q'}$ and 0 otherwise. In the structural model each endogenous LV is linked to the other LVs

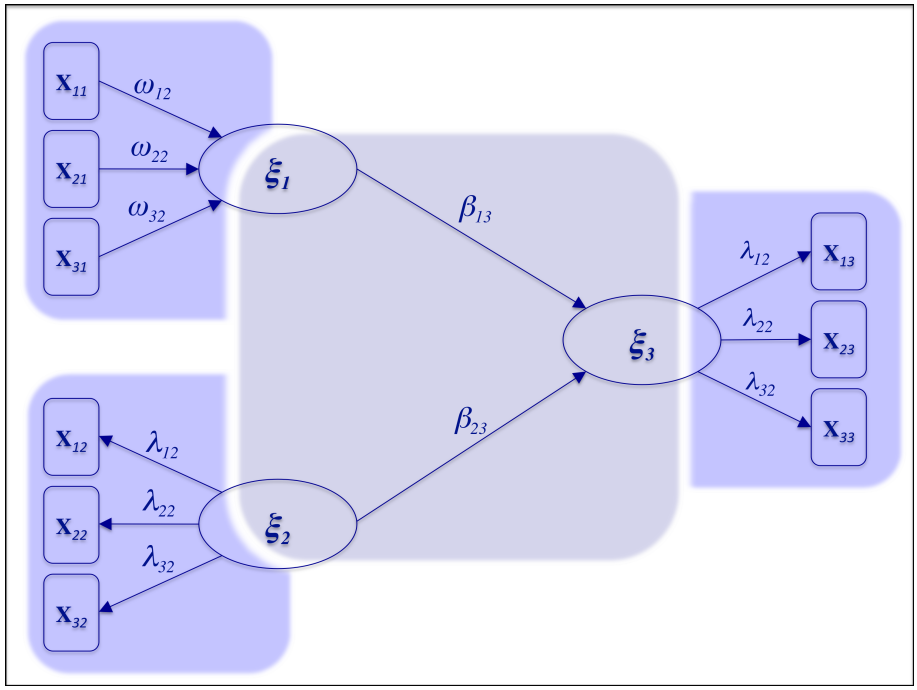

Figure 1.2: *PLS Path Model representation. The inner model is painted in blue grey, the outer model in sky blue*

by the following multiple regression model

$$\boldsymbol{\xi}_q = \beta_{q0} + \sum_{q=1}^{Q} l_{qq'}\beta_{qq'}\boldsymbol{\xi}_{q'} + \boldsymbol{\nu}_q \tag{1.21}$$

where $\beta_{qq'}$ is the so-called path coefficient expressing the impact on

the endogenous LV $\boldsymbol{\xi}_q$ of the connected exogenous LVs. The only hypothesis of this model is what Wold named *prediction specification hypothesis* [Wold 1982]: residual vector $\boldsymbol{\nu}_q$ has zero mean and is not correlated with predictors.

The measurement model relates the MVs to their own LV. A block is defined *reflective* if the LV is assumed to be a common factor that reflects itself in the MVs. This implies that the relation between each MV $\boldsymbol{x}_{pq}$ $(p = 1, \ldots, Pq)$ and the corresponding LV is modeled as

$$\boldsymbol{x}_{pq} = \lambda_{pq0} + \lambda_{pq}\boldsymbol{\xi}_q + \boldsymbol{\epsilon}_{pq} \tag{1.22}$$

Also in this model the *prediction specification hypothesis* is required. In the reflective case, MVs should be highly correlated, due to the fact that they are correlated with the LV of which they are expression. In other words, the block has to be homogeneous. There are four empirical rules for assessing the homogeneity of a reflective block. All of these rules assume, without loss of generality, that LVs are standardized and all correlations between MVs of the block show the same sign. When this last hypothesis is not verified, it is possible just to state that block is unidimensional, but not homogeneous.

a) *Principal component analysis rule*: a block is considered homogeneous if the first eigenvalue of the correlation matrix is higher than 1, while the others are smaller;

b) *Cronbach's alpha rule*: a block is considered homogeneous if this

index is larger than 0.7

$$\alpha = \frac{\sum_{p \neq p'} \text{cor}(\boldsymbol{x}_{pq}, \boldsymbol{x}_{p'q})}{P_q + \sum_{p \neq p'} \text{cor}(\boldsymbol{x}_{pq} \boldsymbol{x}_{p'q})} \times \frac{P_q}{P_q - 1}, \tag{1.23}$$

where $\boldsymbol{x}_{pq}$ and $\boldsymbol{x}_{pq'}$ are two MVs of the q-th block;

c) *Dillon-Goldstein's rho rule* (or *Jöreskog's*): it measure the composite reliability of the block. A block is considered homogeneous if its composite reliability is larger than 0.7

$$\rho_{DG} = \frac{(\sum_{p=1}^{P_q} \hat{\lambda}_{pq}^{\text{PCA}})^2}{(\sum_{p=1}^{P_q} \hat{\lambda}_{pq}^{\text{PCA}})^2 + \sum_{p=1}^{P_q} (1 - (\hat{\lambda}_{pq}^{\text{PCA}})^2)}, \tag{1.24}$$

where $\hat{\lambda}_{pq}^{\text{PCA}}$ is the estimate of standardized loading λ_{pq}, obtained *ex-ante* as the loading $\hat{\lambda}_{pq}^{\text{PCA}} = \text{cor}(\boldsymbol{x}_{pq}, \boldsymbol{t}_q)$ of the first principal component $\boldsymbol{t}_q$ of $\boldsymbol{X}_q$. For an analysis *ex-post*, PLS-PM estimate $\hat{\lambda}_{pq} = \text{cor}(\boldsymbol{x}_{pq}, \widehat{\boldsymbol{\xi}}_q)$ replaces $\hat{\lambda}_{pq}^{\text{PCA}}$, where $\widehat{\boldsymbol{\xi}}_q$ is PLS-PM estimate of $\boldsymbol{\xi}_q$.

According to Chin [1998] the *Dillon-Goldstein's rho* is considered to be a better indicator of the homogeneity of a block than the *Cronbach's alpha*;

d) *Confirmatory TETRAD Analysis* [Bollen & Ting 1993, Gudergan, Ringle, Wende & Will 2008]: A TETRAD is the difference of the products of two pairs of covariances between MVs of the block. All non redundant TETRADs are tested be different from zero with a Bootstrap-based test. If all null hypotheses are ac-

cepted, block is assumed to be reflective; in the opposite case, block is considered formative.

In the *formative* case, each manifest variable or each sub-block of MVs represents different dimensions of the underlying concept: in other words they are in a strictly causative relation with the LV, and this relation is modeled as

$$\boldsymbol{\xi}_q = \sum_{p=1}^{P_q} \omega_{pq} \boldsymbol{x}_{pq} + \boldsymbol{\delta}_{pq} \tag{1.25}$$

where the error term $\boldsymbol{\delta}_{pq}$ is that part of $\boldsymbol{\xi}_q$ variability not accounted by the MVs which is supposed to satisfy the *prediction specification hypothesis*.

Finally, a block can be composed of both reflective and formative MVs: this is the MIMIC (multiple effect indicators for multiple causes) case.

1.4.2 PLS Path Modeling algorithm

In PLS path model external weights ω_{pq}, linking each MV to corresponding LV, are estimated by an iterative procedure in which the latent variable scores are obtained through the alternation of *outer* ($\boldsymbol{v}_q$) and *inner* ($\boldsymbol{z}_q$) estimations of the LVs. This procedure is referred to as PLS Path Modeling (PLS-PM) algorithm.

No formal proof of convergence of the general algorithm has been provided until now, but in number of particular cases the PLS-PM

loop is proven to monotonicallly converge versus a criterion (see section 1.4.5). However, convergence is always assured in practice.

In the next we suppose, without loss of generality, that each $\boldsymbol{x}_{pq}$ is centered. The procedure starts by choosing arbitrary weights w_{pq}. Then, each LV is calculated as a linear combination of its own centered MVs (outer estimation)

$$\boldsymbol{v}_q \propto \sum_{p=1}^{P_q} w_{pq}\boldsymbol{x}_{pq} = \boldsymbol{X}_q\boldsymbol{w}_q \tag{1.26}$$

The symbol $\propto$ means that the left side of the equation corresponds to the normalized ($\boldsymbol{v'}_q\boldsymbol{v}_q = N$) right side.

In the inner estimation, each LV is obtaines as a normalized linear combination of the outer estimations of the connected LVs. Weights e of this linear combination are called inner weights. Let $c_{qq'}$ be the generic element of the square matrix $\boldsymbol{C} = \boldsymbol{L} + \boldsymbol{L}'$ of order Q. $c_{qq'} = 1$ if $\boldsymbol{\xi}_q$ is connected to $\boldsymbol{\xi}_{q'}$ in path diagram and $c_{qq'} = 0$ otherwise; then, the inner estimation $\boldsymbol{z}_q$ can be expressed as

$$\boldsymbol{z}_q \propto \sum_{q'=1}^{Q} c_{qq'}e_{qq'}\boldsymbol{v}_q \tag{1.27}$$

There are three way for calculating the inner weights ($e_{qq'}$):

1. the centroid scheme (the Wold's original scheme), where $e_{qq'}$ is equal to sign of the correlation between $\boldsymbol{v}_q$ and $\boldsymbol{v}_{q'}$;

2. the factorial scheme (the Löhmoller scheme), where $e_{qq'}$ is equal

to correlation between $\boldsymbol{v}_q$ and $\boldsymbol{v}_{q'}$;

3. the path weighting scheme, where, for each $\boldsymbol{\xi}_q$: if $\boldsymbol{\xi}_{q'}$ is a latent predictor of $\boldsymbol{\xi}_q$, $e_{qq'}$ is equal to $\boldsymbol{v}_{q'}$ coefficient in the multiple regression of $\boldsymbol{v}_q$ on the inner estimations of its latent predictors; if $\boldsymbol{\xi}_{q'}$ is a latent response variable of $\boldsymbol{\xi}_q$, $e_{qq'}$ is equal to correlation between $\boldsymbol{v}_q$ and $\boldsymbol{v}_{q'}$.

Once a first inner estimation of the latent variables is obtained, the algorithm goes on by updating the outer weights w_{pq}.
Two different ways are available to update the outer weights:

- *Mode A*: each outer weight w_{pq} is the regression coefficient in the simple regression of the p-th manifest variable of the q-th block ($\boldsymbol{x}_{pq}$) on the inner estimate $\boldsymbol{z}_q$ of the q-th latent variable. As a matter of fact, since $\boldsymbol{z}_{pq}$ is standardized, the generic outer weight w_{pq} is obtained as

$$w_{pq} = \operatorname{cov}(\boldsymbol{x}_{pq}, \boldsymbol{z}_q) \tag{1.28}$$

- *Mode B*: the vector $\boldsymbol{w}_q$ of the weights w_{pq} associated to the manifest variables of the q-th block is the regression coefficient vector in the multiple regression of the inner estimate $\boldsymbol{z}_q$ of the q-th latent variable on its manifest variables $\boldsymbol{X}_q$:

$$\boldsymbol{w}_q = \left(\boldsymbol{X}_q'\boldsymbol{X}_q\right)^{-1}\boldsymbol{X}_q'\boldsymbol{z}_q \tag{1.29}$$

The choice of the external weight estimation mode is strictly related

to the nature of the model. For a *reflective model* the *Mode A* is more appropriate, while *Mode B* is better for the *formative model*. Furthermore, *Mode A* is suggested for endogenous latent variables, while *Mode B* for the exogenous ones.

In a completely data-driven approach, a further alternative to update outer weights is *Mode PLS* [Esposito Vinzi, Trinchera & Amato 2009, Esposito Vinzi & Russolillo 2010]. In *Mode PLS* $\boldsymbol{w}_q$ is the regression coefficient vector in a PLS Regression of $\boldsymbol{z}_q$ on $\boldsymbol{X}_q$. If PLS-PM algorithm converges on a single-component PLS-R, then *Mode PLS* weights will equal *Mode A* weights: data are definitively expression of a reflective model. The case where PLS-PM algorithm converges on a several-component PLS-R, have to be interpreted in a formative sense: each sub-block of MVs represents different dimensions of the concept underlying the LV. Whereas PLS-PM algorithm converges on a P_q-component PLS-R, *Mode PLS* weights will equal *Mode B* weights: each MV represents a different dimension of the concept underlying the LV.

Inner and outer estimation steps are alternated till convergence on the weights.

Once final weights w_{pq} are obtained, the LVs scores are finally calculated as normalized weighted aggregates of the manifest variables

$$\widehat{\boldsymbol{\xi}}_q \propto \boldsymbol{X}_q \boldsymbol{w}_q \tag{1.30}$$

In the last step of PLS-PM algorithm structural (or path) coefficients are estimated through an OLS multiple regression among the

estimated latent variable scores, according to path diagram structure. Denoting $\boldsymbol{\xi}_j$ $(j = 1 \dots J)$ the generic endogenous LV and $\boldsymbol{\Xi}_{\rightarrow j}$ the matrix of the corresponding latent predictors, the path coefficient vector for each $\boldsymbol{\xi}_j$ is

$$\boldsymbol{\beta}_j = \left(\widehat{\boldsymbol{\Xi}'}_{\rightarrow j}\widehat{\boldsymbol{\Xi}}_{\rightarrow j}\right)^{-1}\widehat{\boldsymbol{\Xi}'}_{\rightarrow j}\widehat{\boldsymbol{\xi}}_j \tag{1.31}$$

In the case of multicollinearity among the estimated latent variable scores, in order to reduce estimation variability, PLS regression can be used instead of OLS regression [Esposito Vinzi et al. 2009].

Wold's original algorithm has been further developed by Lohmöller [Lohmöller 1987, Lohmöller 1989]. In Lohomöller's algorithm, inner estimation $\boldsymbol{z}_q^{(s)}$ in the s-th iteration is always a function of $\boldsymbol{v}_{q'}^{(s)}$. In the original Wold's algorithm, instead, $\boldsymbol{z}_q^{(s)}$ is a function of $\boldsymbol{v}_{q'}^{(s+1)}$ if $q' < q$ and is a function of $\boldsymbol{v}_{q'}^{(s)}$ if $q' > q$. As showed by Hanafi [2007], Wold's algorithm, if estimation *mode B* is used, converges in a monotonic (and consequently faster) way. Here, both the Lohmöller's and Wold's iterative procedures are presented respectively in algorithms 7 and 8.

1.4.3 Model assessment

PLS-PM aims to find LVs that at the same time summarize well their own block and are correlated between them (following the path diagram). So, the quality of a PLS Path Model depends on the quality

Algorithm 7 Lohmöller's PLS Path Modeling algorithm

Input: $\boldsymbol{X} = [\boldsymbol{X}_1, \ldots, \boldsymbol{X}_q \ldots, \boldsymbol{X}_Q], \boldsymbol{C}$

Output: $\boldsymbol{\beta}_j$, $\boldsymbol{w}_q$, $\widehat{\boldsymbol{\xi}}_q$;

Step 0: Initialization

$\boldsymbol{w}_q = \boldsymbol{w}_q^{(0)}$

Step 1: Iteration

repeat

Step 1.1: Outer estimation of the LVs

$\boldsymbol{v}_q^{(s)} \propto \sum_{p=1}^{P_q} w_{pq}^{(s)} \boldsymbol{x}_{pq} = \boldsymbol{X}_q \boldsymbol{w}_q^{(s)}$

Step 1.2: Computation of the inner weights

$e_{qq'}^{(s)} = f\left(\boldsymbol{v}_q^{(s)}, \boldsymbol{v}_{q'}^{(s)}\right)$, according to the chosen scheme

Step 1.3: Inner estimationof the LVs

$\boldsymbol{z}_q^{(s)} \propto \sum_{q'=1}^{Q} c_{qq'} e_{qq'}^{(s)} \boldsymbol{v}_{q'}^{(s)}$

Step 1.4: Computation of the outer weights

$\boldsymbol{w}_q^{(s+1)} = (1/N) \boldsymbol{X}_q^{'} \boldsymbol{z}_q^{(s)}$ (Mode A) or

$\boldsymbol{w}_q^{(s+1)} = (\boldsymbol{X}_q^{'} \boldsymbol{X}_q)^{-1} \boldsymbol{X}_q^{'} \boldsymbol{z}_q^{(s)}$ (Mode B)

until convergence of $\boldsymbol{w}_q$

Step 2: Computation of the LVs

$\widehat{\boldsymbol{\xi}}_q \propto \boldsymbol{X}_q \boldsymbol{w}_q$

Step 3: Computation of the Path Coefficients

$\boldsymbol{\beta}_j = \left(\widehat{\boldsymbol{\Xi}'}_{\rightarrow j} \widehat{\boldsymbol{\Xi}}_{\rightarrow j}\right)^{-1} \widehat{\boldsymbol{\Xi}'}_{\rightarrow j} \widehat{\boldsymbol{\xi}}_j$

Algorithm 8 Wold's PLS Path Modeling algorithm

Input: $\boldsymbol{X} = [\boldsymbol{X}_1, \ldots, \boldsymbol{X}_q \ldots, \boldsymbol{X}_Q], \boldsymbol{C}$

Output: $\boldsymbol{\beta}_j$, $\boldsymbol{w}_q$, $\widehat{\boldsymbol{\xi}}_q$;

Step 0: Initialization

Step 0.1: $\boldsymbol{w}_q = \boldsymbol{w}_q^{(0)}$

Step 0.2: $\boldsymbol{v}_q^{(0)} \propto \sum_{p=1}^{P_q} w_{pq}^{(0)} \boldsymbol{x}_{pq} = \boldsymbol{X}_q \boldsymbol{w}_q^{(0)}$

Step 1: Iteration

repeat

Step 1.1: Computation of the inner weights

if $q' < q$ **then** $e_{qq'}^{(s)} = f\left(\boldsymbol{v}_q^{(s)}, \boldsymbol{X}_q \boldsymbol{w}_{q'}^{(s+1)}\right)$

end if

if $q' > q$ **then** $e_{qq'}^{(s)} = f\left(\boldsymbol{v}_q^{(s)}, \boldsymbol{v}_{q'}^{(s)}\right)$

end if

Step 1.2: Inner estimation of the LVs

$\boldsymbol{z}_q \propto (\sum_{q'<q} c_{qq'} e_{qq'}^{(s)} \boldsymbol{X}_q \boldsymbol{w}_{q'}^{(s+1)} + \sum_{q'>q} c_{qq'} e_{qq'}^{(s)} \boldsymbol{v}_{q'}^{(s)})$

Step 1.3: Computation of the outer weights

$\boldsymbol{w}_q^{(s+1)} = (1/N) \boldsymbol{X}_q' \boldsymbol{z}_q^{(s)}$ (Mode A) or

$\boldsymbol{w}_q^{(s+1)} = (\boldsymbol{X}_q' \boldsymbol{X}_q)^{-1} \boldsymbol{X}_q' \boldsymbol{z}_q^{(s)}$ (Mode B)

Step 1.4: Outer estimation of the LVs

$\boldsymbol{v}_q^{(s+1)} \propto \sum_{p=1}^{P_q} w_{pq}^{(s+1)} \boldsymbol{x}_{pq} = \boldsymbol{X}_q \boldsymbol{w}_q^{(s+1)}$

until convergence of $\boldsymbol{w}_q$

Step 2: Computation of the LVs

$\widehat{\boldsymbol{\xi}}_q \propto \boldsymbol{X}_q \boldsymbol{w}_q$

Step 3: Computation of the Path Coefficients

$\boldsymbol{\beta}_j = \left(\widehat{\boldsymbol{\Xi}'}_{\rightarrow j} \widehat{\boldsymbol{\Xi}}_{\rightarrow j}\right)^{-1} \widehat{\boldsymbol{\Xi}'}_{\rightarrow j} \widehat{\boldsymbol{\xi}}_j$

of both measurement and structural model.
In a good measurement model each MV is well summarized by its own LV. So, for each block, a Communality index is computed as

$$Com_q = \frac{1}{P_q} \sum_{p=1}^{P_q} \text{cor}^2 \left(\boldsymbol{x}_{pq}, \widehat{\boldsymbol{\xi}}_q \right) = \frac{1}{P_q} \sum_{p=1}^{P_q} \widehat{\lambda}_{pq}^2 \tag{1.32}$$

that is the average of the communalities between each MV of the q-th block and $\widehat{\boldsymbol{\xi}}_q$. A similar index is the Average Variance Extracted (AVE) [Fornell & Larcker 1981], that express the part of variance of the block explained by $\widehat{\boldsymbol{\xi}}_q$

$$AVE_q = \frac{\sum_{p=1}^{P_q} \widehat{\lambda}_{pq}^2}{\sum_{p=1}^{P_q} \text{var}(\boldsymbol{x}_{pq})} \tag{1.33}$$

If we work on standardized MVs ($\text{var}(\boldsymbol{x}_{pq}) = 1$), AVE and Communality coincide for less than the constant $1/P_q$.
Goodness of the whole measurement model is measured by *Average Communality* index, *i.e.* the weighted average of all the Q blocks specific Communality indexes, with weights equal to the number of MVs in each block

$$\overline{Com} = \frac{\sum_{q:P_q>1} P_q Com_q}{\sum_{q:P_q>1}} \tag{1.34}$$

This index does not take into account blocks composed of just one MV because their communality is systematically equal to 1.

Goodness of the structural model depends on the portion of variability of each endogenous LVs explained by the corresponding exogenous LV, measured by the multiple determination coefficient (R^2); nevertheless, also communalities of the endogenous LVs must be taken into account. So, for each endogenous LV, the following Redundancy index measures the portion of variability of MVs, related to an endogenous LV $\boldsymbol{\xi}_j$, explained by its latent predictors

$$Red_j = Com_j \times R_j^2 \tag{1.35}$$

The *Average Redundancy* index measures the quality of the whole structural model. It is the average of the redundancies in the model. If J is the number of the endogenous LVs,

$$\overline{Red} = \frac{1}{J} \sum_{j=1}^{J} Red_q \tag{1.36}$$

The global quality of the model is assessed by the Goodness of Fit (GoF) index [Tenenhaus, Amato & Esposito Vinzi 2004]. The GoF is computed as the geometric mean of the Average Communality and the average $\overline{R^2} = (1/J) \sum_{j=1}^{J} R_j^2$ of the J coefficients of determination coefficients

$$GoF = \sqrt{\overline{Com} \times \overline{R^2}} \tag{1.37}$$

A normalized version of the same index has been proposed by Tenenhaus et al. [2004]. In the normalized GoF, $\sum_{p=1}^{P_q} \widehat{\lambda}_{pq}^2$ and R_j^2 are divided

for their theoretical maximum. In particular, the sum of the squared loadings $(\widehat{\lambda}_{pq})^2$ can not be greater than the sum of the squared loadings $(\lambda_{pq}^{\mathrm{PCA}})^2$ corresponding to the first Principal Component of block q; on the other hand, maximum value of R_j^2 is given by the square of canonical correlation ρ_j between $\boldsymbol{\xi}_j$ and the MVs associated to the latent predictors explaining $\boldsymbol{\xi}_j$.

$$GoF_{norm} = \sqrt{\frac{1}{\sum_{q:P_q>1} P_q} \sum_{q:P_q>1} \frac{\sum_{p=1}^{P_q} (\widehat{\lambda}_{pq})^2}{\sum_{p=1}^{P_q} \left(\lambda_{pq}^{\mathrm{PCA}}\right)^2} \times \frac{1}{J} \sum_{j=1}^{J} \frac{R_j^2}{\rho_j^2}} \tag{1.38}$$

This index is bounded between 0 and 1.

Communalities, Redundancies and GoF measure respectively the capacity of outer, inner and global models in explaining the MVs. By cross-validating these indexes, it is possible to evaluate the capacity of inner, outer and global models in predicting the MVs,*i.e.* in explaining observations that do not participate to the PLS estimation procedure. Wold [1982] proposed to use Stone-Geisser's approach [Stone 1974, Geisser 1974] to cross validation. This approach follows a blindfolding procedure: data matrix is divided in G groups and a PLS Path Model is run G times by excluding each time one of the groups (for a review on missing data handling in PLS-PM, see Tenenhaus et al. [2005]). Once model parameters are estimated, any missing data can be predicted as

$$\widehat{x}_{pq(-i)} = \widehat{\xi}_{q(-i)} \widehat{\lambda}_{pq(-i)},$$

where $\widehat{\xi}_{q(-i)}$ and $\widehat{\lambda}_{pq(-i)}$ are computed in models considering i-th value ($i = 1, \ldots, N$) of variable $\boldsymbol{x}_{pq}$ as missing data. Assuming that the variances of MVs are close each other (or simply working on standardized data), the cross-validated communality is obtained as

$$H_q^2 = 1 - \frac{\sum_{p=1}^{P_q} \sum_{i=1}^{N} (x_{pqi} - \widehat{x}_{pq(-i)})^2}{\sum_{p=1}^{P_q} \sum_{i=1}^{N} x_{pqi}^2} \tag{1.39}$$

On the other side, the cross-validated Redundancy index is

$$F_j^2 = 1 - \frac{\sum_{p=1}^{P_q} \sum_{i=1}^{N} (x_{pqi} - \widehat{\lambda}_{pq(-i)} \widehat{\widehat{\boldsymbol{\xi}}}_{j(-i)})^2}{\sum_{p=1}^{P_q} \sum_{i=1}^{N} x_{pqi}^2} \tag{1.40}$$

where $\widehat{\widehat{\boldsymbol{\xi}}}_{j(-i)}$ is the prediction of $\widehat{\boldsymbol{\xi}}_{j(-i)}$ in the structural model computed without including the i-th observation of $\boldsymbol{x}_{pq}$. This index is built under the further assumption that the regression coefficient of $\widehat{\widehat{\boldsymbol{\xi}}}_j$ in the regression of $\boldsymbol{x}_{pq}$ on $\widehat{\widehat{\boldsymbol{\xi}}}_j$ is close to $\widehat{\lambda}_{pqi}$.

If the cross-validated communality indices for each block are all positive, their mean can be used to measure the quality of the measurement model. At the same way, the mean of the cross-validated redundancy indexes can be used to assess the quality of the whole model.

1.4.4 Model validation

Since PLS-PM is a *Soft Modeling* approach, model validation regards only the way relations are modeled, in both the structural and the measurement model; in particular, the following null hypotheses should be rejected:

- $\lambda_{pq} = 0$, as each reflective MV is supposed be correlated to corresponding LV;
- $\omega_{pq} = 0$, as each LV is supposed be affected by all MVs of its block;
- $\beta_{qq'} = 0$, as each latent predictor is assumed be causative with respect to its latent response;
- $R_j^2 = 0$, as each endogenous LV $\boldsymbol{\xi}_j$ is assumed be explained by its latent predictors;
- $\text{cor}(\boldsymbol{\xi}_q, \boldsymbol{\xi}_{q'}) = 1$, as LVs are assumed to measure concepts that are different between them. Rejecting this hypotesis means to assess the *Discriminant Validity* of the PLS Path Model;
- AVE_q and $AVE_{q'} \leq \text{cor}(\boldsymbol{\xi}_q, \boldsymbol{\xi}_{q'})$, as a LV should be related more strongly with its block of indicators than with another LV representing a different block of indicators.

If some of these hypotheses is not rejected, something was wrong in choosing variables or in model specification.

Since PLS-PM avoids any distributional hypothesis on MVs, it is not possible to extend sample information to population through a classic inferential procedure. In order to get confidence intervals for model parameters, resampling techniques, such as Jackknife and Bootstrap [Efron 1982], can be used. However, when relations between LVs are modeled in a OLS framework, confidence intervals for parameters of the structural model can be obtained by means of the usual Student's t test. In the practice, in fact, decision rules yielded by this parametric test are similar to decision rules yielded by resampling based tests, as Student's t test is robust to deviations from normality hypothesis.

1.4.5 Optimizing criteria

PLS-PM is a very flexible technique. PLS Path models may differ in number of LVs, in the path linking them, and in the way we calculate both inner and outer weights. As a consequence, there is not an overall scalar function optimized by whatever model. However, the stationary equation for most of the models has been found out in recent years, showing that PLS-PM generalizes most of the Multivariate Analysis techniques. In the following, a brief recap of the criteria optimized by PLS-PM is given, distinguishing between the two-block case and the multi-block case.

Two-block case

In a path model with two blocks $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$, PLS-PM algorithm converges to three different stationary equations [Lyttkens, Areskoug & Wold 1975], depending on the way the outer weights are calculated (the scheme used in inner estimation does not affect the results). In particular:

- if both $\boldsymbol{\omega}_1$ and $\boldsymbol{\omega}_2$ are estimated with *Mode A*, the covariance between the LVs is maximized. As a consequence, PLS-PM algorithm converges to the first component of Inter-battery Analysis [Tucker 1958] and PLS Regression (see section 1.3.10);

- if both $\boldsymbol{\omega}_1$ and $\boldsymbol{\omega}_2$ are estimated with *Mode B*, the correlation between the LVs is maximized: PLS-PM algorithm converges to the first component of the Canonical Correlation Analysis (CCA) [Hotelling 1936];

- if $\boldsymbol{\omega}_1$ is estimated with *Mode A* and $\boldsymbol{\omega}_2$ is estimated with *Mode B*, the redundancy of $\boldsymbol{X}_1$ on $\boldsymbol{X}_2$ is maximized: PLS-PM algorithm converges to the first component of the corresponding Redundancy Analysis [Van de Wollemberg 1977, D'Ambra & Lauro 1982].

Multi-block case

PLS Path Modeling can be also seen as the generalization of a number of multi-block methods, and in particular:

- *hierarchical model*: each block $\boldsymbol{X}_q$ is connected to a super-block $\boldsymbol{X}_{q+1}$ obtained by juxtaposing $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_q$ (see figure 1.3). Using *Mode B* for all of the blocks in such path model finds out different approaches to generalization of CCA to multi-block case. In particular, depending on the inner estimation scheme, PLS-PM algorithm converges to stationary equations of Horst's and Carrol's Generalized Canonical Correlation analyses ([Horst 1961, Carroll 1968]). On the other hand, using *Mode A* and path weighting scheme may lead to different techniques, among them Multiple Factor Analysis [Escofier & Pagés 1994], depending on slight transformations of original data. For a a complete review on multi-block hierarchical case, see Tenenhaus et al. [2005].

- *confirmative model*: each LV is related to a single block, and it is connected to all the LVs related to the other blocks (see figure 1.4). This path model leads to the stationary equations of Ketterning's generalized CCA [Kettenring 1971]. For further interpretations of the multi-block confirmative case, refer toTenenhaus & Hanafi [2009].

- *Mode B general model*: All outer weights are calculated by means of *Mode B* estimation process. Following Glang [1988] and Mathes [1993], the Lagrange equations associated to the optimization of the criterion

$$\sum_{q \neq q'} c_{qq'} |\mathrm{cor}(\boldsymbol{X}_q \boldsymbol{w}_q, \boldsymbol{X}_{q'} \boldsymbol{w}_{q'})| \tag{1.41}$$

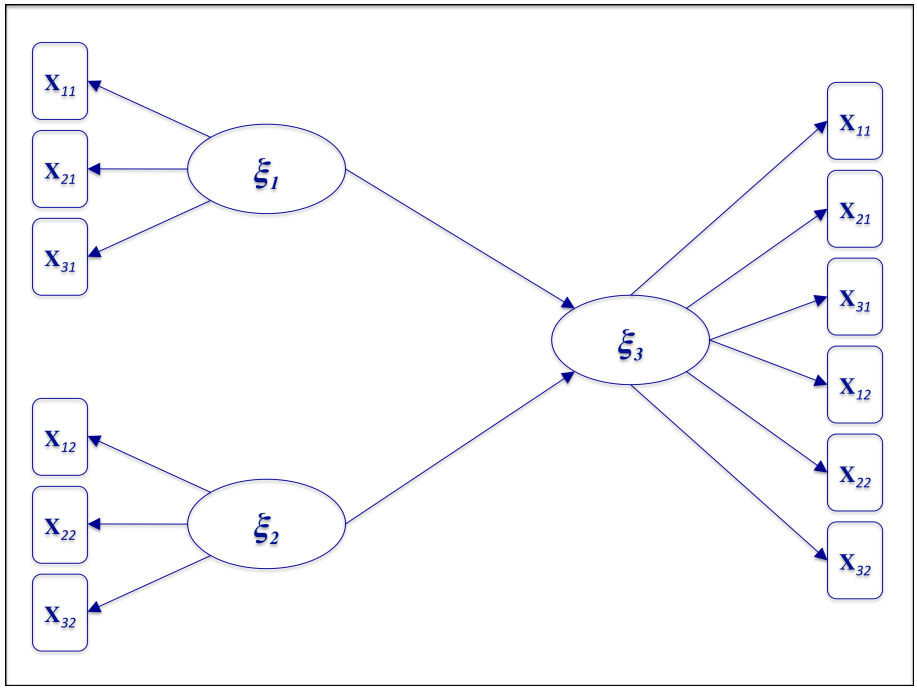

Figure 1.3: *An example of Hierarchical Path Model with three reflective blocks*

with respect to $\|\boldsymbol{w}_q = 1\|$, give exactly the stationary equation of PLS-PM algorithm when estimation *Mode B* is used in all the blocks and the centroid scheme is used in inner estimation of the LVs. They showed also that the Lagrange equations associated

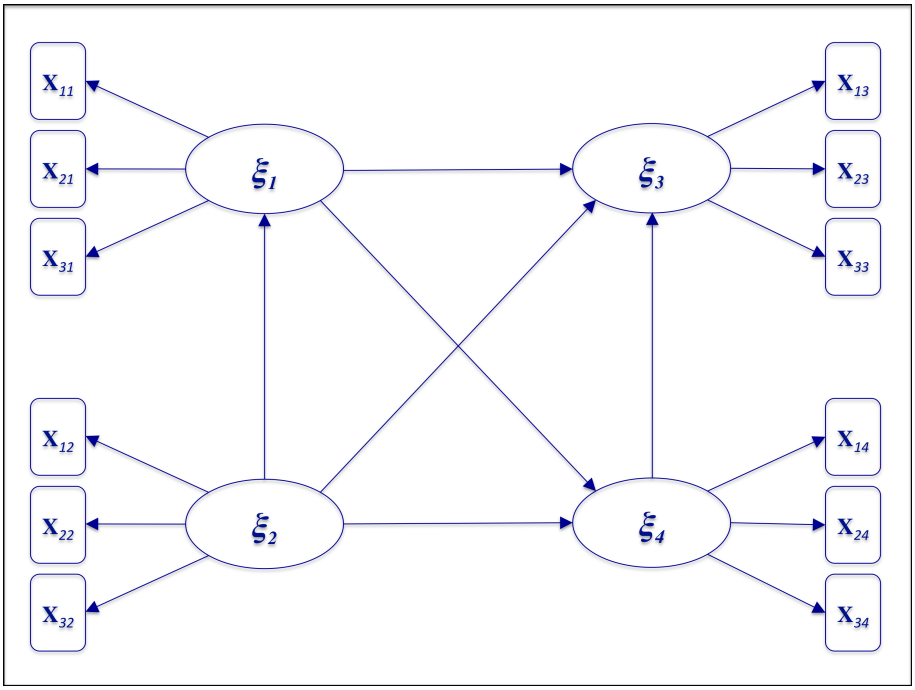

Figure 1.4: *An example of Confirmatory Path Model with four reflective blocks*

to the optimization of the criterion

$$\sum_{q \neq q'} c_{qq'} \mathrm{cor}^2(\boldsymbol{X}_q \boldsymbol{w}_q, \boldsymbol{X}_{q'} \boldsymbol{w}_{q'}) \tag{1.42}$$

with respect to $\|\boldsymbol{w}_q = 1\|$, give exactly the stationary equation of PLS-PM algorithm when estimation *Mode B* is used in all the blocks and the factorial scheme is used in inner estimation of the LVs. Hanafi [2007] proved that Wold's iterative procedure is monotonically convergent to these criteria.

- *new Mode A general model*: All outer weights are calculated by means of the so-called *new Mode A* estimation process. *Mode A* general PLS Path Model seems do not optimize any criterion, as Krämer [2007] showed that *Mode A* Wold's algorithm is not based on stationary equations related to the optimization of a twice differentiable function. However, Tenenhaus [2009] has recently extended the results of Hanafi to a slightly adjusted *Mode A* in which a normalization constraint is put on the weights. In particular, he showed that Wold's procedure, applied to a PLS Path Model where the *new Mode A* is used in all the blocks, monotonically converges to the criterion

$$\arg\max_{\|\boldsymbol{w}_q=1\|} \sum_{q\neq q'} c_{qq'} |\mathrm{cov}(\boldsymbol{X}_q \boldsymbol{w}_q, \boldsymbol{X}_{q'} \boldsymbol{w}_{q'})| \tag{1.43}$$

when centroid scheme is used for the inner estimation of the LVs. Analogously, it converges to the criterion

$$\arg\max_{\|\boldsymbol{w}_q=1\|} \sum_{q\neq q'} c_{qq'} \mathrm{cov}^2(\boldsymbol{X}_q \boldsymbol{w}_q, \boldsymbol{X}_{q'} \boldsymbol{w}_{q'}) \tag{1.44}$$

when factorial scheme is used for the inner estimation of the LVs.

At the present, hence, if all blocks are modeled in the same way, PLS-PM seems to be an heuristic only when path weighting scheme is used.

Chapter 2

Theory of Scales of Measurement and Scaling

2.1 Introduction

2.2 Theory of scales of measurement

Properties of data which have to be analyzed are important because they determine which mathematical operations one can perform on the data. This, in its turn, determines which statistics are allowed for the data.

A scale is an ordered set of values or a set of categories to which an attribute is mapped. The scale defines the range of possible values that can be produced by executing the measurement method.

In its seminal paper on the theory of the scales of measurement, Stevens [1946] classified measurements into four different types of scales. Even nowadays, Stevens' theory is widely adopted, even if there has been, and continues to be, debate about the merits of his classification scheme [Velleman & Wilkinson 1993].

In table 2.1 are listed the four type of scales proposed by Stevens: nominal, ordinal, interval and ratio. Each of these scales is characterized by some of the following properties:

- *Equality* (or *Grouping*): each number defines a particular group of units.

- *Order* (or *Magnitude*): numbers have an inherent order from smaller to larger.

- *Equal intervals*: differences between numbers anywhere on the scale are the same (e.g., the difference between values 4 and 5 is the same as the difference between 7 and 8).

- *Absolute/true zero*: the zero point represents the absence of the property being measured. This property implies that equalities between ratios can be assessed.

Here, with the word *number* we mean a numeric label which may refer to both a qualitative (category or group) and a quantitative concept (value). In the following, we review the features of the possible types of scales, as classified by Stevens.

Scale	**Basic empirical operations**	**Mathematical group structure**	**Permissible statistics**
NOMINAL	Determination of equality	Permutation group	mode, chi square
ORDINAL	Determination of greater or less	Isotonic group	median, percentile
INTERVAL	Determination of equality of intervals or differences	General linear group	mean, standard deviation, product moment and rank order correlations
RATIO	Determination of equality or ratio	Similarity group	geometric mean, harmonic mean, coefficient of variation

Table 2.1: Stevens different type of measurement scales.

- *Nominal* scale. It is the lowest scale of measurement. Numbers are assigned to categories (or groups) of units as labels. Which number is assigned to which group is completely arbitrary: the scale remains invariant under the general substitution or permutation group. The number just identifies the group to which units are assigned. Therefore, the only property of the nominal scale of measurement is equality (or grouping). The only mathematical operation we can perform with nominal data is to count how many units belong to a group. As a consequence permissible statistics for nominal data is the mode, at an univariate level, and chi-squares, at a bivariate level.

- *Ordinal* scale. It is the most common scale in psychometrics. Ordinal scale has the property of magnitude (order) as well as equality. The numbers represent an attribute being measured

(group membership) and can tell us whether a case has more (or less) of the observed attribute than another case. The distance between scale points is not equal. Ranked preferences are an example of ordinal scales encountered in everyday life. We also address the concept of unequal distance between scale points. Because of the property of magnitude (or order), the numbers are no longer considered arbitrary as they are in nominal scales. Since any order-preserving transformation will leave the scale form invariant, the structure of this scale can be called isotonic. Further statistics, like median and percentiles, can be calculated on data measured on this scale,

- *Interval* scale: Equal distances on this scale correspond to equal quantities of the attribute without the use of 0 values. Interval scales have the following properties: equality, magnitude, and equal distance. The equal distance between scale points allows us to know how many measurement units one case is greater or smaller from another on the measured characteristic. So, we can always be confident that the meaning of the distance between 25 and 35 is the same as the distance between 65 and 75. Interval scales do not have a true zero point; the zero point on an interval scale is a matter of convention or convenience; this is shown by the fact that the scale form remains invariant when a constant is added. The interval scale of measurement permits mathematical operations of addition and subtraction. Ratios between numbers on the scale are not meaningful, so operations such as multiplica-

tion and division cannot be directly carried out. However, ratios of differences can be expressed; for example, one difference can be twice another. Hence, we can speak about equal *relative* distance (or spacing) property. This property implies that central tendency indexes can be computed on interval data. Statistical dispersion can be measured in most of the usual ways, which just involves differences or averaging, such as range and interquartile range. Also standardized moments are permissible, since ratios of differences are meaningful, but one cannot define coefficient of variation, since the mean is a moment about the origin, unlike the standard deviation, which is (the square root of) a central moment.

- *Ratio* scale: it is the most common encountered in physics. Ratio scale has all of the properties previously listed. These properties allow us to apply all mathematical operations (addition, subtraction, multiplication, and division) used in data analysis. The absolute/true zero allows us to know how many times greater one case is than another. Once such a scale is erected, its numerical values can be transformed only by multiplying each value by a constant if we want that scale form remains invariant. All types of statistics are applicable to ratio scales.

Characteristics of the four measurement scales lead us to divide scales, from the mathematical-statistical point of view, in two great classes: *non-metric* and *metric*.

We define the variables observed on non-metric measurement scales

as *non-metric variables*, and the variables observed on metric measurement scales as *metric variables*.

Non-metric variables can be unordered (*i.e. nominal*, for example religion or marital status) or ordered (*i.e. ordinal*, for example judgments or Likert scales). The categories of nominal variables have labels while the categories of ordinal variables have ordered labels (such as low, medium, high, or never, sometimes, always) or numbers. In this last case, we prefer to speak about *pseudo-numbers*. From the mathematical-statistical point of view, these *pseudo-numbers* are just labels representing an ordered sequence, as they can not be added or subtracted. Hence, they can not be regarded as numeric values.

From the mathematical point of view, nominal and ordinal variables are respectively ordered and ordered *sets*. Interval and ratio variables, instead, are *metric structures*, *i.e.* sets where notion of distance (metric) between elements of the set is defined. As a consequence, metric variables have an unit of measurement, while non-metric data do not. The main implication from the statistical point of view is that all the standard factorial analyses can be performed only on metric variables.

2.3 The Scaling approach

Once a set of measurements have been made on a particular scale, it is possible to transform the measurements to yield a new set of measurements at a different level. As a matter of fact, it is always possible to transform from a stronger level to a weaker level. For

example, a temperature measurement in degrees Kelvin is at the ratio level. If we convert the measurements to degrees Celsius, the level is interval. If we rank the measurements, the level becomes ordinal.

Scaling techniques allow us to convert a weaker measurement scale to a stronger measurement scale. Scaling a variable means to provide non-metric variables with a metric; however, also a metric variable can be re-scaled in the case where it is provided with a new metric.

We define a non-metric scaling as a scaling which does not depend on the metric properties of the variable. A metric scaling, instead, depends on the metric properties of the variable, *i.e.* is obtained as a functional transformation of the variable.

In scaling approach, each observed category (which can be represented by a label, a pseudo-number or a numeric value) of the raw (*i.e.* to be scaled) variable $\boldsymbol{x}^*$ is replaced by a numerical value. The new scale is an interval scale, independently on the original measurement scale of $\boldsymbol{x}^*$.

In the scaling process certain properties of the raw data are preserved. The researcher must decide which of the properties of the old measurement scale have to be retained in the new metric scale. This means to choose which properties of the original variable must be preserved in the scaled variable $\tilde{\boldsymbol{x}}$.

It is noteworthy that the scaling level is the level at which a variable is analyzed, which does not need to retain all of the properties of measurement level of the variable.

If the researcher wants to preserve in scaled variable all of the

properties of the raw variable, the scaling level should be chosen in accordance with the measurement level of the variable. With nominal scaling level, only the grouping property is preserved, while ordinal scaling level preserves grouping and order properties, and the linear scaling level preserves grouping, ordering, and equal relative spacing.

However, the researcher can decide to do not keep all of the properties of the raw variable in the new measurement scale. This implies that it is possible choose among different levels of scaling analysis for the same variable, depending on which properties of its measurement scale we want to preserve. For example, an interval variable can be scaled in such a way to retain just its grouping property.

Metrics of new interval scales are constrained depending on the scale level at which a raw variable is measured and on which of its properties the researcher wants to preserve. These restrictions, together with the different levels at which a variable can be scaled, will be discussed in detail in the next section.

2.4 Scaling levels and corresponding restrictions

A variable measured on a nominal scale carries information just about the group membership: observations sharing the same category belong to the same group. Hence, a nominal variable can be scaled only at a nominal level. In order to respect grouping property, the scaled

variable $\hat{\boldsymbol{x}}$ must be constrained in such a way that

$$(x_i^* \sim x_{i'}^*) \Rightarrow (\hat{x}_i = \hat{x}_{i'}), \tag{2.1}$$

where the symbol $\sim$ indicates membership in the same category and x^* represent a raw variable.

An ordinal variable has a further property, as its categories are unequivocally ordered. Hence, it can be scaled both at nominal and ordinal level. If we choice the last option, in order to preserve the order property, we must add an order constraint which further reduces the space of all its possible quantifications. Guttman [1968] proposed an approach providing strictly monotone quantification, such that

$$(x_i^* \sim x_{i'}^*) \Rightarrow (\hat{x}_i = \hat{x}_{i'}) \text{ and } (x_i^* \prec x_{i'}^*) \Rightarrow (\hat{x}_i < \hat{x}_{i'}), \tag{2.2}$$

where symbol $\prec$ indicates empirical order. These restrictions impose that each category must be quantified by an individual numerical value, as well as that different categories must be quantified by different values. A weaker approach to monotonicity, instead, allows unequal categories to be quantified with the same value, following the rules:

$$(x_i^* \sim x_{i'}^*) \Rightarrow (\hat{x}_i = \hat{x}_{i'}) \text{ and } (x_i^* \prec x_{i'}^*) \Rightarrow (\hat{x}_i \leq \hat{x}_{i'}). \tag{2.3}$$

These restrictions correspond to which used in the weak monotonicity approach by Kruskal [1964*b*]. They are used also in discrete-ordinal scaling by de Leeuw, Young & Takane [1976].

Even a numeric variable can be re-scaled. It can be transformed regardless of its metric properties and preserving grouping and ordering properties, or just grouping property. In these cases it is treated as it was a non-metric variable, with the number of categories equal to the number of distinct values of the variable (thus, values of the numeric variable will also be referred to as category values). Another possibility is to keep into account in the scaling its metric properties by imposing some functional restrictions, for example requiring that scaled and raw variables are related by the following polynomial rule

$$\hat{x}_i = \sum_{d=0}^{D} \alpha_d \hat{x}_i, \tag{2.4}$$

being D the degree of the polynomial function [Young 1981]. We will call this level of scaling as functional. A particular case of this scaling is the linear scaling level, obtained for $D = 1$.

Another discriminant in data scaling is tie handling. The question is: should equal categories be quantified by equal values? This question regards data measurement process, which is different from their measurement level. It implies two possible scaling solutions, called by Kruskal [1964*b*] primary and secondary approach to the ties. So far we considered Kruskal's primary approach, where all observations in a particular category are represented by the same real number; this implies that condition 2.1 is always verified. Young [1981] suggests to use this approach if we want to reflect in the scaling the fact that variable is measured through a discrete measurement process.

In the secondary approach to the ties the same category can be replaced by several values

$$(x_i^* \sim x_{i'}^*) \Rightarrow (\hat{x}_i \lesseqgtr \hat{x}_{i'}) \tag{2.5}$$

According to Young [1981], this approach adopts a continuous assumption about the measurement process underlying data. It has been proposed in several papers [de Leeuw et al. 1976, Young 1981], where it is called continuous option: a real number selected from a closed interval of real numbers is assigned to each of the observations within a particular category. The continuous option is suggestive from the theoretical point of view, as it include in the model the measurement process underlying the data. However, in the practice it gives trivial results when (as it usually happens) there are a small number of categories with respect to the number of the observations [Gifi 1990]. Moreover, in the reality all of the measurements involve a finite number of categories, because of the finite precision of the measurement process. Hence, difference in data coming from different measurement processes, is an issue much more philosophical than concrete.

Chapter 3

A PLS approach to Optimal Scaling: the Non Metric PLS methods

3.1 Motivation

As we showed in Chapter 2, PLS methods are component based techniques. Components (or latent variables) are obtained as linear combination of the corresponding block of indicators (or variables). The main parameters in all PLS models are the weights associated to variables to build the components. In NIPALS the weight associated to a generic variable measures the relation between the variable and the component. In PLS-R it measures the relation between the variable and a linear combination of the variables in the other space. Finally,

in PLS-PM weights measure the relation between each manifest variable and a linear combination (the corresponding latent variable inner estimate) of linear combinations (the outer estimates of connected latent variables) of manifest variables belonging to connected blocks. In NIPALS, PLS-R and PLS-PM, when we work on standardized variables, weights are expressed as a function of Pearson product-moment correlation coefficient. This leads to two basic hypotheses underlying PLS models:

- Each variable is measured on interval (or ratio) scale.
- Relations between variables and latent constructs are linear and, as a consequence, monotone.

As a consequence, standard PLS methods can not handle data which are measured on a scale which has not metric properties.

There exists a simple way to overcome this problem: replacing each non-metric variable with the corresponding indicator matrix. Most of the softwares currently used to perform PLS analyses use such a coding in order to handle categorical variables; however, in author's opinion, this is not an effective solution to the problem (see section 3.2).

In this chapter an alternative approach to handling non-metric variables in PLS framework is proposed: the Optimal Scaling (OS). OS has been extensively proposed in multidimensional data analysis to generalize MDA methods in a way that they can handle variables measured on a variety of different scales (see section 3.3.1). In all of these methodology, OS is implemented by Alternating Least Squares

algorithms [de Leeuw et al. 1976]. In the next, it will be proven that also properly adjusted PLS algorithms can be used as OS algorithms.

3.2 Drawbacks of binary coding in PLS framework

A simple approach to cope with the quantification problem, which can be easily used in whatever multidimensional data analysis method, is to replace each non-metric variable $\boldsymbol{x}^*$ with the corresponding N by K indicator matrix $\widetilde{\boldsymbol{X}}$. $\widetilde{\boldsymbol{X}}$ has a row for each of the N observations and a column for each of the K categories; its element $\widetilde{x}_{i,k}$ equals the unit if the i-th observation belongs to the k-th category and it is null otherwise.

PLS are strongly component based methods. PLS components are always built in order to well represent the variables, because the fundamental task of PLS is exploring data. The exploration of the relations between latent concepts makes sense in PLS framework only if they effectively summarize the variables. Starting from these considerations, it is easy to understand why, in author's opinion, the main outputs of any PLS algorithm are the weights assigned to each variable to build the component.

Binary coding presents some relevant drawbacks which affect the interpretability of the resulting weights. First of all, a binary coded gives up the idea of the variable as a whole, while it considers categories as they were variables in themselves. As a consequence, whatever PLS

analysis on binary variables yields a weight for each category, and not for the whole categorical variable. PLS weights measure the intensity of the relation between indicators and latent constructs. Using the binary coding, instead, such weights measure the impact of each individual category on the latent construct; this makes it impossible to evaluate the importance of the whole variable in the model, as well as to compare the weight of a variable with the ones of the other variables.

Secondly, the binary coding affects the dimensionality of the data matrix, as each categorical variable is coded in as many binary variables as the number of its categories. Hence, the number of categories affects the relative impact of the categorical variables with respect to the other variables. Moreover, if the number of categories is large, binary coding generates sparse matrices.

Finally, the weight of binary variable representing a category mainly associated to central values of the corresponding latent variable (or component) score distribution is systematically underestimated. In fact, such binary variables are always linked to the latent construct by a non-monotonic relation. This type of relation can not be reflected in the weights associated to these variables, as they are expressed in terms of linear correlation coefficient.

These considerations lead us to discard this approach and to propose the OS approach in order to handle non-metric variables and investigate non linearity in PLS framework.

3.3 Optimal Scaling

According to Young [1981], "*Optimal scaling is a data analysis technique which assigns numerical values to observation categories in a way which maximizes the relation between the observations and the data analysis model while respecting the measurement character of the data*".

Hence, in order to define a scaling process as optimal, the resulting scaling must be:

- suitable, as it must respect the constraints defining which among the properties of the original measurement scale we want to preserve.
- optimal, as it must optimize the same criterion of the analysis in which the Optimal Scaling is involved.

Optimal scaling means to apply *ad hoc* scaling functions to non-metric variables in such a way to transform them in numerical variables. This process is usually called *quantification*. However, optimal scaling can be applied also to numerical variables, in order to detect and handle non-linear relations. In this last case, it is more correct to speak about *transformation* of numerical variables.

Optimal scaling methods are able to handle variables regardless of their measurement scale. Non-metric variables, in fact, are quantified in such a way that they can be analyzed as they were measured on an interval scale. Metric variables, instead, can be properly transformed in order to investigate non linearity in data.

3.3.1 Alternating Least Squares approach to Optimal Scaling

OS has been extensively implemented in multivariate analysis by iterative algorithms of the family Alternating Least Squares (ALS) [de Leeuw et al. 1976, Young 1981, Gifi 1990]: for this reason, these algorithms are called also ALSOS (Alternating Least Squares approach to Optimal Scaling) [Young 1981]. Each ALSOS algorithm optimizes an objective loss function by using an algorithm based on the ALS and OS principles.

The OS principle involves viewing observations as categorical, and then representing each observation category by a parameter. This parameter is subject to constraints implied by the measurement characteristics of the variables.

In ALS approach all of the parameters are divided into two mutually exclusive and exhaustive subsets: the parameters of the model and the parameters of the data (or scaling parameters). Then, the loss function is optimized by alternately optimizing with respect a subset, then the other.

In particular, the ALSOS flow proceeds as follows: least squares estimates of model parameters are obtained while assuming that the scaling parameters are constants. These least squares estimates are defined *conditional*, since the least squares nature is conditional on the values of the parameters in the other subsets. Successively, conditional least squares estimates of scaling parameters is obtained for fixed model parameters. This ALS procedure is iterated until conver-

gence.

The ALSOS approach has been applied to the most various fields of multivariate data analysis: analysis of variance [de Leeuw et al. 1976, Gifi 1990], multiple regression and canonical correlation analysis [Young, de Leeuw & Takane 1976], discriminant analysis [Gifi 1990], principal component analysis (see section 4.3) [Young, Takane & de Leeuw 1978, de Leeuw & Van Rijckevorsel 1980, Gifi 1990], path analysis [Gifi 1990], common factor analysis [Takane, Young & de Leeuw 1979] and multidimensional scaling [Takane, Young & de Leeuw 1977].

3.4 Partial Least Squares for Optimal Scaling

Quoting Young [1981], "*Certain strong correspondences exist between an ALSOS procedure and the NILES approach developed by Wold and Lyttkens .. The main difference between these metric algorithms and the nonmetric ALSOS algorithms is the optimal scaling features of the ALSOS algorithm. .. The scaling feature permits the analysis of qualitative data, whereas the previous procedures can only analyze quantitative data*".

The computational core of any PLS method, as well as the one of ALSOS algorithms, is based an iterative least squares procedure used to compute model parameters. In particular, in NIPALS approach to PCA, PLS iteration is used for analyzing the covariance within a

block of variables; in PLS-R it is used to analyze the cross-covariance between two blocks of variables; in PLS-PM, PLS iteration is used to analyze the cross-covariance among different blocks of variables. All these algorithms exploit the PLS iteration in order to handle multicollinearity, missing data, landscape matrices and to explore data regardless of distributional hypothesis.

In the following we will investigate a peculiarity of PLS iteration which has not been fully exploited yet. We refer in particular to the potentiality of PLS algorithms to yield data scaling with optimal features.

In order to exploit this potentiality we propose to adjust the PLS iteration to device an optimal scaling procedure, calculating iteratively both scaling and model parameters. This new PLS procedure leads to a new class of algorithms which implement methods that generalize the standard PLS methods. We call them Non-Metric PLS (NM-PLS) methods [Russolillo & Trinchera 2009a], because they are able to provide data with a new metric structure, which does not depend on the metric properties of the data. In other words, NM-PLS methods yield a metric to non-metric data, and a new metric to metric data, making relationships between variables and latent constructs linear, as required by the hypothesis of standard PLS models (see section 3.1). These methods could be named non-linear PLS methods as well, since they discard the intrinsic linearity hypothesis in standard PLS methods. However, by naming them Non-Metric PLS methods, we preferred to highlight their ability to work just on non-metric features

of data.

NM-PLS methods overcome the limits of PLS techniques remarked in section 3.1. In fact, they are able to quantify non-metric variables, in such a way to make possible analyses by means of standard statistical indexes; this allows us to analyze non-metric variables together with variables measured at a higher scale level. Moreover, since in NM-PLS methods the variables are scaled in a way to linearize their relations with latent constructs, NM-PLS methods are not affected by violations of linearity hypothesis intrinsic in each PLS model.

In the following section, the optimality of scalings provided by NM-PLS methods will be discussed. In particular, NM-PLS scaling will be proven to be suitable, as it respect the constraints depending on which ones we want to preserve among the properties of its measurement scale, as well as optimal, as it optimizes the same criterion of the analysis in which the Optimal Scaling is involved.

3.4.1 Optimal scaling with respect to a latent criterion

According to Hayashi [1952], methods of quantification can be divided in two main classes [Tanaka 1979]. The first class contains the methods for the case where an external criterion is present. These methods are aimed to the prediction of the external criterion or to the analysis of the effects of factors. The other class contains the methods for the case where no external criterion is present. These methods are used to construct a spatial configuration so as to grasp the mutual

relationships of the data.

In PLS framework, like in all component-based methods, there is not an external criterion to which to relate the quantifications. So, the non-metric approach to PLS falls in the second class of Hayashi's methods. However, NM-PLS methods face the problem from another point of view: optimal quantifications (scalings) are found out with respect to a latent construct, that we will call Latent Criterion (LC).

The LC is an unknown vector of order N, centered by construction. For each PLS method different LCa are considered, depending on the way the weights linking each variables to the corresponding component are calculated.

In NIPALS, the weight of variable $\boldsymbol{x}_p$ is calculated in such a way to maximize the squared correlation of the variable with a linear combination $\boldsymbol{t}$ of all the variables (see section 1.2). Hence, in NIPALS, the LC to keep into account in the scaling process is the score vector $\boldsymbol{t}_1$.

In PLS-R, for each variable, the corresponding weight is calculated in such a way to maximize the squared correlation with a linear combination of the variables belonging to the other space. To be clear, the weight of $\boldsymbol{x}_p$ maximizes the squared correlation of the variable with $\boldsymbol{u}$, while the weight of $\boldsymbol{y}_r$ maximizes the squared correlation of the variable with $\boldsymbol{t}$ (see section 1.3). As a consequence, in PLS-R we have to keep into account two LCa: $\boldsymbol{u}_1$ for the predictors and $\boldsymbol{t}_1$ for the responses.

In *Mode A* PLS-PM, the weight of a variable $\boldsymbol{x}_{pq}$ maximizes the squared correlation of the variable with a linear combination $\boldsymbol{z}_q$ of

the variables belonging to connected blocks (see section 1.4). So, in a PLS-PM framework, different LCa $\boldsymbol{z}_q$ have to be considered for each block of variables.

In the next, we will refer to all of these LCa with the generic notation $\boldsymbol{\gamma}$, in order to find out the general criterion optimized by the weight w corresponding to a generic variable $\boldsymbol{x}$ in all the above cited algorithms, that is

$$\text{cor}^2(\boldsymbol{x}, \boldsymbol{\gamma}) \tag{3.1}$$

NIPALS, PLS-R and *Mode A* PLS-PM work on a set of raw variables, measured on various measurement scales. It is possible to show (see chapters 4, 5 and 6) that in all of NM-PLS methods the optimality condition of the scaling with respect to the model criterion is satisfied if, for each raw variable, the resulting quantified variable $\widehat{\boldsymbol{x}}$ optimizes the criterion

$$\text{cor}^2(\widehat{\boldsymbol{x}}, \boldsymbol{\gamma}) \tag{3.2}$$

under the constraints defined by the scaling level of the analysis chosen for $\boldsymbol{x}^*$ (see section 2.4).

NM-PLS methods conceive four levels of scaling analysis: nominal, ordinal, functional and linear. The linear scaling level just implies the standardization of the raw variable, like in ordinary PLS methods. In the next (section 3.4.3) it will be explained why we consider the linear scaling level as a particular case of the functional one. To each of these

level implies an *ad hoc* scaling function, maximizing criterion 3.2. A scaling function $\mathcal{Q}()$ is a real function applied to $\boldsymbol{x}^*$ which generates a numeric value (i.e. the optimally scaled value) $\widehat{x}_i$ for each observation. In the following, we will describe the optimal scaling functions used in NM-PLS methods both for non-metric and metric variables. These functions have already been proposed by [Young 1981] in ALSOS framework for data coming from a discrete measurement process.

3.4.2 Optimal scaling functions for non-metric variables

In NM-PLS methods, a non-metric variable can be analyzed at a nominal or at an ordinal scaling level.

If we want to analyze $\boldsymbol{x}^*$ at a nominal scaling level, we must find, among all its possible quantifications, the one the most correlated to $\boldsymbol{\gamma}$, under the grouping constraint.

From the geometrical point of view, the scalings of the K categories of $\boldsymbol{x}^*$ satisfying the grouping constraint are the points in the space spanned by the columns of the corresponding indicator matrix $\widetilde{\boldsymbol{X}}$. This space is a closed convex cone, denoted $\mathcal{C}_{\mathrm{n}}$[1], in $\Re^N$. The nominal quantifications of $\boldsymbol{x}^*$ are geometrically represented by the rays of $\mathcal{C}_{\mathrm{n}}$. In fact, any vector-ray in the cone respects the constraint for which observations belonging to the same group assume the same value.

Following Hayashi's first quantification criterion, the optimal ray is

[1]A set $\mathcal{C} \subset \Re^N$ is called a cone if $0 \in \mathcal{C}$ and $qx \in \mathcal{C}$ for every $q \geq 0$ and every $x \geq \mathcal{C}$. For further deepening on convex spaces, see Barvinok [2003].

the closest one to the LC $\boldsymbol{\gamma}$, that is the orthogonal projection of $\boldsymbol{\gamma}$ on the cone

$$\widetilde{\mathcal{Q}}(\boldsymbol{x}^*, \boldsymbol{\gamma}) : \widehat{\boldsymbol{x}} = \widetilde{\boldsymbol{X}}(\widetilde{\boldsymbol{X}}'\widetilde{\boldsymbol{X}})^{-1}\widetilde{\boldsymbol{X}}'\boldsymbol{\gamma}. \tag{3.3}$$

The resulting scaling values for the categories of $\boldsymbol{x}^*$ are the K least squares regression coefficients of $\widetilde{\boldsymbol{X}}$ on $\boldsymbol{\gamma}$, which correspond to the averages of $\boldsymbol{\gamma}$ conditioned to $\boldsymbol{x}^*$ categories.
Moreover, the quantified variable contains the LC values predicted by the regression of $\boldsymbol{\gamma}$ on $\widetilde{\boldsymbol{X}}$. The determination coefficient of this regression equals the squared Pearson's correlation ratio between the original categorical variable and the LC. Hence, the relation between $\boldsymbol{\gamma}$ and $\boldsymbol{x}^*$ in terms of linear correlation can be expressed as the Pearson's correlation ratio $\eta_{\boldsymbol{\gamma}|\boldsymbol{x}^*}$

$$\text{cor}\boldsymbol{\gamma}, \widehat{\boldsymbol{x}}) = \eta_{\boldsymbol{\gamma}|\boldsymbol{x}^*}. \tag{3.4}$$

In the next, this equivalence will serve to give an nice interpretation of weights of quantified variables in NM-PLS methods.

If $\boldsymbol{x}^*$ is an ordinal variable, and we want analyze it at an ordinal scale level, we must search for our quantifications in the conic subspace of $\mathcal{C}_{\text{o}}$ respecting the constraint 2.3.

From the operational point of view it would seem to be a better procedure to search for a subset $\widetilde{\widetilde{\boldsymbol{X}}}$ of $\widetilde{\boldsymbol{X}}$ for which order constraints are respected. This subset represent the conic space $\mathcal{C}_{\text{o}}$. The ray of this cone the closest to $\boldsymbol{\gamma}$ is the projection of $\boldsymbol{\gamma}$ into $\widetilde{\widetilde{\boldsymbol{X}}}$. This means that

the quantified variable $\widehat{\boldsymbol{x}}$ is the vector of the predictions of $\boldsymbol{\gamma}$ elements obtained by regressing the LC on $\widetilde{\widetilde{\boldsymbol{X}}}$. Hence, following Young [1975], the ordering scaling function is

$$\widetilde{\widetilde{\mathcal{Q}}}(\boldsymbol{x}^*, \boldsymbol{\gamma}) : \widehat{\boldsymbol{x}} = \widetilde{\widetilde{\boldsymbol{X}}}(\widetilde{\widetilde{\boldsymbol{X}}}'\widetilde{\widetilde{\boldsymbol{X}}})^{-1}\widetilde{\widetilde{\boldsymbol{X}}}'\boldsymbol{\gamma}, \tag{3.5}$$

where $\widetilde{\widetilde{\boldsymbol{X}}}$ is build by Kruskal's secondary least squares monotonic transformation [Kruskal 1964*a*]. The vector of the regression coefficient $(\widetilde{\widetilde{\boldsymbol{X}}}'\widetilde{\widetilde{\boldsymbol{X}}})^{-1}\widetilde{\widetilde{\boldsymbol{X}}}'\boldsymbol{\gamma}$ contains the unnormalized optimal scaling values which preserve the order of the categories of $\boldsymbol{x}^*$, as required by the condition 2.3.

In NM-PLS methods, Kruskal's up-and-down block algorithm (also known as pool-adjacent-violators algorithm) [Kruskal 1964*b*] is implemented in order to obtain $\widetilde{\widetilde{\boldsymbol{X}}}$. This algorithm consists of a suite of regressions of $\boldsymbol{\gamma}$ on indicator matrices. In the first regression the indicator matrix is $\widetilde{\boldsymbol{X}}$. Then, another indicator matrix is obtained by merging adjacent columns of $\widetilde{\boldsymbol{X}}$ representing categories whose quantification does not respect the order. In the following step of the algorithm $\boldsymbol{\gamma}$ is regressed on this matrix. This procedure is repeated until the regression coefficients respect the monotonicity condition.

Kruskal's algorithm implements the monotonic regression of $\boldsymbol{\gamma}$ on $\boldsymbol{x}^*$. The residual variance of the regression is, as a consequence, an index of departure from monotonicity. In fact, it is equal to Kruskal's raw *STRESS* index [Kruskal 1964*a*]. This leads us to state that correlation between $\boldsymbol{\gamma}$ and $\widehat{\boldsymbol{x}}$ can be calculated as a function of *STRESS*. In

particular, if an increasing monotone regression is implemented, then

$$\text{cor}(\boldsymbol{\gamma}, \widehat{\boldsymbol{x}}) = \sqrt{1 - STRESS^2_{(\boldsymbol{\gamma}, \boldsymbol{x}^*)}} \tag{3.6}$$

while, if a decreasing monotone regression is implemented, then

$$\text{cor}(\boldsymbol{\gamma}, \widehat{\boldsymbol{x}}) = -\sqrt{1 - STRESS^2_{(\boldsymbol{\gamma}, \boldsymbol{x}^*)}}, \tag{3.7}$$

where notation *STRESS* indicates the normalized *STRESS* index proposed by Kruskal [1964*a*].

Hence, $\text{cor}(\boldsymbol{\gamma}, \widehat{\boldsymbol{x}})$ can be interpreted as a measure of the approaching to monotonicity of the relation between $\boldsymbol{x}^*$ and the LC; it equals the unity if it exists a perfect increasing monotonicity and it is equal to -1 when it exists a perfect decreasing monotonicity.

3.4.3 Optimal transformation functions for metric variables

As previously discussed, (see section 2.4), a raw numerical variable $\boldsymbol{x}^*$ can be handled both at a non-metric and a non-linear scaling level. In the first case, we consider the different values assumed by the variable as distinct categories, and we use the scaling functions 3.3 and 3.5. In the latter, the different values assumed by the variable are considered as numeric values and non-linear relations between $\boldsymbol{x}^*$ and $\boldsymbol{\gamma}$ can be investigated by projecting $\boldsymbol{\gamma}$ into a particular conic space.

Let's suppose to know the degree of a polynomial relation between

the variable and the LC. The aim is to find an optimal transformation constrained to such functional restriction. Following Young [1981], optimal parameters for the polynomial transformation are found by means of the projection of $\boldsymbol{\gamma}$ in the conic space $\mathcal{C}_{\mathrm{p}}$ spanned by the columns of matrix $\dot{\boldsymbol{X}}$. Matrix $\dot{\boldsymbol{X}}$ is built with a row for each observation and with $D+1$ columns, each column being an integer power of the vector $\boldsymbol{x}^*$. The first column is the zero-th power (that is, all ones), the second one is the first power and so on until the last column, which is the D-th power of $\boldsymbol{x}$. Hence, the transformation function yielding optimal scaled variables will be

$$\dot{\mathcal{Q}}(\boldsymbol{x}^*, \boldsymbol{\gamma}) : \widehat{\boldsymbol{x}} = \dot{\boldsymbol{X}}(\dot{\boldsymbol{X}}'\dot{\boldsymbol{X}})^{-1}\dot{\boldsymbol{X}}'\boldsymbol{\gamma} \tag{3.8}$$

If we suppose that the variable and the LC are linked by a linear relation, we have just to pose $D = 1$.

In the next chapters, we will show how to embed scaling functions 3.3, 3.5 and 3.8 in NM-NIPALS, NM-PLSR and NM-PLSPM algorithms, as well as how to interpret the properties of these scaling function in a PLS framework.

Chapter 4

A Non-Metric PLS algorithm for Principal Component Analysis

4.1 Introduction

Among the PLS techniques reviewed in chapter 1, NIPALS (see section 1.2) is the only which is not a methodology in itself, but an algorithmic tool born to implement a Principal Component Analysis (PCA) in presence of missing data without an *a priori* imputation and avoiding to work on the correlation matrix.

In literature there exist a huge suite of methods performing non-metric PCA. All these methods are implemented by ALS algorithm or gradient methods.

Here a new PLS method for non-metric PCA, called Non-Metric NIPALS (NM-NIPALS) is proposed. It represents the statistical base on which more complex Non-Metric PLS methods lie. Moreover, just because of the exstensive literature, NM-NIPALS can be compared to the other well known methods, in order to assess its theoretical consistency. This can not be done with non-metric approaches to PLS Regression and PLS-PM, since they are methods in itself, maximizing specific criteria.

In this chapter the history (section 4.2) and the methodology (sections 4.3 and 4.4) of the multivariate descriptive analysis of categorical variables by non parametric techniques will be reviewed. Then, the Non-Metric NIPALS will be presented (section 4.5), its properties investigated (section 4.5.1 and section 4.5.2), and its connections to other non-metric approaches to PCA will be discussed (section 4.5.3).

4.2 Multivariate descriptive analysis of categorical variables: an historical review

Multivariate descriptive analysis of two or several categorical variables has a long tradition in statistical literature. Early work on multidimensional quantification of categorical data can be traced back to Richardson & Kuder [1933]. They used a method that was successively called by Horst [1935] *method of the reciprocal averages* in order

to build a rating scale of employees, introducing the idea of scalig categories and individuals, both of them connected with averaging relations. In the same year the *Simoultaneous linear regression approach* for the quantitative analysis of a two way table was proposed by Hirschfield [1935]. Some year later, basic principles of both these approaches were rediscovered by Fisher [1940] as Discriminant Analysis methods. Later, the problem was quite completely formalized by Guttman [1941], who successively introduced the concept of *Scalogram Analysis* [Guttman 1950].

In the 1950s Hayashi launched a series of studies on the *theory of quantification*, where he dealt with symmetric data matrices [Hayashi 1950], non-symmetric matrices [Hayashi 1952] and multidimensional tables [Hayashi 1954].

The term *Optimal Scaling* was introduced by Bock [1960] to unify "*all the approaches which assign numerical values to alternatives, or categories, so as to discriminate optimally among the objects ... in some sense. Usually it is the Least Squares sense, and the values are chosen so that the variance between objects after scaling is maximum with respect to that within objects*".

In the same decade the Benzécri's French school of *Analyse des Données* (data analysis) was born. Benzécri developed a new approach to the optimal scaling, called *Correspondence Analysis*, pinpointing the geometrical interpretation and visualization of the outputs [Benzécri 1973]. A *non-symmetrical* approach to correspondence analysis was proposed by Lauro & D'Ambra [1984].

An alternative approach to the problem was yielded by the *Homogeneity Analysis*, deceived by the de Leeuw's datatheory group of the Leiden University [Michalidis & de Leeuw 1998].

The 1970s saw an increasing interest in the joint analysis of variables at different scale levels. Kruskal & Shepard [1974] proposed a non-metric approach in order to bypass the linearity hypothesis in factor analysis. Nishisato & Arri [1975] yielded a contribute in handling partially ordered variables. De Leeuw, Young and Takane published a suite of papers in which proposed non-metric methods and softwares for implementing almost all the multivariate methods (analysis of variance, principal component analysis, canonical correlation analysis and multidimensional scaling among them) on variables measured at different scale levels by ALS algorithms (see section 3.3.1). A similar technique for implementing PCA on both nominal and numeric variables was presented by Tenenhaus [1977].

In the following years the challenge was summing up all the approaches, by discussing their similarities and differences. Tenenhaus & Young [1985] synthesized the different methods for the analysis of a block of nominal variables. Young [1981] reviewed the ALS approaches to optimal scaling. Gifi [1990] made the same by resuming the large contribute of the data theory group of the Leiden University on the issue. Other overviews of Correspondence Analysis are by Greenacre [Greenacre 1984, Greenacre 2007] and Murtagh [2005], while Nishisato reviewed similar issues with the name of *Dual Scaling* [Nishisato 1980, Nishisato 2007].

Non-parametric methods allowing the quantitative analysis of a block of non-metric variables by means of a suitable scaling (or quantification) of their categories will be review in the next section. Following Nishisato [1980], we will call them Dual Scaling (DS) methods.

4.3 Dual Scaling techniques for the multivariate descriptive analysis of categorical variables

Let N units (observations) be described by a set of P categorical variables $\boldsymbol{x}_1^* \dots \boldsymbol{x}_p^* \dots \boldsymbol{x}_P^*$ with $K_1 \dots K_p \dots K_P$ categories, and with $\sum_p K_p = K$. We denote pk the generic category k of the variable p. The generic element $\widetilde{x}_{ikp}$ of the N by K_p indicator binary matrix $\widetilde{\boldsymbol{X}}_p$ associated to $\boldsymbol{x}_p^*$ is one if individual i is in category k of variable p and zero otherwise. The N by K indicator matrix $\widetilde{\boldsymbol{X}} = [\widetilde{\boldsymbol{X}}_1 \dots \widetilde{\boldsymbol{X}}_P]$ is obtained by horizontally adjoining the several $\widetilde{\boldsymbol{X}}_p$.

Any DS method looks for multiple orthogonal quantifications (scalings) for each observation and each category of $\widetilde{\boldsymbol{X}}$ (from here the name of *Dual Scaling* proposed by Nishisato [1980]). The scaling value ϕ_{pk} of category pk is a real number associated with the category. The numerical scaled variable $\widehat{\boldsymbol{x}}_p$ is the quantification of the variable $\boldsymbol{x}_p^*$

induced by the K_p scale values ϕ_{pk}, elements of the vector $\boldsymbol{\phi}_{(p)}$

$$\widehat{\boldsymbol{x}}_p = \sum_{i=1}^{K_p} \phi_{pk} \widetilde{\boldsymbol{x}}_{pk} = \widetilde{\boldsymbol{X}}_p \boldsymbol{\phi}_{(p)} \tag{4.1}$$

The scaling vector of all the K categories is denoted $\boldsymbol{\phi}' = \boldsymbol{\phi}'_{(1)} \ldots \boldsymbol{\phi}'_{(P)}$. Finally, the scaling $\boldsymbol{\psi}$ of units is defined by associating a real number ψ_i to the unit i.

The DS problem consists in searching for optimal solution for $\boldsymbol{\psi}$ and $\boldsymbol{\phi}$. It has been approached from a number of different point of view along the previous century. Here, these approaches are divided in three principal groups: the *Classic* approach, the *French* approach, and the *Dutch* approach, yielding substantially the same results, or results very closely related.

4.3.1 The classical approach

The classical approach to DS, introduced by Richardson & Kuder [1933] and fully formalized by Guttman [1941], involves the concept of *Internal Consistency* of the scaling of categories and individuals.

Internal Consistency means that:

- Globally, over the individuals, the optimal scale values of categories associated with each unit vary as little as possible;
- Globally, over the categories, the optimal scale values of units sharing the same category vary as little as possible.

More formally, Guttman's criterion consists in finding out, on a hand, a category scaling vector $\boldsymbol{\phi}$ maximizing

$$\frac{\boldsymbol{\phi}'\widetilde{\boldsymbol{X}}'\widetilde{\boldsymbol{X}}\boldsymbol{\phi}}{P\boldsymbol{\phi}'\boldsymbol{D}\boldsymbol{\phi}} \tag{4.2}$$

under the normalization requirements

$$\mathbf{1}'_K\boldsymbol{D}\boldsymbol{\phi} = 0 \text{ and } \frac{1}{NP}\boldsymbol{\phi}'\boldsymbol{D}\boldsymbol{\phi} = 1, \tag{4.3}$$

where $\boldsymbol{D} = \text{diag}(\widetilde{\boldsymbol{X}}'\widetilde{\boldsymbol{X}})$ is the diagonal matrix of the frequencies n_{pk}, and $\mathbf{1}'_K$ a vector of K ones.

On the other, an unit scaling vector $\boldsymbol{\psi}$ maximizing

$$\frac{\boldsymbol{\psi}'\widetilde{\boldsymbol{X}}\boldsymbol{D}\widetilde{\boldsymbol{X}}'\boldsymbol{\psi}}{P\boldsymbol{\psi}'\boldsymbol{\psi}} \tag{4.4}$$

The optimal scaling values are respectively the eigenvectors $\boldsymbol{\phi}_1$ and $\boldsymbol{\psi}_1$ corresponding to the dominant (non trivial) eigenvalue of the matrices

$$(1/P)\boldsymbol{D}^{-1}\widetilde{\boldsymbol{X}}'\widetilde{\boldsymbol{X}} \tag{4.5}$$

and

$$(1/P)\widetilde{\boldsymbol{X}}\boldsymbol{D}^{-1}\widetilde{\boldsymbol{X}}' \tag{4.6}$$

Each of these two matrices have identical non zero eigenvalues, all of

which are between zero and one. The number of positive common eigenvalues equals the minimum of N and $(K - P)$.

A second pair of scaling vectors (quantifications) can be obtained by maximizing 4.2 and 4.4 under the further restrictions

$$\boldsymbol{\phi}'\boldsymbol{D}\boldsymbol{\phi}_1 = 0 \text{ and } \boldsymbol{\psi}'\boldsymbol{\psi}_1 = 0 \tag{4.7}$$

The solutions to this optimization problem are the eigenvectors $\boldsymbol{\phi}_2$ and $\boldsymbol{\psi}_2$ associated to the second largest non trivial eigenvalues of matrices 4.5 and 4.6.

At the same way it is possible to extract successive quantifications $\boldsymbol{\psi}_h$ and $\boldsymbol{\phi}_h$, linked by the following transition relations

$$\boldsymbol{\psi}_h = \mu_h^{-1/2}(1/P)\widetilde{\boldsymbol{X}}\boldsymbol{\phi}_h \tag{4.8}$$

and

$$\boldsymbol{\phi}_h = \mu_h^{-1/2}\boldsymbol{D}^{-1}\widetilde{\boldsymbol{X}}'\boldsymbol{\psi}_h, \tag{4.9}$$

where μ_h is the h-th non-zero eigenvector associated to both matrices 4.5 and 4.6.

4.3.2 The French approach

The French approach to DS aims to find a category scaling vector $\boldsymbol{\phi}$ maximizing

$$\text{var}\left(\frac{1}{P}\sum_{p=1}^{P}\sum_{k_p=1}^{K_p}\phi_{pk}\boldsymbol{x}_{pk}\right) \tag{4.10}$$

This criterion can be maximized in two different ways.

The first is a Principal Component Analysis (PCA) on a suitable transformation of $\widetilde{\boldsymbol{X}}$ [Burt 1950, Benzécri 1973]. Denoting $\boldsymbol{O^c}$ the centering operator $(1/N)(\boldsymbol{I}_N - \boldsymbol{I}_N\boldsymbol{I}'_N)$, $\boldsymbol{\psi}_h$ can be obtained as the left eigenvector corresponding to the h-th largest singular value of

$$(1/N)\boldsymbol{O^c}\widetilde{\boldsymbol{X}}\boldsymbol{D}^{1/2}. \tag{4.11}$$

Moreover, if we define $\boldsymbol{a}_h$ the right eigenvector corresponding to the h-th largest singular value of the same matrix, $\boldsymbol{\psi}_h$ can be obtained as

$$\boldsymbol{\psi}_h \propto \boldsymbol{D}^{-1/2}\boldsymbol{a}_h \tag{4.12}$$

where the symbol $\propto$ indicates that the right part of the equation must be normalized to unitary norm.

Another way to maximize the criterion 4.10 is to implement a Generalized Canonical Correlation Analysis (GCCA) to the P blocks of indicator matrices $\widetilde{\boldsymbol{X}}_p$ [Bouroche, Saporta & Tenenhaus 1975]; From GCCA point of view, we wish to maximize the sum of the squared

correlations between the scaled variables and the scaling of the units. Hence, the DS problem can be reformulated as the research of a N by H matrix of observation scalings (components) $\boldsymbol{\Psi} = [\boldsymbol{\psi}_1 \dots \boldsymbol{\psi}_h \dots \boldsymbol{\psi}_H]$ and a K by H matrix of category scalings $\boldsymbol{\Phi} = [\boldsymbol{\phi}_1 \dots \boldsymbol{\phi}_h \dots \boldsymbol{\phi}_H]$ maximizing

$$\frac{1}{P} \sum_{h=1}^{H} \sum_{p=1}^{P} \text{cor}^2(\widetilde{\boldsymbol{X}}_p \boldsymbol{\phi}_{(p)h}, \boldsymbol{\psi}_h) \tag{4.13}$$

under the constraints

$$\boldsymbol{\Phi}'\boldsymbol{\Phi} = \boldsymbol{I} \text{ and } \boldsymbol{\Psi}'\boldsymbol{\Psi} = \boldsymbol{I},$$

where $\boldsymbol{I}$ indicates the identity matrix and $\widetilde{\boldsymbol{X}}_p \boldsymbol{\phi}_{(p)h}$ is the scaling of variable $\boldsymbol{x}^p$ yielded by the h-th component.

Criterion 4.13 is maximized with respect to $\boldsymbol{\phi}_{(p)h}$ for

$$\boldsymbol{\phi}_{(p)h} = (\widetilde{\boldsymbol{X}}'_p \widetilde{\boldsymbol{X}}_p)^{-1} \widetilde{\boldsymbol{X}}'_p \boldsymbol{\psi}_p, \tag{4.14}$$

which implies that, for these values of $\boldsymbol{\phi}_{(p)h}$,

$$\text{cor}^2(\widetilde{\boldsymbol{X}}_p \boldsymbol{\phi}_{(p)h}, \boldsymbol{\psi}_h) = (1/N) \boldsymbol{\psi}'_p \widetilde{\boldsymbol{X}}_p (\widetilde{\boldsymbol{X}}'_p \widetilde{\boldsymbol{X}}_p)^{-1} \widetilde{\boldsymbol{X}}'_p \boldsymbol{\psi}_h \tag{4.15}$$

and:

$$\boldsymbol{\phi}_h = \boldsymbol{D}^{-1} \boldsymbol{X}' \boldsymbol{\psi}_h \tag{4.16}$$

Therefore, criterion 4.13 can be rewritten as

$$\frac{1}{NP}\sum_{h=1}^{H}\boldsymbol{\psi}'\widetilde{\boldsymbol{X}}\boldsymbol{D}^{-1}\widetilde{\boldsymbol{X}}\boldsymbol{\psi} \tag{4.17}$$

and it is maximized by means of eigenvectors associated with the eigenvalues of 4.6.

The link between DS and GCCA allows us to apply to DS an interesting result derived by Kettenring [1971] in the context of Carroll's GCCA [Carroll 1968]: if we use normalized scaled variables, $\boldsymbol{\psi}_1$ is the first normalized principal component of $\widehat{\boldsymbol{X}}$.

A further interpretation of criterion 4.13 in terms of sum of correlation ratios was provided by Saporta [1980], who noticed the equality

$$\mathrm{cor}^2(\widetilde{\boldsymbol{X}}_p\boldsymbol{\phi}_{(p)h}, \boldsymbol{\psi}_h) = \eta^2_{(\boldsymbol{x}^*_p, \boldsymbol{\psi}_h)}, \tag{4.18}$$

where η denotes the Pearson's correlation ratio, *i.e.* the part of the variance of $\boldsymbol{\psi}_h$ explained by the categories of $\boldsymbol{x}^*_p$. As a consequence, the factors of DS are also solution of the problem

$$\frac{1}{P}\sum_{h=1}^{H}\sum_{p=1}^{P}\eta^2(\boldsymbol{x}^*_p, \boldsymbol{\psi}_h) \tag{4.19}$$

under the constraint $\boldsymbol{\Psi}'\boldsymbol{\Psi} = \boldsymbol{I}$.

4.3.3 The Dutch approach

Under the pseudonym of Albert Gifi, a group led by Jan de Leeuw, was involved with an important development of DS. This group mostly explored the use of DS as a quantification technique embedded in classical multivariate analysis to achieve nonlinear generalizations of multivariate methods. The work of the Gifi group is amply described in Gifi [1990].

According to Gifi [1990], it is possible to use two different approaches to multivariate descriptive analysis of categorical variables: HOMALS (HOMogeneity analysis by Alternating Least Squares) and PRINCALS (PRINCipal component analysis by Alternating Least Squares). HOMALS [Gifi 1990] uses the multiple approach, in which multiple transformations for each variable are found. PRINCALS [de Leeuw & Van Rijckevorsel 1980] uses the single approach, in which we look for a single transformation for each variable. In the next we focus on HOMALS, while PRINCALS approach will be deepened in the next section.

HOMALS finds multiple sets of rank-one quantifications (i.e. the best approximation in an unidimensional space) $\boldsymbol{\Phi}_1 \dots \boldsymbol{\Phi}_H$ for each variable $\boldsymbol{x}_p^*$.

Let $\boldsymbol{\Phi}_{(p)}$ be the K_p by H sub-matrix of the optimal scalings of the categories of variable $\boldsymbol{X}_p^*$ induced by the H observation scalings; HOMALS algorithm minimizes the loss function

$$\frac{1}{P} \sum_p \mathrm{SSQ}(\boldsymbol{\Psi} - \widetilde{\boldsymbol{X}}_p \boldsymbol{\Phi}_{(p)}) \tag{4.20}$$

where the operator SSQ() denotes the sum of the squares of the elements of the matrix (or vector) to which it is applied. Loss function 4.20 is minimized under the constraints $\mathbf{1}'_N\boldsymbol{\Psi} = 0$ and $\boldsymbol{\Psi}'\boldsymbol{\Psi} = N\boldsymbol{I}$ (the observation scalings must be centered and normalized to unitary variance) by means of a three-step Alternating Least Squares (ALS) algorithm.

In the first step, criterion 4.20 is minimized with respect to $\boldsymbol{\Phi}_{(p)}$ for fixed $\boldsymbol{\Phi}$. The optimal $\boldsymbol{\Phi}_{(p)}$ is

$$\boldsymbol{\Phi}_{(p)} = (\widetilde{\boldsymbol{X}}'_p\widetilde{\boldsymbol{X}}_p)^{-1}\widetilde{\boldsymbol{X}}_p\boldsymbol{\Psi} \tag{4.21}$$

In the second step, 4.20 is minimized with respect to $\boldsymbol{\Psi}$ for fixed $\boldsymbol{\Phi}_{(p)}$. The optimal $\boldsymbol{\Psi}$ is obtained as

$$\boldsymbol{\Psi} = \frac{1}{P}\sum_p \widetilde{\boldsymbol{X}}_p\boldsymbol{\Phi}_{(p)} \tag{4.22}$$

In the third step $\boldsymbol{\Psi}$ is centered and orthonormalized following the Gram-Schmidt procedure [Golub & Loan 1996].

The ALS algorithm iterates these three steps until convergence. It is possible to show (see Gifi [1990]) that also HOMALS optimal observation scaling correspond to the left eigenvectors of the Singular Value Decomposition of 4.11.

4.4 The non-metric approach to PCA

Principal Component Analysis (PCA) [Hotelling 1933] postulates that an N by P matrix $\boldsymbol{X}$ of N observations on P variables can be approximated by the bilinear structure $\boldsymbol{TP'}$ of rank H obtained as the matricial product of a N by H component score matrix $\boldsymbol{T} = [\boldsymbol{t}_1 \dots \boldsymbol{t}_H]$ and a P by H loading matrix $\boldsymbol{P} = [\boldsymbol{p}_1 \dots \boldsymbol{p}_H]$. For identification purposes, $\boldsymbol{T}$ and $\boldsymbol{P}$ are constrained such that $\boldsymbol{T'T} = N\boldsymbol{I}$ and $\boldsymbol{P'P}$ is diagonal. Hotelling's method find finds $\boldsymbol{T}$ and $\boldsymbol{P}$ such that

$$\frac{1}{P}\sum_p \sum_i (x_{ip} - \boldsymbol{t}_i'\boldsymbol{p}_p)^2 \tag{4.23}$$

is minimized under a prescribed number of components.
With notations $\boldsymbol{t}_i'$ and $\boldsymbol{p}_p$ we denote respectively the i-th row of the matrix $\boldsymbol{T}$ intended as a H order row vector and the p-th row of the matrix $\boldsymbol{P}$ intended as a H order column vector. Note that quantity 4.23 may be expressed also as

$$\frac{1}{P}\sum_p \mathrm{SSQ}(\boldsymbol{x}_p - \boldsymbol{T}\boldsymbol{p}_p) \tag{4.24}$$

The purpose of non-metric PCA is to apply certain non-linear transformation (or scaling) functions to $\boldsymbol{X}$-variables, in order to obtain a new set of transformed variables $\widehat{\boldsymbol{X}}$ that minimizes the loss function

$$\frac{1}{P}\sum_p \mathrm{SSQ}(\widehat{\boldsymbol{x}}_p - \boldsymbol{T}\boldsymbol{p}_p) \tag{4.25}$$

This quantity is a function of the model parameters (component scores and loadings) and the scaling parameters, *i.e.* the vectors $\boldsymbol{q}_1 \dots \boldsymbol{q}_P$ of the category scaling values of each variable, such that $\widetilde{\boldsymbol{X}}_p \boldsymbol{q}_p = \widehat{\boldsymbol{x}}_p$.

It's noteworthy that in this and in the following sections we denote $\boldsymbol{t}_h$ the generic score vector, whereas we noted $\boldsymbol{\psi}_h$ the generic units scaling vector in previous section. This change in notation is due to the fact that we prefer to keep traditional PCA notations, in order to yield a clearer comparison between standard and non-metric PCA methods.

Differently from the DS techniques, that require for each category of the p-th variable multiple quantifications $\boldsymbol{\Phi}_{(p)k}$ (one for each component), in non-metric PCA methods, for each category pk we have a single quantification q_{pk}.

Another important difference is that while ordinary PCA and DS are nested methods, these non-metric PCA methods are not. This means that if one requires a h-dimensional solution and then a second h'-dimensional solution, with $(h' > h)$, then the first h dimensions of the latter solution are not necessary identical to the h-dimensional solution.

An additional feature of non-metric PCA methods is that their solutions are not eigenvalue problems. So, this methods are implemented by iterative procedures such as ALS algorithms or gradient methods. Unfortunately, these algorithm may converge to local *optima* if the initialization of the parameters is not properly chosen.

In 1970's a number of methods and respective softwares were born

to implement non-metric PCA. The three more important approaches are the non-metric PCA by Kruskal and Shepard, PRINCIPALS by Young, Takane and De Leeuw, PRINQUAL by Tenenhaus and PRINCALS by De Leeuw and Van Rijckevorsel. In the next, these methods will be briefly reviewed.

4.4.1 Kruskal&Shepard's non-metric PCA

J.J Thurstone, the father of the *theory of multiple factors* [Thurstone 1931], wrote that "*one of the principal assumptions underlying factorial theory is that the scores are monotonic increasing or decreasing scores functions of the scores on the primary factors or parameters*", and indeed that the further assumption of linearity was adopted only "*as a first approximation*". Hence, "*it would probably be ... profitable to develop non-metric methods of factor analysis*" [Thurstone 1947].

Starting from this idea and from their previous studies on ordinal data [Kruskal 1964*b*, Kruskal 1964*a*, Shepard 1966], Kruskal & Shepard [1974] developed a non-metric variant of linear factor analysis. This method keeps the monotonicity assumption cited by Thurstone, but ignores the further assumption of linearity. It is based on normalized *STRESS*, which measures the departure from monotonicity

$$STRESS = \sqrt{\frac{1}{P}\sum_{p}\frac{\mathrm{SSQ}(\boldsymbol{T}\boldsymbol{p}_p - \widehat{\boldsymbol{x}}_p)}{\mathrm{SSQ}(\widehat{\boldsymbol{x}}_p - (1/N)\sum_i \widehat{x}_{pi})}} \tag{4.26}$$

This function must be minimized under the constraints $\boldsymbol{T}'\boldsymbol{T} = N\boldsymbol{I}$

and $\boldsymbol{p}_p'\boldsymbol{p}_p = 1$.

Given these constraints, except irrelevant details, *STRESS* function equals 4.25, and can be interpreted as a normalized loss function, defined by the variance of the residuals divided by the variance of the transformed data.

The method is implemented in two steps; firstly, for fixed $\boldsymbol{T}$ and $\boldsymbol{P}$, optimal $\widehat{\boldsymbol{x}}_p$ are found performing a least squares monotone regression [Kruskal 1964*b*] for each variable p. Then, for fixed values of $\widehat{\boldsymbol{x}}_p$, *STRESS* is minimized with respect to $\boldsymbol{T}$ and $\boldsymbol{P}$ using the method of gradients (see *e.g.* Kelley [1962]).

4.4.2 PRINCIPALS

PRINCIPal component analysis by Alternating Least Squares (PRINCIPALS) [Young et al. 1978] has been the first procedure able to extend PCA to the case where the variables have a variety of measurements charateristics. Some may be nominal, others ordinal and the rest interval. Furthermore, some may be discrete and others continuous.

PRINCIPALS optimizes the loss function 4.25, formulated as

$$\mathrm{tr}(\widehat{\boldsymbol{X}} - \boldsymbol{T}\boldsymbol{P}')(\widehat{\boldsymbol{X}} - \boldsymbol{T}\boldsymbol{P}') \tag{4.27}$$

under standardization contraints $\widehat{\boldsymbol{x}}_p'\mathbf{1} = 0$ and $\widehat{\boldsymbol{x}}_p'\widehat{\boldsymbol{x}}_p = N$ on the quantified variables.

PRINCIPALS is implemented by an iterative algorithm that alter-

nates model parameters estimation and scaling parameter estimation. For fixed values of $\widehat{\boldsymbol{X}}$, the loss function is optimized with respect to $\boldsymbol{T}$ and $\boldsymbol{P}$ calculated as the H largest terms of a singular value decomposition of $\widehat{\boldsymbol{X}}$. Since the quantifications are column conditional and, for fixed $\boldsymbol{T}$ and $\boldsymbol{P}$, function 4.25 is separable with respect to the optimally scaled data for each variable, the optimal scaling of data can be performed for each variable separately and independently.

This procedure is proved to be monotonically convergent (see de Leeuw et al. [1976]). Hence, in each step the value of 4.27 decreases; the procedure is arrested when the improvement of the fit is irrelevant.

4.4.3 PRINQUAL

PRINQUAL is a method (and a program) which perform a PCA on a set of nominal and numerical variable devised by Tenenhaus [1977]. PRINQUAL solves 4.25 rewritten the maximization problem of the function

$$\frac{1}{P}\sum_{p}\sum_{h}\mathrm{cor}^2(\widehat{\boldsymbol{x}}_p, \boldsymbol{t}_h) \tag{4.28}$$

under standardization contraints $\mathbf{1}'\widehat{\boldsymbol{x}}_p = 0$ and $\widehat{\boldsymbol{x}}_p'\widehat{\boldsymbol{x}}_p = N$ on the quantified variables.

Each quantification of a nominal variable is calculated as the product between a suitable standardization $\boldsymbol{\Delta}$ of its indicator matrix (see formula 4.11) and a vector of order K_p. Also this procedure is implemented by alternatively estimating scaling and model parameters.

In PRINQUALS, an optimal initial solution for $\widehat{\boldsymbol{X}}$ is obtained performing a Carrol's GCCA [Carroll 1968] on P blocks of variables, each block being composed just of a numerical variable or a nominal variable properly recoded. Then, for given $\widehat{\boldsymbol{X}}$, scaling parameters are calculated as the first canonical vector associated to $\boldsymbol{\Delta}$ in the Canonical Correlation Analysis of $\boldsymbol{\Delta}$ and $\boldsymbol{T}$. Finally, for given scaling parameters, the maximum of 4.28 with respect to $\boldsymbol{t}_h$ is obtained by the first H principal components of $\widehat{\boldsymbol{X}}$.

This procedure is shown to monotonically converge to a maximum (see Tenenhaus [1977].

4.4.4 PRINCALS

PRINCALS (PRINcipal Component analysis by means of Alternating Least Squares) algorithm [de Leeuw & Van Rijckevorsel 1980] is a generalization of HOMALS, where the set of quantifications $\boldsymbol{\Phi}_p$ for the variable p is constrained to be of unitary rank by means of the restriction

$$\boldsymbol{\Phi}_{(p)} = \boldsymbol{q}_p \boldsymbol{p}'_p \tag{4.29}$$

with $\boldsymbol{q}_p$ being the vector of the single quantifications for the categories of variable p. Moreover, $\boldsymbol{q}_p$ is constrained to be centered ($\mathbf{1}'\widetilde{\boldsymbol{X}}'_p\widetilde{\boldsymbol{X}}_p\boldsymbol{q}_p = 0$) and normalized ($\boldsymbol{q}'_p\widetilde{\boldsymbol{X}}'_p\widetilde{\boldsymbol{X}}_p\boldsymbol{q}_p = 1$) following the metric $\widetilde{\boldsymbol{X}}'_p\widetilde{\boldsymbol{X}}_p$. The introduction of the rank-1 restriction allows the existence of multidimensional solutions for the observations (unit scalings

in DS language, components in PCA language) with a single category quantification, and also makes it possible to incorporate the measurement level for the variables into the analysis.

PRINCALS optimizes HOMALS loss function 4.20 that, under the additional rank-1 restriction, becomes

$$\frac{1}{P}\sum_{p}\mathrm{SSQ}(\widetilde{\boldsymbol{X}}_p\boldsymbol{q}_p\boldsymbol{p}_p'-\boldsymbol{T}). \tag{4.30}$$

This function can be decomposed in a sum of two functions, of which the first one, called *multiple* loss, is exactly 4.20 and the latter, called *single* loss is

$$\frac{1}{P}\sum_{p}\mathrm{tr}(\boldsymbol{q}_p\boldsymbol{p}_p'-\boldsymbol{\Phi}_{(p)})\widetilde{\boldsymbol{X}}_p'\widetilde{\boldsymbol{X}}_p(\boldsymbol{q}_p\boldsymbol{p}_p'-\boldsymbol{\Phi}_{(p)})'. \tag{4.31}$$

The *single* loss corresponds to the additional loss incurred by imposing the rank-1 restriction.

PRINCALS algorithm consists in a double ALS loop: the outer loop and the inner loop. In the outer loop, corresponding to HOMALS loop, the multiple loss is minimized for fixed $\boldsymbol{\Phi}_{(p)}$. In the inner loop the *single* loss function is minimized with respect to $\boldsymbol{q}_p$ and $\boldsymbol{p}_p$. Once its optimum obtained, $\boldsymbol{q}_p$ is projected in the conic space corresponding to all possible transformations given the restriction imposed by the measurement level of the variable. In the practice, since the value of the loss function is smaller after a single iteration of the inner loop, the inner loop is iterated just one time each iteration of the outer

loop. Hence, PRINCALS algorithm can be summarized in the following steps:

1. Initialize $\boldsymbol{T}$, so that $\mathbf{1}'\boldsymbol{T} = 0$ and $\boldsymbol{T}'\boldsymbol{T} = N\boldsymbol{I}$
2. For fixed $\boldsymbol{T}$: $\boldsymbol{\Phi}_{(p)} = (\widetilde{\boldsymbol{X}}'_p\widetilde{\boldsymbol{X}}_p)^{-1}\widetilde{\boldsymbol{X}}_p\boldsymbol{T}$
3. $\boldsymbol{p}_p = (\boldsymbol{\Phi}'_p\widetilde{\boldsymbol{X}}'_p\widetilde{\boldsymbol{X}}'_p\boldsymbol{q}_p)/(\boldsymbol{q}_p\widetilde{\boldsymbol{X}}'_p\widetilde{\boldsymbol{X}}'_p\boldsymbol{q}_p)$
4. $\boldsymbol{q}_p = \boldsymbol{\Phi}_{(p)}\boldsymbol{p}_p/(\boldsymbol{p}'_p\boldsymbol{p}_p)$
5. Account for the measurement level of the p-th variable by performing a suitable conic regression
6. $\boldsymbol{\Phi}_{(p)} = \boldsymbol{q}_p\boldsymbol{p}'_p$
7. $\boldsymbol{T} = (1/P)\sum_p \widetilde{\boldsymbol{X}}_p\boldsymbol{\Phi}_{(p)}$
8. Center and orthonormalize $\boldsymbol{T}$
9. Check the convergence criterion

In PRINCALS, for each variable the preferred level of analysis can be chosen: multiple nominal, single nominal, ordinal or numerical. PRINCALS is a generalization of all previous methods, and in particular, if all the variables are analyzed at a multiple nominal level (*i.e.* if we search for multiple quantifications of the variable), it yields the same results as HOMALS.

4.5 A new PLS method for non-metric PCA: the Non-Metric NIPALS

Non-Metric NIPALS (NM-NIPALS) is a new algorithm developed by the author that performs a non-metric PCA on a N by P matrix $\boldsymbol{X}^*$ representing a mixed set of P variables observed at different measurement scales on N units.

In NM-NIPALS, each raw variable $\boldsymbol{x}_p$ is transformed in a interval variable $\widehat{\boldsymbol{x}}_p$ by means of an OS procedure implemented by an algorithm of the PLS family. This procedure depends on the scaling level at which one want to analyze each variable. The scaling level is strictly connected to the relation we suppose to exist between the variable and the first component. If we make no assumption about this relation (or if we simply can not make any assumption, because we are handling a nominal raw variable), then the suitable scaling level of the analysis is the nominal one. If we deal with a raw variable measured on an ordinal or interval measurement scale, and we suppose a non functional but monotonic relation, we can analyze the variable at ordinal scaling level. In these two cases we do a non-metric analysis of the variable: we just preserve its category ordering and/or grouping properties. Finally, NM-NIPALS can model a non linear functional relation between a numeric raw variable and the first PC by means of a polynomial rule of degree D. In the particular case where $D = 1$, it means that all of the assumptions of PCA model are verified (see section 3.1) and a standard linear analysis is allowed. In this last case, we keep in the

analysis all of the measurement properties of the variable.

In any case we assume that the measurement process is discrete, *i.e.* that the variable can assume a finite number of different values that corresponds to the number of different observed values.

NM-NIPALS scalings can be defined optimal, as they maximize the same criterion of a PCA.

4.5.1 The algorithm

First Principal Component (PC) can be defined as the linear combination $\boldsymbol{t}_1 = \boldsymbol{X}\boldsymbol{p}_1$ of the variables the most correlated with the variables themselves. In mathematical terms, the first PC maximizes

$$\sum_p \text{cor}^2(\boldsymbol{x}_p, \boldsymbol{X}\boldsymbol{p}_1) \tag{4.32}$$

with respect to $\boldsymbol{p}_1$, under the restriction $\boldsymbol{p}_1'\boldsymbol{p}_1 = 1$.

In non-metric approach to NIPALS we find a matrix $\widehat{\boldsymbol{X}}$ of quantified variables maximizing

$$\sum_p \text{cor}^2(\widehat{\boldsymbol{x}}_p, \widehat{\boldsymbol{X}}\boldsymbol{p}_1) \tag{4.33}$$

under the constraints $\boldsymbol{p}_1'\boldsymbol{p}_1 = 1$, $\mathbf{1}'\widehat{\boldsymbol{x}}_p = 0$ and $\widehat{\boldsymbol{x}}_p'\widehat{\boldsymbol{x}}_p = N$.

Criterion 4.33 is a function of the quantified variables $\widehat{\boldsymbol{x}}_p$ and the model parameters $\boldsymbol{p}_1$ and $\boldsymbol{t}_1$, where further restrictions on the quantification of each variable depend on the chosen scaling level.
The optimal value for $\boldsymbol{p}_1$, given $\widehat{\boldsymbol{x}}_p$, is the right singular vector corre-

sponding to the dominant singular value of $\widehat{\boldsymbol{X}}$. This value, in standard NIPALS algorithm, is iteratively found using the power method principle (see section 1.2).

For fixed $\boldsymbol{t}_1 = \widetilde{\boldsymbol{X}}\boldsymbol{p}_1$, function 4.33 is separable with respect to the optimal scaled data for each variable $\widehat{\boldsymbol{x}}_1 \dots \widehat{\boldsymbol{x}}_P$. Criterion 4.33 can be decomposed into a sum of components each of which is a function only of the scaling parameters of one variable. Hence, the problem can be solved by separately maximizing each function

$$\text{cor}^2(\widehat{\boldsymbol{x}}_p, \boldsymbol{t}_1) \tag{4.34}$$

with respect to $\widehat{\boldsymbol{x}}_p$.

Remembering the discussion in section 3.4.1, and considering $\boldsymbol{t}_1$ as our LC, optimal $\widehat{\boldsymbol{x}}_p$ can be found as one of the following conic projections, where the type of cone depends on the restriction imposed by the chosen scaling level of the analysis for each variable. In particular, for each $\widehat{\boldsymbol{x}}_p$ analyzed at nominal scaling level we have

$$\widetilde{\mathcal{Q}}(\boldsymbol{x}_p^*, \boldsymbol{t}_1) : \widehat{\boldsymbol{x}}_p \propto \widetilde{\boldsymbol{X}}_p(\widetilde{\boldsymbol{X}}_p'\widetilde{\boldsymbol{X}}_p)^{-1}\widetilde{\boldsymbol{X}}_p'\boldsymbol{t}_1 \tag{4.35}$$

while, for each $\widehat{\boldsymbol{x}}_p$ analyzed at ordinal scaling level the quantification function is

$$\widetilde{\widetilde{\mathcal{Q}}}(\boldsymbol{x}_p^*, \boldsymbol{t}_1) : \widehat{\boldsymbol{x}}_p \propto \widetilde{\widetilde{\boldsymbol{X}}}_p(\widetilde{\widetilde{\boldsymbol{X}}}_p'\widetilde{\widetilde{\boldsymbol{X}}}_p)^{-1}\widetilde{\widetilde{\boldsymbol{X}}}_p'\boldsymbol{t}_1 \tag{4.36}$$

To conclude, for each $\widehat{\boldsymbol{x}}_p$ analyzed at a functional scaling level we have

$$\dot{\mathcal{Q}}(\boldsymbol{x}_p^*, \boldsymbol{t}_1) : \widehat{\boldsymbol{x}}_p \propto \dot{\boldsymbol{X}}_p (\dot{\boldsymbol{X}}_p' \dot{\boldsymbol{X}}_p)^{-1} \dot{\boldsymbol{X}}_p' \boldsymbol{t}_1 \tag{4.37}$$

Symbol $\propto$ in equations 4.35, 4.36 and 4.37 means that the left part of the equation is standardized to unitary variance.

In criterion 4.34 quantified variables are functions of $\boldsymbol{t}_1 = \widetilde{\boldsymbol{X}} \boldsymbol{p}_1$. Since $\boldsymbol{t}_1$, on its turn, is a function of the quantified variables $\widehat{\boldsymbol{x}}_p$, the only way to obtain both is estimating the quantifications simultaneously with the estimation of the NIPALS parameters by alternating maximization of criterion 4.34 with respect to scaling and model parameters.

The problem is solved by NM-NIPALS algorithm, in which a quantification step is added to the classic NIPALS loop.

In the first step of the NM-NIPALS loop the chosen quantification function is applied to each raw variable. In this quantification step, for fixed model parameters, each optimal $\widetilde{\boldsymbol{x}}_p$ is calculated as a function of $\boldsymbol{t}_1$ and one among the matrices $\widetilde{\boldsymbol{X}}_p$, $\widetilde{\widetilde{\boldsymbol{X}}}_p$ and $\dot{\boldsymbol{X}}_p$, depending on the scaling restrictions. Then, for fixed $\widetilde{\boldsymbol{x}}_p$, optimal model parameters are obtained one as function of the other. Since in each iteration the value of criterion 4.34 will be larger after the quantification step, the procedure will monotonically converge to a *maximum*. This procedure yields the first PC and the optimally quantified variable matrix $\widetilde{\boldsymbol{X}}$.

Further components are added by implementing the standard NIPALS iteration to the deflated matrix $\boldsymbol{E}_1 = \widehat{\boldsymbol{X}} - \boldsymbol{t}_1 \boldsymbol{p}_1'$.

The pseudo-code of NM-NIPALS algorithm is shown in algorithm

9.

Algorithm 9 NM-NIPALS algorithm

Input: $\boldsymbol{X}^*$
Output: $\boldsymbol{P}_H = [\boldsymbol{p}_1, \ldots, \boldsymbol{p}_H], \boldsymbol{T}_H = [\boldsymbol{t}_1, \ldots, \boldsymbol{t}_H], \widehat{\boldsymbol{X}}$
  Step 1.0: Initialize $\boldsymbol{t}_1$
  Step 1.1:
  repeat
    Step 1.1.1: $\widehat{\boldsymbol{x}}_p = \mathcal{Q}(\boldsymbol{x}_p^*, \boldsymbol{t}_1)$
    Step 1.1.2: $\widehat{\boldsymbol{X}} = [\widehat{\boldsymbol{x}}_1 \ldots \widehat{\boldsymbol{x}}_P]$
    Step 1.1.3: $\boldsymbol{p}_1 = \widehat{\boldsymbol{X}}' \boldsymbol{t}_1 / (\boldsymbol{t}_1' \boldsymbol{t}_1)$
    Step 1.1.4: $\boldsymbol{p}_1 = \boldsymbol{p}_1 / \|\boldsymbol{p}_1\|$
    Step 1.1.5: $\boldsymbol{t}_1 = \widehat{\boldsymbol{X}} \boldsymbol{p}_1 / (\boldsymbol{p}_1' \boldsymbol{p}_1)$
  until convergence of $\boldsymbol{p}_1$
  Step 1.2: $\boldsymbol{E}_1 = \widehat{\boldsymbol{X}} - \boldsymbol{t}_1 \boldsymbol{p}_1^{'}$
  for all $h = 2, \ldots, H$ **do**
    Step 2.0: Initialize $\boldsymbol{t}_h$
    Step 2.1:
    repeat
      Step 2.1.1: $\boldsymbol{p}_h = \boldsymbol{E}_{h-1}^{'} \boldsymbol{t}_h / (\boldsymbol{t}_h^{'} \boldsymbol{t}_h)$
      Step 2.1.2: $\boldsymbol{p}_h = \boldsymbol{p}_h / \|\boldsymbol{p}_h\|$
      Step 2.1.3: $\boldsymbol{t}_h = \boldsymbol{E}_{h-1} \boldsymbol{p}_h / (\boldsymbol{p}_h^{'} \boldsymbol{p}_h)$
    until convergence of $\boldsymbol{p}_h$
    Step 2.2: $\boldsymbol{E}_h = \boldsymbol{E}_{h-1} - \boldsymbol{t}_h \boldsymbol{p}_h^{'}$
  end for

4.5.2 Interpretation of the outputs

Non-metric PCA implemented by NM-NIPALS yields the same results of a metric PCA on the optimally scaled variables implemented

by NIPALS. Thus, all of the relations between the quantified variables, which are measured on an interval scale, and the factors can be interpreted and represented as in the classic PCA. However, this rule has an exception, as if $\boldsymbol{x}_p^*$ is analyzed at a nominal scaling level, $\text{cor}(\widehat{\boldsymbol{x}}_p, \boldsymbol{t}_1)$ is positive by construction. As a consequence, relation between the factor and such a variable can be interpreted in terms of intensity, but not in terms of sign. This is due to the fact that, for non-metric scaling analyses, statistical relation between $\boldsymbol{t}_1$ and $\widehat{\boldsymbol{x}}_p$ can be nicely interpreted also in terms of the statistical relation between the factor and the original raw variable. In fact, if a raw variable $\boldsymbol{x}_p^*$ is quantified at nominal scale level, the following relation holds

$$\text{cor}(\widehat{\boldsymbol{x}}_p, \boldsymbol{t}_1) = \eta_{(\boldsymbol{t}_1|\boldsymbol{x}_p^*)}, \tag{4.38}$$

where η is the Pearson's correlation ratio, *i.e.* the part of variability of $\boldsymbol{t}_1$ explicated by the categories of $\boldsymbol{x}_p^*$.

Since $0 \leq \eta \leq 1$, this correlation is always not negative. This implies that the relation of a quantified variable $\widehat{\boldsymbol{x}}_p$ generated by a nominal scaling level analysis with the first PC can be interpreted in terms of intensity, but not in terms of sign. This makes sense, as it is conceptually wrong expecting a sign in the relation between a numerical variable (the score vector) and a nominal variable, since a nominal variable neither increases, nor decreases.

If a variable is quantified at an ordinal level, instead, the sign of corresponding weight can be interpreted as an index of approaching

to monotonicity. In fact, the following equations hold

$$\text{cor}(\widehat{\boldsymbol{x}}_p, \boldsymbol{t}_1) = \begin{cases} \sqrt{1 - STRESS^2_{(\boldsymbol{t}_1, \boldsymbol{x}^*_p)}} & \text{if } \text{cor}(\widehat{\boldsymbol{x}}_p, \boldsymbol{t}_1) \geq 0 \\ -\sqrt{1 - STRESS^2_{(\boldsymbol{t}_1, \boldsymbol{x}^*_p)}} & \text{if } \text{cor}(\widehat{\boldsymbol{x}}_{pq}, \boldsymbol{t}_1) < 0 \end{cases}$$

These equations show that, when $\boldsymbol{x}^*_p$ is analyzed at a nominal scaling level:

- if p_{1p} approaches to one, relation between $\boldsymbol{x}^*_p$ and $\boldsymbol{t}_1$ is increasing monotone;
- if p_{1p} approaches to zero, relation between $\boldsymbol{x}^*_p$ and $\boldsymbol{t}_1$ is absolutely non-monotone;
- if p_{1p} approaches to minus one, relation between $\boldsymbol{x}^*_p$ and $\boldsymbol{t}_1$ is decreasing monotone.

4.5.3 Links with other non-metric approaches to PCA

If all the variables are analyzed at a nominal scaling level, NM-NIPALS solution for $\boldsymbol{t}_1$ corresponds, for less than a scale factor, to dominant eigenvector of $\widetilde{\boldsymbol{X}}\boldsymbol{D}^{-1}\widetilde{\boldsymbol{X}}'$. As a consequence, NM-NIPALS first component equals, for less then a scale factor, to the one dimensional solution for the unit scaling $\boldsymbol{\psi}_1$ in all the methods discussed in section 4.3.

In fact, if all of the variables are nominal, step 1.1.1 of algorithm

9, for each p, becomes

$$\widehat{\boldsymbol{x}}_p \propto \widetilde{\boldsymbol{X}}_p(\widetilde{\boldsymbol{X}}'_p\widetilde{\boldsymbol{X}}_p)^{-1}\widetilde{\boldsymbol{X}}'_p\boldsymbol{t}_1$$

For these scaling values

$$\text{cor}^2(\boldsymbol{t}_1, \widehat{\boldsymbol{x}}_p) \propto \boldsymbol{t}'_1\widetilde{\boldsymbol{X}}_p(\widetilde{\boldsymbol{X}}'_p\widetilde{\boldsymbol{X}}_p)^{-1}\widetilde{\boldsymbol{X}}'_p\boldsymbol{t}_1$$

and criterion 4.33 can be written

$$\sum_p \boldsymbol{t}'_1\widetilde{\boldsymbol{X}}_p(\widetilde{\boldsymbol{X}}'_p\widetilde{\boldsymbol{X}}_p)^{-1}\widetilde{\boldsymbol{X}}'_p\boldsymbol{t}_1$$

which is equal to

$$\boldsymbol{t}'_1\widetilde{\boldsymbol{X}}\boldsymbol{D}^{-1}\widetilde{\boldsymbol{X}}'\boldsymbol{t}_1$$

When the analysis is generalized at a variety of scaling levels, NM-NIPALS considers well known restrictions, as:

- In nominal scaling analysis, we retrieve the restrictions used by the discrete-nominal option in PRINCALS and PRINCIPALS and by PRINQUALS quantification.
- In ordinal scaling analysis, we retrieve the restriction used by the discrete-ordinal option in PRINCALS and PRINCIPALS, and by secondary approach to monotonicity in Kruskal&Shepard's non-metric PCA.
- In linear scaling analysis, we retrieve the restrictions implicated in standard PCA. Other non-linear functional restrictions can

be introduced by means of polynomial rule, as proposed in de Leeuw et al. [1976] and in Young [1981].

Since both NM-NIPALS solution and the one dimensional solution of methods discussed in section 4.4 maximize PCA criterion, on equal scaling levels, these methods yields the same results for less than a constant factor due to the different normalization applied to the scaling functions.

Chapter 5

Non-Metric PLS Regression

5.1 Motivation

One of the main applications of PLS-R is consumer preference analysis. The aim of consumer preferences analysis is the prediction of consumer preferences from the product attributes, as well as mapping consumer preferences on factorial planes. Typically, data-set consists of two sets of variables organized in two matrices. In the first matrix the variables are the preferences of the consumers or not trained judges, expressed as rating or ranking, for each of N products. In the other matrix chemical or sensorial characteristics (attributes) are measured on the products. Alternatively, it is a design matrix where a suite of qualitative attributes are observed on the products. Many of these attributes are qualitative, such as packaging, color, shape, marketing strategies and so on. So, these variable must be quan-

titatively coded in order to be introduced in the analysis. This is commonly done by using the binary coding, where each variable is replaced by its indicator matrix (see section 3.2). As a matter of fact, even preference variables should not be handled as numeric variables, as, to be rigorous, they are just ordinal variables and they do not have metric properties. Moreover, serious non-monotonicity problems arise when Just-About-Right (JAR) scales [Rothman & Parker 2009] are used to assess consumer expectations of a product attribute [Xiong & Meullenet 2008].

Here we propose a new method, called Non-Metric PLS Regression (NM-PLSR), as a flexible and comprehensive tool for performing a PLS-R on non-metric data in consumer preference analysis and other fields, such as the analysis of genetic marker-phenotype relationships.

NM-PLSR is an Optimal Scaling method based on PLS Regression algorithm. It allows us to analyze predictor and response variables measured on a variety of measurement scales, as well to handle non-linear relations in PLS-R framework.

5.2 State of the art

The treatment of non-metric variables has been explored in PLS-R framework almost always in classification problems. A number of proposals exist in literature for using adjusted PLS-R with the aim to discriminate categories of a non-metric response from a set of quantitative predictors [Tenenhaus, Giron, Viennet, Bera, Saporta

& Fertil 2007, Bastien, Esposito Vinzi & Tenenhaus 2005, Fort & Lambert-Lacroix 2005, Barker & Rayens 2003]. However, at author's knowledge, there not exists a comprehensive approach aimed to the analysis of predictor and response variables measured on a variety of measurement scales in PLS framework. In this more general case, non-metric variables are usually replaced by a suitable indicator matrix.

Another issue very discussed in PLS-R literature is non linearity. Several approaches have been proposed to provide non-linear models that retain the properties of a linear PLS Regression. Some of these approaches work on the functional form of the inner relation; others work on a suitable transformation of the predictor variables.

Methods belonging to the first family are based on a non-linear inner relation linking the predictor PLS components with the response PLS components:

$$\widehat{\mathbf{u}}_h = f(\mathbf{t_h}) \tag{5.1}$$

Various forms have been proposed for $f(\mathbf{t_h})$, such as a quadratic form [Wold, Kettaneh-Wold & Skagerberg 1989, Baffi, Martin & Morris 1999a, Höskuldsson 1992], a smoothing procedure [Frank 1990], a spline function [Wold 1992] and a neural network [Qin & McAvoy 1992, Baffi et al. 1999a]. A more general approach, based on the principles of continuum regression, was proposed in Taavitsainen & Korhonen [1992] and Haario & Taavitsainen [1994].

The alternative way to cope with non-linearity in PLS-R is based on the transformation of the explanatory variables [Rosipal & Trejo 2001,

Durand & Sabatier 1997, Berglund & Wold 1997, Berglund, Kettaneh, Uppgard, Wold, Bendwell & Cameron 2001, Durand 2001]. All methods following the latter approach are based on a priori transformation function of the predictors. Rosipal applied the theory of kernel-based learning to PLS. According to the Durand's proposals, predictors are transformed by spline functions, while Berglund's proposals involve the transformation of quantitative variables to a set of dichotomous variables similar to the binary coding of qualitative variables. For an extensive review of most of these methods, see Vivien [2002].

Optimal scaling has been proposed in non-linear approaches to OLS Regression [Young et al. 1976, Breiman & Friedman 1985] Canonical Correlation Analysis [Young et al. 1976], as well as in Redundancy Analysis [Israels 1984]. Lovaglio proposed a new methodology for a non-parametric factorial modeling of two blocks of variables linked by a causal relation, which works also with variables observed on different measurement scales [Lovaglio 2001, Lovaglio 2002]. The author was the first to propose using optimal scaling approach in PLS-R, in order to overcome drawbacks due to binary coding: the PLS algorithm for CAtegorical Predictors (PLS-CAP). This approach is useful for handling nominal predictors [Russolillo & Lauro 2010], as well non linearity [Russolillo, Trinchera & Esposito Vinzi 2009b], in PLS Regression. PLS-CAP was born to handle categorical predictors in a PLS regression. It involves an optimal quantification of predictors, finding out scaling parameters that maximize covariance between the first PLS components in predictor and response space. PLS-CAP as-

signs a numerical value to each bin of the variable properly discretized. This implies that the number of predictors does not change when variables are tranformed. Another advantage of this technique is that the trasformation of the variables is internal to the algorithm and not a priori. As a consequence, transformations are more prediction oriented and coherent with the model.

However, PLS-CAP can handle predictors only at a nominal scaling level. Here, we propose an extension of PLS-CAP, which can analyze both non-metric predictor and response variables at a variety of different measurement level: the Non-Metric PLS Regression (NM-PLSR). NM-PLSR can handle also metric variables affected by non linearity problem by non-metric quantifications, as well as by polynomial transformations of metric variables.

5.3 Non-Metric PLS Regression optimization criterion

Let two blocks $\boldsymbol{Y}^*$ of and $\boldsymbol{X}^*$ of raw variables measured on a variety of measurement scales. The generic variable of $\boldsymbol{Y}^*$ is $\boldsymbol{y}_r^*$ (with $r = 1 \dots R$), while the generic variable of of $\boldsymbol{X}^*$ is $\boldsymbol{x}_p^*$ (with $p = 1 \dots P$).

NM-PLSR finds out optimally scaled data matrices $\widehat{\boldsymbol{X}}$ and $\widehat{\boldsymbol{Y}}$ maximizing criterion

$$\text{cov}^2(\widehat{\boldsymbol{Y}}\boldsymbol{c}_1, \widehat{\boldsymbol{X}}\boldsymbol{w}_1) \tag{5.2}$$

Criterion 5.2 depends on two sets of parameters; the first set consists of model parameters, constrained to unitary norm ($\|\boldsymbol{w}_1\| = 1$ and $\|\boldsymbol{c}_1\| = 1$). The other set consists of scaling parameters, constrained to the restrictions due to the scaling level chosen for each variable and normalized to unitary variance ($\text{var}(\widehat{\boldsymbol{y}}_r) = 1$ and $\text{var}(\widehat{\boldsymbol{x}}_p) = 1$).

The aim of NM-PLSR is to maximize criterion 5.2 with respect to model and scaling parameters. It is noteworthy that, since criterion 5.2 involves just the first component parameters, NM-PLSR logic can be applied also to methods discussed in section 1.3.10.

In the next, the problem of the maximization of 5.2 will be solved with respect to model parameters, keeping fixed the scaling parameters; afterwards, we will solve the same problem with respect to the scaling parameters, keeping fixed the model parameters. In order to find out a global optimization, we will propose a modified PLS loop, which is he core of the MN.PLSR algoritm.

The maximization of 5.2 for fixed scaling parameters is yielded by the classic PLS-R solution (see section 1.3.5), given by the eigenvectors $\boldsymbol{w}_1$ and $\boldsymbol{c}_1$ corresponding to the common largest eigenvalue $\lambda = f(\boldsymbol{w}, \boldsymbol{c})$ verifying the equations

$$(\widehat{\boldsymbol{Y}}'\widehat{\boldsymbol{X}}\widehat{\boldsymbol{X}}'\widehat{\boldsymbol{Y}})\boldsymbol{c}_1 = \lambda \boldsymbol{c}_1 \tag{5.3}$$

and

$$(\widehat{\boldsymbol{X}}'\widehat{\boldsymbol{Y}}\widehat{\boldsymbol{Y}}'\widehat{\boldsymbol{X}})\boldsymbol{w}_1 = \lambda \boldsymbol{w}_1 \tag{5.4}$$

For given model parameters, instead, the problem involves searching for the largest $\lambda = f(\widehat{\boldsymbol{Y}}, \widehat{\boldsymbol{X}})$ among all the dominant eigenvalues resulting from the eigenvalue analysis of the matrices obtained from different quantifications respecting the constraints due to the level of the scaling for each variable. In order to find optimal scaling parameters for the response variables, we maximize $\lambda = f(\widehat{\boldsymbol{Y}}, \widehat{\boldsymbol{X}})$, for fixed $\boldsymbol{Y}$-weights in 5.3. It can be easily done maximizing, for each p

$$\mathrm{cor}^2(\widehat{\boldsymbol{x}}_p, \boldsymbol{u}_1), \tag{5.5}$$

where, following the notations in section 1.3, $\boldsymbol{u}_1$ is the first PLS-R component in the response space. In fact,

$$\begin{aligned}
\lambda &= \boldsymbol{c}_1' \widehat{\boldsymbol{Y}}' \widehat{\boldsymbol{X}} \widehat{\boldsymbol{X}}' \widehat{\boldsymbol{Y}} \boldsymbol{c}_1 \\
&= \boldsymbol{u}_1' \widehat{\boldsymbol{X}} \widehat{\boldsymbol{X}}' \boldsymbol{u}_1 \\
&= (\widehat{\boldsymbol{X}}' \boldsymbol{u}_1)^2 \\
&= \sum_p \mathrm{cov}^2(\widehat{\boldsymbol{x}}_p, \boldsymbol{u}_1) \\
&= \sum_p \mathrm{cor}^2(\widehat{\boldsymbol{x}}_p, \boldsymbol{u}_1) \mathrm{var}(\widehat{\boldsymbol{x}}_p) \mathrm{var}(\boldsymbol{u}_1)
\end{aligned}$$

Since $\mathrm{var}(\widehat{\boldsymbol{x}}_p) = 1$ and $\mathrm{var}(\boldsymbol{u}_1)$ is fixed with respect to the sum, optimal solutions for scaling parameter are the ones maximizing

$$\lambda = \sum_p \mathrm{cor}^2(\widehat{\boldsymbol{x}}_p, \boldsymbol{u}_1) \tag{5.6}$$

For fixed $\boldsymbol{u}_1 = \widetilde{\boldsymbol{Y}}\boldsymbol{c}_1$, this function is separable with respect to the optimal scaled data for each variable $\widehat{\boldsymbol{x}}_p$.

Criterion 5.6 can be decomposed into a sum of components each of which is a function only of the scaling parameters of one variable. Hence, the problem can be solved by separately maximizing with respect to $\widehat{\boldsymbol{x}}_p$ the squared correlation of each quantified predictor and the first PLS component in $\boldsymbol{Y}$-space, taking into account the level of scaling analysis for $\widehat{\boldsymbol{x}}_p$.

Starting from equation 5.3, a specular reasoning can be done in order to find the optimal quantifications for response variables, leading to the maximization of each

$$\text{cor}^2(\widehat{\boldsymbol{y}}_r, \boldsymbol{t}_1). \tag{5.7}$$

with respect to $\widehat{\boldsymbol{y}}_r$.

Optimization of the quantifications in PLS-R can be interpreted also as a Least Squares minimization of scaling parameters with respect to a LC, which depends on the role played by the variable. In the quantification of a predictor variable the LC is the component $\boldsymbol{u}_1$, and the aim is to minimize the least squares function

$$(\widehat{\boldsymbol{x}}_p - \boldsymbol{u}_1)'(\widehat{\boldsymbol{x}}_p - \boldsymbol{u}_1). \tag{5.8}$$

In the quantification of a response variable, the LC is the component

$\boldsymbol{t}_1$, and the objective becomes the minimization of

$$(\widehat{\boldsymbol{y}}_r - \boldsymbol{t}_1)'(\widehat{\boldsymbol{y}}_r - \boldsymbol{t}_1). \tag{5.9}$$

This specular treatment is due to the fact that the single component PLS-R model is symmetric. In fact, PLS-R model becomes asymmetric only when successive latent dimensions are computed, as both predictors and responses are deflated on the components in the $\boldsymbol{X}$-space.

The optimization of criterion 5.5 must be constrained to the scaling level chosen for the quantification. In order to obtain scalings which are coherent with the scaling level of the analysis, we use the scaling functions discussed in section 3.4.1.

If one wants to quantify a predictor $\boldsymbol{x}_p$ at a nominal scaling level, the corresponding indicator matrix $\widetilde{\boldsymbol{X}}_p$ must be computed. Then, the following scaling function have to be used

$$\widetilde{\mathcal{Q}}(\boldsymbol{x}_p^*, \boldsymbol{u}_1) : \widehat{\boldsymbol{x}}_p \propto \widetilde{\boldsymbol{X}}_p(\widetilde{\boldsymbol{X}}_p'\widetilde{\boldsymbol{X}}_p)^{-1}\widetilde{\boldsymbol{X}}_p'\boldsymbol{u}_1. \tag{5.10}$$

If one wants to quantify a predictor at a ordinal scaling level, the following scaling function have to be applied

$$\widetilde{\widetilde{\mathcal{Q}}}(\boldsymbol{x}_p^*, \boldsymbol{u}_1) : \widehat{\boldsymbol{x}}_p \propto \widetilde{\widetilde{\boldsymbol{X}}}_p(\widetilde{\widetilde{\boldsymbol{X}}}_p'\widetilde{\widetilde{\boldsymbol{X}}}_p)^{-1}\widetilde{\widetilde{\boldsymbol{X}}}_p'\boldsymbol{u}_1. \tag{5.11}$$

Finally, if we know *a priori* that relation between a numerical predictor and $\boldsymbol{u}_1$ can be modeled by means of a polynomial rule, we have to

choose the transformation function

$$\dot{\mathcal{Q}}(\boldsymbol{x}_p^*, \boldsymbol{u}_1) : \widehat{\boldsymbol{x}}_p \propto \dot{\boldsymbol{X}}_p(\dot{\boldsymbol{X}}'_p\dot{\boldsymbol{X}}_p)^{-1}\dot{\boldsymbol{X}}'_p\boldsymbol{u}_1. \tag{5.12}$$

It is noteworthy that choosing a first order polynomial means to suppose that standard PLS-R linearity requirements are respected, and the effect of the transformation 5.12 is just the standardization of the raw variable $\boldsymbol{x}_p^*$.

The optimal quantification process regarding the responses variables is analogous, but referred to the component $\boldsymbol{t}_1$. The optimal quantification of $\boldsymbol{y}_r^*$ is the one minimizing the Least Squares function

$$(\widehat{\boldsymbol{y}}_r - \boldsymbol{t}_1)'(\widehat{\boldsymbol{y}}_r - \boldsymbol{t}_1). \tag{5.13}$$

respecting the normalization constraint $\text{var}(\widehat{\boldsymbol{y}}_r) = 1$ and the constraint linked to the scaling level.

For each response analyzed at a nominal scaling level, indicator matrix $\widetilde{\boldsymbol{Y}}_r$ has to be computed. Then, each response variable $\boldsymbol{y}_r^*$ is optimally quantified by means of the quantification function

$$\widetilde{\mathcal{Q}}(\boldsymbol{y}_r^*, \boldsymbol{t}_1) : \widehat{\boldsymbol{y}}_r \propto \widetilde{\boldsymbol{Y}}_r(\widetilde{\boldsymbol{Y}}'_r\widetilde{\boldsymbol{Y}}_r)^{-1}\widetilde{\boldsymbol{Y}}'_r\boldsymbol{t}_1 \tag{5.14}$$

For each response analyzed at an ordinal scaling level analysis, we compute matrix $\widetilde{\widetilde{\boldsymbol{Y}}}_r$ by merging adjacent columns of $\widetilde{\boldsymbol{Y}}_r$ representing categories whose quantification does not respect the order constraint.

Then, we use the quantification function

$$\widetilde{\widetilde{\mathcal{Q}}}(\boldsymbol{y}_r^*, \boldsymbol{t}_1) : \widehat{\boldsymbol{y}}_r \propto \widetilde{\widetilde{\boldsymbol{Y}}}_r (\widetilde{\widetilde{\boldsymbol{Y}}}_r' \widetilde{\widetilde{\boldsymbol{Y}}}_r)^{-1} \widetilde{\widetilde{\boldsymbol{Y}}}_r' \boldsymbol{t}_1. \tag{5.15}$$

For each response analyzed at a functional scaling level analysis, the transformation function

$$\dot{\mathcal{Q}}(\boldsymbol{y}_r^*, \boldsymbol{t}_1) : \widehat{\boldsymbol{y}}_r \propto \dot{\boldsymbol{Y}}_r (\dot{\boldsymbol{Y}}_r' \dot{\boldsymbol{Y}}_r)^{-1} \dot{\boldsymbol{Y}}_r' \boldsymbol{t}_1 \tag{5.16}$$

is applied to variable $\boldsymbol{y}_r^*$.

5.4 Non-Metric PLS Regression algorithm

In the last section optimally quantified variables have been showed to be functions of the first PLS component in predictor and response spaces. Since the PLS component is, on its turn, a function of the quantified variables, it is not possible to obtain both by means of a one-step algorithm.

The Non-Metric PLS Regression (NM-PLSR) algorithm is a modified PLS-R algorithm where the first component is obtained by means of a new loop involving computation of both scaling and model parameters.

NM-PLSR loop starts with the quantification of each predictor that is optimal with respect to an initial vector $\boldsymbol{u}_1$ by means of the quan-

tification functions 5.10-5.12. Once obtained a first approximation of the matrix of the quantified predictors $\widehat{\boldsymbol{X}}$, $\boldsymbol{w}_1$ is calculated as a function of $\widehat{\boldsymbol{X}}$ and $\boldsymbol{u}_1$. After having obtained $\boldsymbol{t}_1$ as function of $\widehat{\boldsymbol{X}}$ and $\boldsymbol{w}_1$, $\boldsymbol{Y}$-variables can be optimally quantified with respct to $\boldsymbol{t}_1$ by means of the quantification functions 5.14-5.16. Then, weights $\boldsymbol{c}_1$ are computed as a function of $\widehat{\boldsymbol{Y}}$ and $\boldsymbol{t}_1$. Finally, the vector $\boldsymbol{u}_1$ is computed as a function of $\widehat{\boldsymbol{Y}}$ and $\boldsymbol{c}_1$ and the loop can restart until convergence.

At the convergence, this iterative process yields in output both model parameters $\boldsymbol{w}_1$, $\boldsymbol{c}_1$, $\boldsymbol{t}_1$ and $\boldsymbol{u}_1$, and quantified variable matrices $\widehat{\boldsymbol{X}}$ and $\widehat{\boldsymbol{Y}}$. MN-PLSR algorithm continues by regressing $\widehat{\boldsymbol{X}}$ and $\widehat{\boldsymbol{Y}}$ on $\boldsymbol{t}_1$. Residuals matrices $\widehat{\boldsymbol{E}}_1$ and $\widehat{\boldsymbol{F}}_1$ are then entered in the standard PLS-R loop in order to extract a second set of component, and so on.

The pseudo-code of NM-PLSR algorithm is provided in algorithm 10. This code is based on PLS-R algorithm described in Tenenhaus [1998] and showed in algorithm 3; however, the version presented in Höskuldsson [1988] (see algorithm 4) can be modified as well.

5.5 Interpretation of the outputs in Non-Metric PLS Regression

In NM-PLSR, the weight of a quantified variable in the construction of the first component can be expressed also as a function of the raw non-metric variables. However, though this feature enriches their informative power, it can leads to misleading interpretation, as some of their properties depend on scaling level of the analysis.

Algorithm 10 NM-PLSR algorithm

Input: $\boldsymbol{X}^*, \boldsymbol{Y}^*$
Output: $\boldsymbol{W}, \boldsymbol{C}, \boldsymbol{T}, \boldsymbol{U}, \widehat{\boldsymbol{X}}, \widehat{\boldsymbol{Y}}$

Step 1.0: Initialize $\boldsymbol{u}_1$
Step 1.1:
repeat
Step 1.1.1: $\widehat{\boldsymbol{x}}_p = \mathcal{Q}(\boldsymbol{u}_1, \boldsymbol{x}_p^*)$
Step 1.1.2: $\widehat{\boldsymbol{X}} = [\widehat{\boldsymbol{x}}_1 \dots \widehat{\boldsymbol{x}}_P]$
Step 1.1.3: $\boldsymbol{w}_1 = \widehat{\boldsymbol{X}}^{'}\boldsymbol{u}_1 / \|\widehat{\boldsymbol{X}}^{'}\boldsymbol{u}_1\|$
Step 1.1.4: $\boldsymbol{t}_1 = \widehat{\boldsymbol{X}}\boldsymbol{w}_1 / (\boldsymbol{w'}_1\boldsymbol{w}_1)$
Step 1.1.5: $\widehat{\boldsymbol{y}}_r = \mathcal{Q}(\boldsymbol{t}_1, \boldsymbol{y}_r^*)$
Step 1.1.6: $\widehat{\boldsymbol{Y}} = [\widehat{\boldsymbol{y}}_1 \dots \widehat{\boldsymbol{y}}_P]$
Step 1.1.7: $\boldsymbol{c}_1 = \widehat{\boldsymbol{Y}}^{'}\boldsymbol{t}_1 / (\boldsymbol{t}_1^{'}\boldsymbol{t}_1)$
Step 1.1.8: $\boldsymbol{u}_1 = \widehat{\boldsymbol{Y}}\boldsymbol{c}_1 / (\boldsymbol{c}_1^{'}\boldsymbol{c}_1)$
until convergence of $\boldsymbol{w}_1$
Step 1.2: $\boldsymbol{p}_1 = \widehat{\boldsymbol{X}}\boldsymbol{t}_1 / (\boldsymbol{t}_1^{'}\boldsymbol{t}_1)$
Step 1.3: $\boldsymbol{E}_1 = \widehat{\boldsymbol{X}} - \boldsymbol{t}_1\boldsymbol{p}_1^{'}$
Step 1.4: $\boldsymbol{F}_1 = \widehat{\boldsymbol{Y}} - \boldsymbol{t}_1\boldsymbol{c}_1^{'}$
for all $h = 2, \dots, H$ **do**
Step 2.0: Initialize $\boldsymbol{u}_h$
Step 2.1:
repeat
Step 2.1.1: $\boldsymbol{w}_h = \boldsymbol{E}_{h-1}^{'}\boldsymbol{u}_h / \|\boldsymbol{E}_{h-1}^{'}\boldsymbol{u}_h\|$
Step 2.1.2: $\boldsymbol{t}_h = \boldsymbol{E}_{h-1}\boldsymbol{w}_h / (\boldsymbol{w}_h^{'}\boldsymbol{w}_h)$
Step 2.1.3: $\boldsymbol{c}_h = \boldsymbol{F}_{h-1}^{'}\boldsymbol{t}_h / (\boldsymbol{t}_h^{'}\boldsymbol{t}_h)$
Step 2.1.4: $\boldsymbol{u}_h = \boldsymbol{F}_{h-1}\boldsymbol{c}_h / (\boldsymbol{c}_h^{'}\boldsymbol{c}_h)$
until convergence of $\boldsymbol{w}_h$
Step 2.2: $\boldsymbol{p}_h = \boldsymbol{E}_{h-1}^{'}\boldsymbol{t}_h / (\boldsymbol{t}_h^{'}\boldsymbol{t}_h)$
Step 2.3: $\boldsymbol{E}_h = \boldsymbol{E}_{h-1} - \boldsymbol{t}_h\boldsymbol{p}_h^{'}$
Step 2.4: $\boldsymbol{F}_h = \boldsymbol{F}_{h-1} - \boldsymbol{t}_h\boldsymbol{c}_h^{'}$
end for

5.5.1 The weights

The generic weight w_{p1} in PLS-R equals the OLS regression coefficient $b^{\text{OLS}}_{(\boldsymbol{u}_1|\boldsymbol{x}_p)}$ of $\boldsymbol{u}_1$ on $\boldsymbol{x}_p$. In NM-PLSR, since the quantified variables are standardized by construction, the weights of the first PLS component can be interpreted in term of correlations with the component: the higher the correlation is, the higher the weight is In particular, for each p, w_{p1} equals the correlation between $\widehat{\boldsymbol{x}}_p$ and $\boldsymbol{u}_1$ for less than the constant $\sqrt{\text{var}(\boldsymbol{u}_1)}/\|\boldsymbol{w}_1\|$. Analogously proceeding for the $\widehat{\boldsymbol{Y}}$-weights leads us to the equations

$$c_{1r} \propto \text{cor}(\widehat{\boldsymbol{y}}_r, \boldsymbol{t}_1) \text{ and } w_{1r} \propto \text{cor}(\widehat{\boldsymbol{x}}_p, \boldsymbol{u}_1), \tag{5.17}$$

where symbol $\propto$ indicates a proportionality factor constant for each r and for each p.

The quantifications listed in the previous section assure a nice interpretation of weights even in term of statistical relation between the component and the original variable analyzed at a non-metric scaling level (for metric scaling level analyses, of course, the problem of the interpretation does not exists).

When a predictor $\boldsymbol{x}^*_p$ is quantified at nominal scaling level, the following relation holds

$$\text{cor}(\widehat{\boldsymbol{x}}_p, \boldsymbol{u}_1) = \eta_{(\boldsymbol{u}_1|\boldsymbol{x}^*_p)}, \tag{5.18}$$

where η is the Pearson's correlation ratio, *i.e.* the part of variability of $\boldsymbol{u}_1$ explicated by the categories of $\boldsymbol{x}^*_p$.

At the same way, if a response $\boldsymbol{x}_p^*$ is quantified at nominal scaling level,

$$\text{cor}(\widehat{\boldsymbol{y}}_r, \boldsymbol{t}_1) = \eta_{(\boldsymbol{t}_1|\boldsymbol{y}_r^*)}. \tag{5.19}$$

It is necessary to pay attention in interpreting these weights, because they are always not negative, since $0 \leq \eta \leq 1$. The weight referred to a variable analyzed at a nominal scaling level can be interpreted in terms of intensity, but not in terms of sign. After all, it is conceptually wrong expecting a sign in the relation between a numerical and a nominal variable, since a nominal variable neither increases, or decreases. To be clear, a sentence like "When income increases, the nationality increase too" does not make sense.

If a variable is quantified at an ordinal level, the sign of corresponding weight can be interpreted. In fact, the following equations hold

$$\text{cor}(\widehat{\boldsymbol{y}}_r, \boldsymbol{t}_1) = \begin{cases} \sqrt{1 - STRESS^2_{(\boldsymbol{t}_1, \boldsymbol{y}_r^*)}} & \text{if cor}(\widehat{\boldsymbol{y}}_r, \boldsymbol{t}_1) \geq 0 \\ -\sqrt{1 - STRESS^2_{(\boldsymbol{t}_1, \boldsymbol{y}_r^*)}} & \text{if cor}(\widehat{\boldsymbol{y}}_r, \boldsymbol{t}_1) < 0 \end{cases}$$

$$\text{cor}(\widehat{\boldsymbol{x}}_p, \boldsymbol{u}_1) = \begin{cases} \sqrt{1 - STRESS^2_{(\boldsymbol{u}_1, \boldsymbol{x}_p^*)}} & \text{if cor}(\widehat{\boldsymbol{x}}_p, \boldsymbol{u}_1) \geq 0 \\ -\sqrt{1 - STRESS^2_{(\boldsymbol{u}_1, \boldsymbol{x}_p^*)}} & \text{if cor}(\widehat{\boldsymbol{x}}_p, \boldsymbol{u}_1) < 0 \end{cases}$$

From these relations it descends that if the weight of a scale variable is positive there is a direct relation between the raw variable and the first component. In the opposite case, it means that this relation is inverse. Moreover, intensity of the weight of a quantified variable measures the

strength of the statistical relations between the component with both the quantified and the raw variable: the first in terms of correlation, the second in terms of approaching to perfect monotonicity, intended as in Kruskal's secondary approach [Kruskal 1964*a*].

5.5.2 The regression coefficients

The one-component PLS regression coefficient $b^{\text{PLS(1)}}_{\widehat{\boldsymbol{y}}_r|\widehat{\boldsymbol{x}}_p}$ of $\widehat{\boldsymbol{x}}_p$ on $\widehat{\boldsymbol{y}}_r$ equals the product of the weights c_{1r} and w_{1p}. Hence, the sign of the weight affect the sign of the regression coefficient in the following way:

- If both $\widehat{\boldsymbol{x}}_p$ and $\widehat{\boldsymbol{y}}_r$ are analyzed at a nominal scaling level, the regression coefficient equals

$$b^{\text{PLS(1)}}_{\widehat{\boldsymbol{y}}_r|\widehat{\boldsymbol{x}}_p} = \eta_{(\boldsymbol{u}_1|\boldsymbol{x}^*_p)} \times \eta_{(\boldsymbol{t}_1|\boldsymbol{y}^*_r)}$$

  and it is always positive.

- If one of the two variable is analyzed at a nominal and the other at ordinal scaling level,

$$b^{\text{PLS(1)}}_{\widehat{\boldsymbol{y}}_r|\widehat{\boldsymbol{x}}_p} = \begin{cases} \eta_{(\boldsymbol{u}_1|\boldsymbol{x}^*_p)} \times \pm\sqrt{1 - STRESS^2_{(\boldsymbol{t}_1,\boldsymbol{y}^*_r)}} \\ \pm\sqrt{1 - STRESS^2_{(\boldsymbol{u}_1,\boldsymbol{x}^*_p)}} \times \eta_{(\boldsymbol{t}_1|\boldsymbol{y}^*_r)} \end{cases}$$

  Hence, it takes the sign of the weight of the variable analyzed at an ordinal scaling level.

- If the two variables are analyzed at an ordinal scaling level, the sign of the regression coefficient is expressed by the product of the signs of the weights c_{1r} and w_{1p}, like in standard PLS-R

$$b^{\mathrm{PLS}(1)}_{\widehat{\boldsymbol{y}}_r|\widehat{\boldsymbol{x}}_p} = \pm\sqrt{1 - STRESS^2_{(\boldsymbol{u}_1,\boldsymbol{x}^*_p)}} \times \pm\sqrt{1 - STRESS^2_{(\boldsymbol{t}_1,\boldsymbol{y}^*_r)}}$$

5.6 A Conjoint Analysis by means of Non-Metric PLS Regression

In this section NM-PLSR is applied to Conjoint Analysis. The aim of this application is to show how Non-Metric PLS Regression can improve data interpretation with respect to the dummy coding.

We use a very known data-set, that is the tea data-set used as example in Kuhfeld [1993] in order to show the procedure Transreg of SAS software. It is a classic example of Conjoint Analysis, where we want to study consumer (or judges) preferences with respect to a set of scenarios described by relevant attributes.

Tea data-set has been already analyzed in Tenenhaus [1998] using the PLS Regression. Tenenhaus handled the nominal attributes using the dummy coding: each nominal variable was replaced by the corresponding indicator matrix. We discussed the drawbacks of such a coding in section 3.2. First of all, dummy coding gives up the idea of the variable as a whole, while it considers all the categories as they were variables in themselves. Furthermore, the dummy coding increases the dimensionality of the data matrix. Finally, the weight of a

dummy variable representing a category mainly associated to central values of the corresponding component distribution is systematically underestimated.

Here we overcome these drawbacks by performing a Non-Metric PLS Regression.

In the following, first we briefly present the data (section 5.6.1). Then we perform three different PLR Regression analyses on these data (sections 5.6.2, 5.6.3 and 5.6.4). These analyses are performed by using an R code developed by the author (see Appendix). In the first analysis, we run a standard PLS Regression according to Tenenhaus [1998]. In particular, we consider judges' preference rankings as they were interval variables, and we replace each attribute with the corresponding indicator matrix. In the second analysis, we implement a NM-PLS Regression analyzing each attribute at a nominal scaling level and considering the responses as interval variables. In the third, we scale also the response variables, using an ordinal scaling level.

5.6.1 A data-set on tea tasting

Tea data-set (see table 5.1) consists of six judges who rank, according to their preferences, 18 teas differing in four attributes: temperature (three levels: "hot", "warm" and "iced"), strength ("light", "medium", "strong"), presence of lemon ("yes", "no") and of sugar ("no" sugar, "one" sugar cube, "two" sugar cubes).

Data are organized in the following way:

- the consumer rankings are the columns J1, J2, .., J6, of a re-

sponse matrix. These ranking are considered in inverse order: in this way, the better the preference is, the higher the value is. For example, the value assigned to the preferred scenario is "18", and the value assigned to the worst one is "1".

- the four attributes observed on the scenarios are the columns of a predictor matrix $\boldsymbol{X}$
- attributes and consumer prefefences are observed on eighteen scenarios (raws), chosen among the 54 possible scenarios so as to build an orthogonal design.

5.6.2 PLS Regression on dummy coded attributes

A standard PLS-R is performed according to Tenenhaus [1998]. In particular, we consider judges' rankings as they were interval variables, and we replace each attribute with the corresponding indicator matrix (see table 5.2). In order to distinguish this analysis from the next, we will name it "linear-dummy".

Using dummy coding heavily enlarges the number of predictors. In the case of the tea data-set the predictor set passes from four categorical variables (the attributes) to eleven binary variables (the levels). Afterwards, all the variables are centered and standardized to unitary variance in order to make the analysis comparable to the next ones.

According to Tenenhaus [1998], we perform a four-component PLS Regression. Since the number of predictors is large, the analysis has a very good explanatory power. In fact, the four components explain

	Judgements						Attributes			
Scenario	J1	J2	J3	J4	J5	J6	Temp	Sugar	Strength	Lemon
1	15	17	15	16	6	14	Hot	Zero	Strong	Yes
2	17	11	18	10	9	11	Hot	One	Medium	Yes
3	13	9	6	1	14	13	Hot	Two	Light	No
4	6	6	9	14	17	7	Warm	Zero	Medium	No
5	5	3	2	7	3	10	Warm	One	Light	Yes
6	4	1	7	4	11	3	Warm	Two	Strong	Yes
7	12	16	5	17	1	17	Iced	Zero	Light	Yes
8	8	13	14	12	16	2	Iced	One	Strong	No
9	9	8	13	6	7	12	Iced	Two	Medium	Yes
10	16	18	8	15	13	15	Hot	Zero	Light	No
11	18	12	17	9	12	5	Hot	One	Strong	Yes
12	14	7	16	2	10	6	Hot	Two	Medium	Yes
13	2	5	3	13	8	1	Warm	Zero	Strong	Yes
14	1	4	10	8	18	8	Warm	One	Medium	No
15	3	2	1	3	4	9	Warm	Two	Light	Yes
16	11	15	11	18	5	18	Iced	Zero	Medium	Yes
17	10	14	4	11	2	16	Iced	One	Light	Yes
18	7	10	12	5	15	4	Iced	Two	Strong	No

Table 5.1: Tea data-set

	Judgements						Attributes										
Sc.	J1	J2	J3	J4	J5	J6	H	W	I	Z	O	T	S	M	L	Y	N
1	15	17	15	16	6	14	1	0	0	1	0	0	1	0	0	1	0
2	17	11	18	10	9	11	1	0	0	0	1	0	0	1	0	1	0
3	13	9	6	1	14	13	1	0	0	0	0	1	0	0	1	0	1
4	6	6	9	14	17	7	0	1	0	1	0	0	0	1	0	0	1
5	5	3	2	7	3	10	0	1	0	0	1	0	0	0	1	1	0
6	4	1	7	4	11	3	0	1	0	0	0	1	1	0	0	1	0
7	12	16	5	17	1	17	0	0	1	1	0	0	0	0	1	1	0
8	8	13	14	12	16	2	0	0	1	0	1	0	1	0	0	0	1
9	9	8	13	6	7	12	0	0	1	0	0	1	0	1	0	1	0
10	16	18	8	15	13	15	1	0	0	1	0	0	0	0	1	0	1
11	18	12	17	9	12	5	1	0	0	0	1	0	1	0	0	1	0
12	14	7	16	2	10	6	1	0	0	0	0	1	0	1	0	1	0
13	2	5	3	13	8	1	0	1	0	1	0	0	1	0	0	1	0
14	1	4	10	8	18	8	0	1	0	0	1	0	0	1	0	0	1
15	3	2	1	3	4	9	0	1	0	0	0	1	0	0	1	1	0
16	11	15	11	18	5	18	0	0	1	1	0	0	0	1	0	1	0
17	10	14	4	11	2	16	0	0	1	0	1	0	0	0	1	1	0
18	7	10	12	5	15	4	0	0	1	0	0	1	1	0	0	0	1

Table 5.2: Tea data-set: the dummy coding

the 89.8% of the variability of $\boldsymbol{Y}$. However, as table 5.1 shows, J1 and J2 preferences are well represented on the first component, while the second component mainly explains J3 and J5 preferences. The third component is important in the description of J4 preferences that are sufficiently represented also on the first component. J6 is the worst modeled judge and his better representation is on the first axis. Therefore, observing the loading plot of first two components, which explain the 70.3% of $\boldsymbol{Y}$ variability, most of the relations between variables can be read (see figure 5.2). The first dimension represents judges J1, J2,

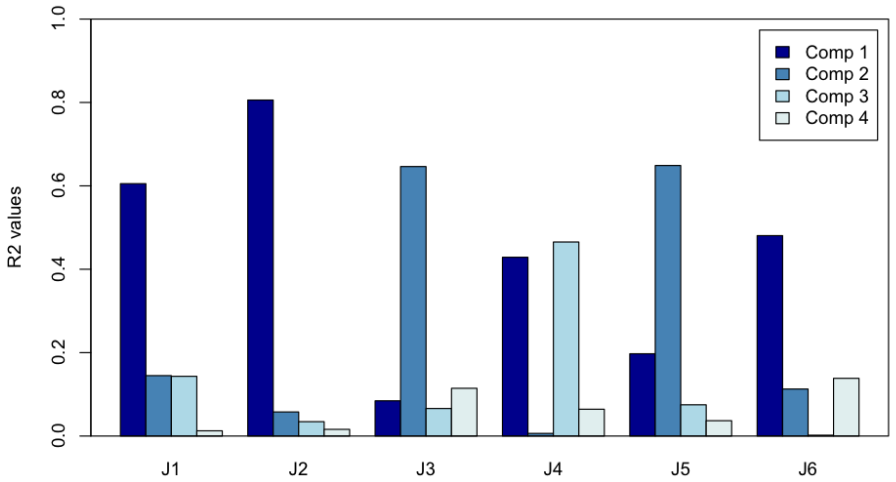

Figure 5.1: PLS Regression on dummy predictors: Response variability explained by the first four components

J4 and J6. On aggregate, they don't like warm tea, while they like tea without sugar. Other attributes have a secondary importance in their preferences. Judges J3 and J5 are well represented on the second dimension. They prefer hot tea, while dislike light tea with lemon. It's noteworthy that all of these information can be read without excessive effort because the number of judges, attributes and their levels is quite low. In most complex cases the judges can be grouped in classes in order to simplify the interpretation of the analysis. However, it makes no sense do the same with levels of different attributes.

This type of analysis does not provide information about which characteristics are the most important drivers for judge preferences.

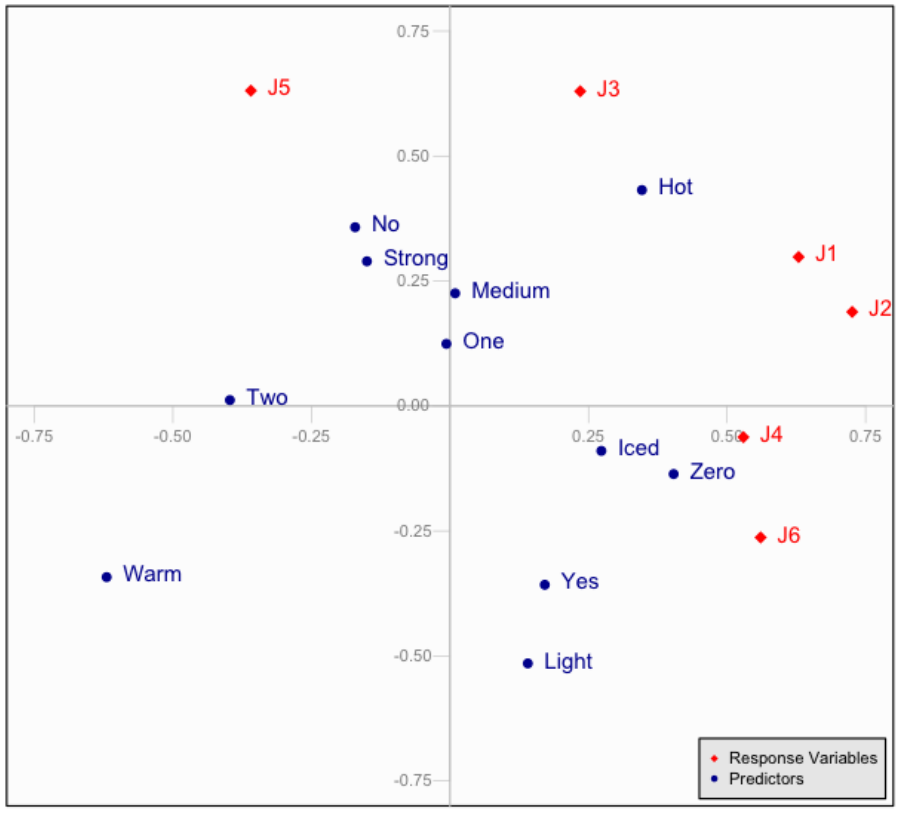

Figure 5.2: PLS Regression on dummy predictors: Loading plot of $(\boldsymbol{r}_1, \boldsymbol{c}_1)$ and $(\boldsymbol{r}_2, \boldsymbol{c}_2)$

In fact, the dummy approach does not consider each attribute as a whole, but considers separately each level of the attributes. In order to overcome these drawbacks, in the next a NM-PLSR is implemented, where each attribute is scaled by assigning a numerical value to each

of its levels.

5.6.3 Non-Metric PLS Regression on optimally scaled attributes

Here data are analyzed *via* NM-PLS Regression. We will call this analysis "linear-nominal", as each attribute is analyzed at a nominal scaling level and the responses are considered at a linear scaling level (*i.e.* they are simply standardized). As a result, we predict six responses from just four optimally scaled predictors, any predictor representing an attribute (temperature, presence of sugar, strength and presence of lemon). This analysis has a good explicative power ($R^2_Y = 0.81$ in the four-component model), but lower than the "linear-dummy" one. However, preference variables are well represented on the first loading plot, of $(\boldsymbol{r}_1, \boldsymbol{c}_1)$ and $(\boldsymbol{r}_2, \boldsymbol{c}_2)$, since each of them is well explained by one of the first two components (see figure 5.3).

The loading plot highlights all the information useful for investigating judges preferences with respect to each attribute (see figure 5.4). In interpreting the loading plot, however, we have to pay attention, as the impact of each attribute on the preferences of each judges can be read only in terms of intensity and not in terms quality. For example, the fact that variables "temperature" and "J2" lie on the same quadrant and they are very close in terms of angle does not mean that judge J2 prefers very hot tea. The right interpretation is that attribute temperature is the most important for J2 preferences; similarly, for J3 and J5, strength and force are the most important attributes; for J4

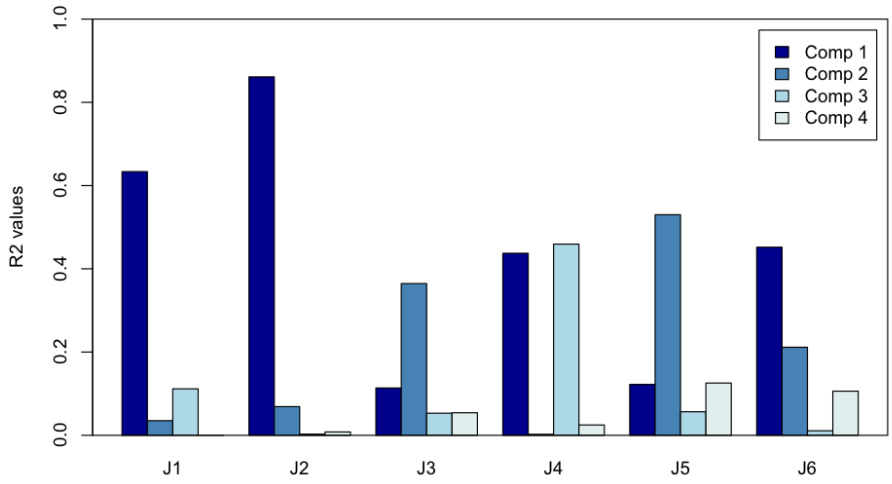

Figure 5.3: Non-Metric PLS Regression on predictors scaled at a nominal scaling level: Response variability explained by the four components

the presence of lemon in the tea is a very discriminant factor for its preference; finally, preferences of J6 depend more or less on all of the attributes but temperature.

The previous loading plot does not describe the way an attribute affects the preferences, that is which level of the attribute is appreciated or disliked by judges. However, this information can be easily recovered if we remember that the scaling value for each level equals the average of the $\boldsymbol{Y}$-scores of observations sharing that level. This implies that levels associated to positive scaling values are globally preferred by the judges and *vice versa*. Hence, we can infer which

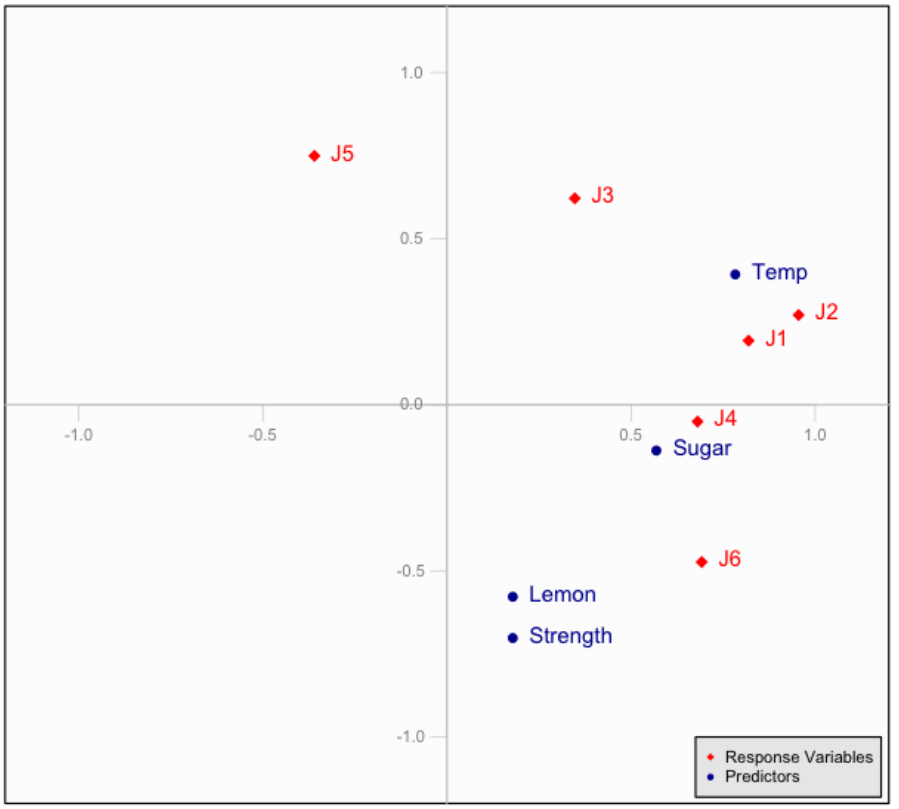

Figure 5.4: Non-Metric PLS Regression on predictors scaled at a nominal scaling level: Loading plot of vectors $(\boldsymbol{r}_1, \boldsymbol{c}_1)$ and $(\boldsymbol{r}_2, \boldsymbol{c}_2)$

levels are preferred by judges from their scaling values. In the case of judges J1 and J2, for example, "iced" and "hot" levels are preferred, as their scaling values are positive (respectively 0.58 and 0.79); on the contrary, they don't like at all warm tea, as scaling value of level

"warm" is -1.37. This information can be effectively represented by plotting these values on the loading plot.

In particular, we propose to represent each level as a point lying on the direction spanned by the point-vector of the corresponding attribute having a distance from the origin equal to own scaling value. In alternative, for each attribute, the averages of the predicted scaling values referred to observations sharing the same level of the attribute can be plotted. Using the predicted values allows us to interpret the variability of the point-levels referring to each attribute in terms of variability of the attribute explained by the model. In the loading plot in figure 5.5 this type of representation is used. As one can clearly deduce by this plot, temperature is important for J1 and J2 in the sense that they strongly dislike warm tea; J4 prefers hot or iced tea too, but the most important for him is that there is no sugar inside; J6 likes light tea with lemon but without sugar; on the contrary, J3 and J5 prefer strong tea without lemon.

5.6.4 Non-Metric PLS Regression of optimally scaled preferences on optimally scaled attributes

In standard PLS Regression model a linear relation between variables and components is supposed. The analysis of variables at an ordinal scaling level allows us discarding this hypothesis and replacing it by a milder hypothesis of monotonicity.

In this last analysis, response variables (*i.e.* the judges' preferences expressed in terms of ranking) are analyzed at an ordinal scaling level

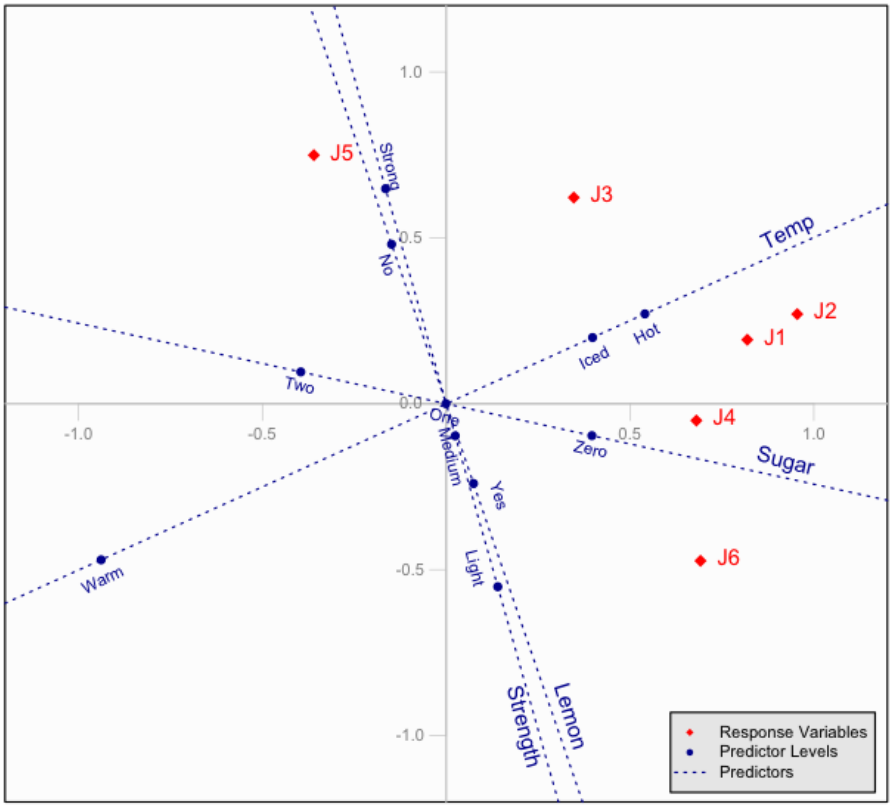

Figure 5.5: Non-Metric PLS Regression of responses scaled at a ordinal level on predictor scaled at a nominal level: Mapping of the levels in the loading plot

in order to detect their non-linear (but monotonic) relations with the first PLS component. Predictor variables (*ie.* the attributes), instead, are analyzed at a nominal scaling level, as in "linear-nominal" analysis. Therefore, we will call this analysis " ordinal-nominal".

Since this analysis discards linear constraint in relation between the responses and the first component, we obtain a very predictive first order model, which explains the 59% of response variability. On aggregate, the first two components explain very well responses J1 ($R^2 = 0.84$), J2 ($R^2 = 0.98$), J3 ($R^2 = 0.86$) and J4 ($R^2 = 0.78$), and adequately responses J5 ($R^2 = 0.58$) and J6 ($R^2 = 0.62$). Higher order components does not seem to add useful information (see figure 5.6)

In figure 5.7 the scaling values of response variables are plotted versus their original rankings. This plot can suggest functional transformations to apply to original variables in order to capture their non-linear relation with the first component. For example, relation between the original ranking variable J3 and the component could be well approximated by an exponential function.

Looking at the correlation between the scaled variables and the first component it is possible to measure the degree of monotonicity of the relation between the component and the original ranking (see section 5.5.1). The same information can be visualized looking at the x-axis values of the response variables in the loading plot of $(\boldsymbol{r}_1, \boldsymbol{c}_1)$ and $(\boldsymbol{r}_2, \boldsymbol{c}_2)$ (see figure 5.8).

The loading plot in figure 5.8 shows the relations already observed

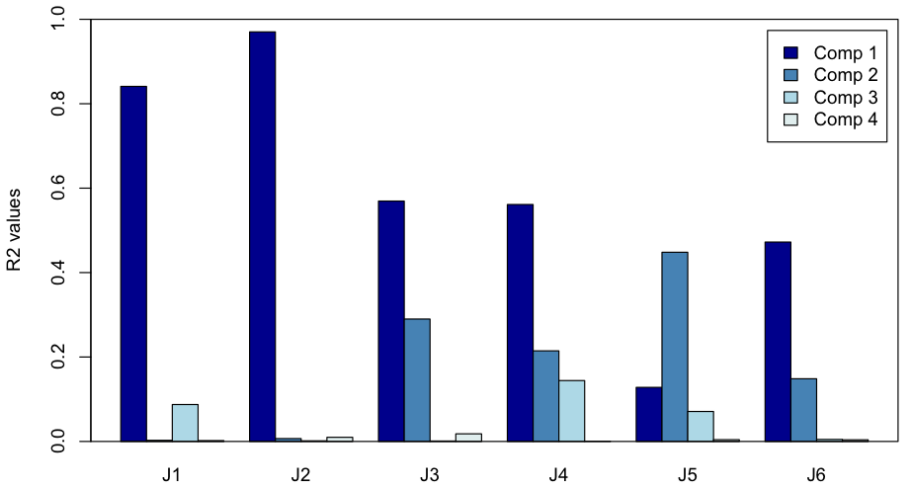

Figure 5.6: Non-Metric PLS Regression of responses scaled at a ordinal level on predictor scaled at a nominal level: Response variability explained by the four components

in previous analyses. J1's preferences are very influenced by attributes presence of lemon and strength; strength, as well as temperature important also for J3. Preferences of J1, J2, J4 and J6 depend on attributes temperature and presence of sugar. Temperature is of main importance for J1 and J2, while sugar is of primary importance for J4 and J6.

In order to investigate the way each attribute influences judge's preferences, levels can be mapped in the same plot, as previously explained. The resulting plot (figure 5.9) is even clearer than the one in "linear-nominal" analysis, as all of the responses lie in the right part

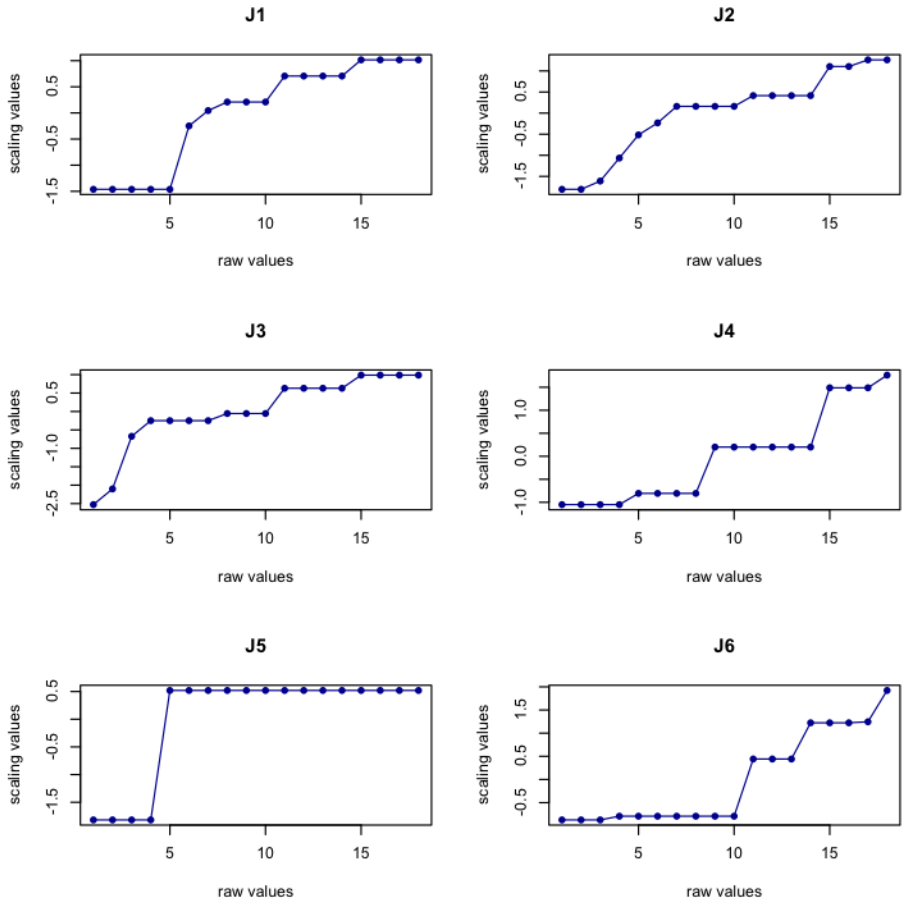

Figure 5.7: Response variables: scaling values vs ranking

of the plot. Hence, in order to satisfy the most of the judges' tastes the tea should be not warm, with neither lemon nor sugar and not too light.

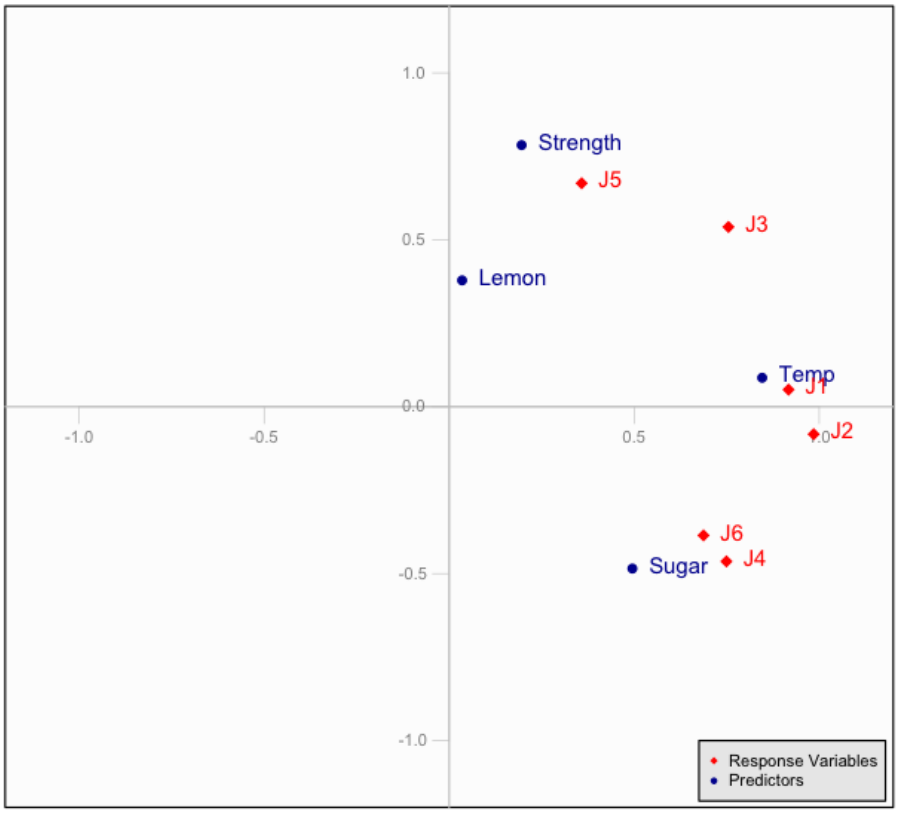

Figure 5.8: Non-Metric PLS Regression of responses scaled at a ordinal level on predictor scaled at a nominal level: Loading plot of $(\boldsymbol{r}_1, \boldsymbol{c}_1)$ and $(\boldsymbol{r}_2, \boldsymbol{c}_2)$

5.6.5 Conclusion

From the comparison of the explicative power of the three analyses, it descends that dummy approach provides better results when we

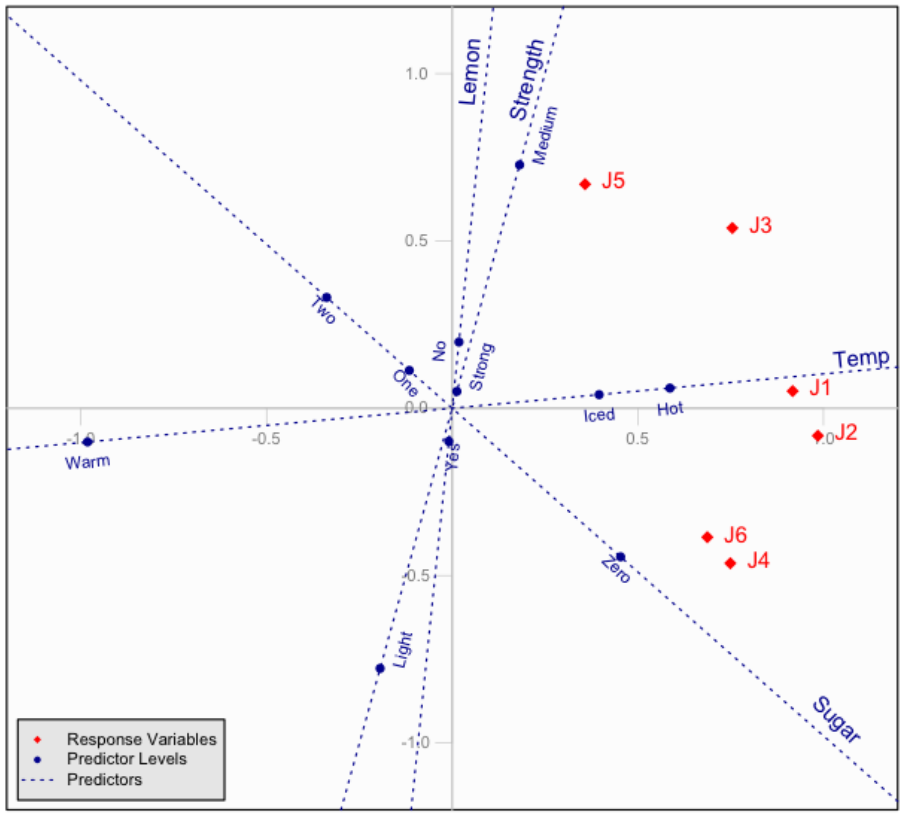

Figure 5.9: Non-Metric PLS Regression of responses scaled at a ordinal level on predictor scaled at a nominal level: Mapping of the levels on the loading plot

build models with several components (see figure 5.10). This can be explained by the fact that "linear-dummy" analysis works with eleven predictors, while the non-metric approaches can exploit the explicative

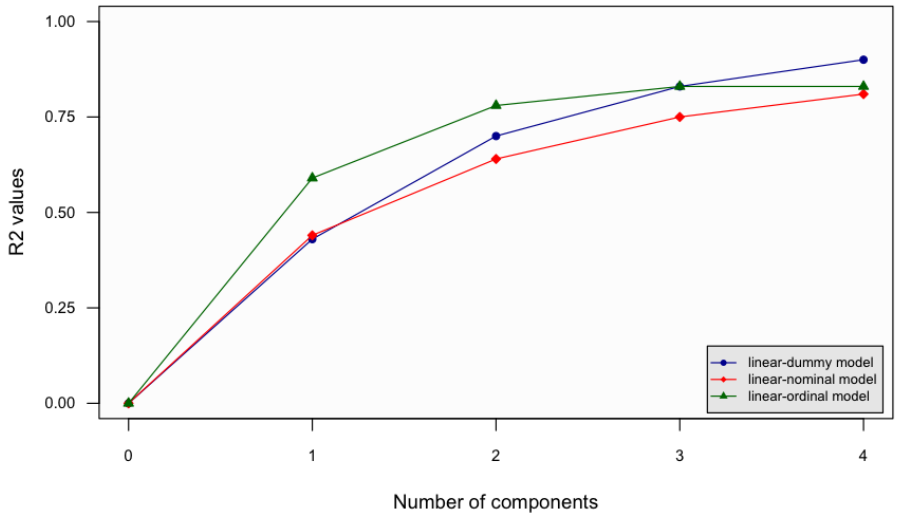

Figure 5.10: Response variability explained by model from one to four component in the three analyses

capability of just four scaled predictors. However, this difference is mainly quantitative, since we can extract the same information in all of the three analyses.

Non-Metric PLS Regression resumes most of the information in fewer components, and allows for a two level analysis. This synthesis ability of NM-PLSR can be greatly useful when we have a large number of attributes with a large number of levels.

In a first level analysis, we observe the impact of each attribute on the judges' preferences. This leads to cleaner and more easily interpretable factorial representations, as we observe relations between the original variables, and not between their levels. Moreover, since in NM-PLSR each attribute is handled as a whole, it is possible to assess which attributes are the most important in judges' preferences by means of regression coefficients and VIP indexes.

VIP index (see section 1.3.8) measures the importance of each predictor in the prediction of the whole response set. In table 5.3, the predictors sorted by VIP of four-component models are shown. Whereas in non-metric analysis is straightforward to observe that the variable "temperature" is the most important in the prediction, in standard PLS-R analysis a precise ranking is not possible because the levels of this attribute take the first, the second but also the ninth place in the VIP ranking. At the same way, levels of variable "sugar" take very different places in the ranking. This variability is due to the fact that importance in the prediction of levels associated to central values of the component distribution is underestimated (see section

(a) "linear-dummy" analysis

Variable	VIP
Warm	1.57
Hot	1.21
Zero	1.17
Two	1.12
Light	1.12
Yes	0.98
No	0.98
Strong	0.76
Iced	0.68
Medium	0.43
One	0.26

(b) "linear-nominal" analysis

Variable	VIP
Temp	1.27
Sugar	1.04
Strength	0.83
Lemon	0.78

(c) "ordinal-nominal" analysis

Variable	VIP
Temp	1.45
Sugar	1.02
Strength	0.84
Lemon	0.41

Table 5.3: Predictor variables ranked by VIP index (four-component models)

3.2).

Through a second level analysis, in MN-PLSR it is possible to investigate also the way an attribute affects the preferences exploiting the *a priori* information we have about the membership of observations to groups defined by the levels of the attributes.

Chapter 6

Non-Metric PLS Path Modeling

6.1 Motivation

PLS-PM is very used in social science for its ability in handling latent concepts like satisfaction, performance, wellness or intelligence, which are not directly observable. All of these variable can not be directly measured, but can be synthesized by a suite of indicators. Due to the need to formalize models relating latent concepts, since 1982 PLS-PM is more an more used in marketing research for the quantitative analysis of consumer satisfaction [Fornell & Bookstein 1982].
In marketing applications, however, latent concepts are expressed as a synthesis of variables which in their turn can not be measured *strictu sensu*. Typically, in fact, it is asked to the consumer to ex-

press the level of agreement to a statement, or a judgement about particular characteristics of the offered product or service. Sometimes interviewee are asked to associate their opinion choosing one between a set of ordered response levels (the so-called Likert item [Likert 1932]). For example, a typic Likert item is : Strongly disagree, Disagree, Neither agree nor disagree, Agree, Strongly agree. Afterwards, responses are replaced by more or less arbitrary values (*e.g.* Strongly disagree=1, Disagree=2, Neither agree nor disagree=3, Agree=4, Strongly agree=5) which are analyzed with standard quantitative methods. Most of times, in order to directly obtain (fake) quantitative values, the interviewer asks to associate the agreement level to one of the values of a certain scale (*e.g.* 1-5, 1-10 or 1-100). Of course, this does not change the actual nature of the variable to be analyzed: data collected in such a way are not based on a metric, and so they could not be handled as they were numeric. Notwithstanding, PLS Path Modeling is often implemented on this kind of data. In other words, categories of non-metric variables are usually arbitrarily quantified and then used as numerical indicators in a PLS Path Model. Generally these quantifications suppose that relative differences between subsequent categories are equals. This approach does not take in account that Likert scales may be subject to distortion from several causes. Namely, respondents may avoid using extreme response categories (central tendency bias) [Couch & Keniston 1960]; agree with statements as presented (acquiescence bias) [Knowles & Nathan 1997]; try to portray themselves in a more favorable light (social desirability

bias) [Ferrando 2008]; or tend to endorse the most extreme response categories regardless of content (extremity bias) [Greenleaf 1992].

A less arbitrary approach is provided by the Item Response Theory, in which data coming from responses to questionnaires are supposed to follow mathematical models defined *ex ante* [Andrich 1978]. However, these model seems do not fit with the soft modeling spirit of PLS-PM, due their strong distributional assumptions.

Here, we propose a new approach to PLS-PM, which provides at the same time specific PLS-PM parameters as well as scaling values for variables to be scaled. Non-Metric PLS Path Modeling (NM-PLSPM) is a data driven approach, that allows for non-metric and non-linear analysis of variables measured at any scale level in reflective Path Models. As in the other NM-PLS methods, in NM-PLSPM quantifications depends on the scale level chosen for each variable and are coherent with the model, as they optimize the same criterion with which model parameters are estimated.

6.2 State of the art

Two proposals have been recently presented in order to handle nominal variable in PLS-PM framework.

Betzin & Henseler [2005] proposed an ex ante transformation of each block of nominal manifest variables (MVs). Starting from the idea that PLS-PM can be interpreted as a multiple eigenvalue problem, Betzin and Hanseler propose to transform any block composed of non-

metric indicators by the transformation 4.11. This approach, however, has some drawbacks:

- it can be used only if a block is composed of all nominal variables.
- it yields outer weights for each category
- it provides *a priori* quantifications, which do not depend on the model

Jakobowicz & Darquenne [2007] proposed a modified PLS algorithm, called Partial Maximum Likelihood (PML), that can be applied also in the case where a block is composed of both nominal and numerical MVs.

PML algorithm works in three steps. In the first steps, an initial inner estimation is computed for each MV. For each block $\boldsymbol{X}_q$ a so-called reference variable belonging to a connected block $\boldsymbol{X}_{q'}$ is considered as a first outer estimation of $\boldsymbol{\xi}_q$. Then, initial outer weights for each numerical MV and for each category of a non-metric MV are calculated following the measurement scale of both the MV and the reference variable. In particular, Least Squares Regression model, ANOVA model, Logistic Regression model, Polytomic Logit model or Generalized Logit model coefficients are calculated considering the MV as explanatory and the reference variable as response. At the end of this step, initial outer estimations for each LV are calculated.

This procedure aims to obtain numerical initial latent variable (LV) estimates, keeping at the same time the original scales of the MVs. It is not clear the reason of all the computational burden in the first step of

PLS aimed to obtain initial outer weights (and consequently initial LV outer estimations) which are coherent with the measurement scales of the MVs. In fact, quoting the authors themselves, " Lohmöller [1989] *has showed that choosing different initial weights does not affects the final estimation of the model"*.

In the second step, a PLS loop is implemented in order to obtain the parameters of the measurement model. Inner and outer estimations of the LVs are alternated as in the standard PLS-PM algorithm. Outer weights, however, are updated following the measurement scale of the MVs. An outer weight for each modality of a non-metric MV is calculated as conditioned mean of the outer estimation of LV related to the reference variable, and it is successively centered and normalized. Weights of numerical MVs, instead, are obtained in the standard way. Hence, in PML a final outer weight for each category is obtained. The authors propose a formula to calculate *ex post* a global outer weight for each non-metric MV.

In the third step, inner model structural relations are estimated by OLS regressions between each response LV and its explanatory LVs.

The authors advice using PML for nominal or binary MVs, as in this cases "*it is not possible to suppose there is any underlying continuous distribution*".

6.3 From Non-Metric PLS Regression to Non-Metric PLS Path Modeling

Since *Mode A* PLS-PM algorithm is a straightforward extension of PLS-R algorithm, basic algorithmic principles of NM-PLSR algorithm can be easily extended to PLS approach to SEM. PLS-R algorithm in fact, for less that normalization constraints, is the same than two blocks PLS-PM algorithm.

Following PLS-PM notations, $\boldsymbol{u}_1$ is outer estimate of the LV associated to the block $\boldsymbol{Y}$ ($\boldsymbol{u}_1 \propto \boldsymbol{Y}\boldsymbol{c}_1$), as well as the inner estimate of the LV of the block $\boldsymbol{X}$ by means of which we calculate the outer weights ($\boldsymbol{w}_1 \propto \boldsymbol{X}'\boldsymbol{u}_1$). Simmetrically, $\boldsymbol{t}_1$ can considered as the outer estimation of the block $\boldsymbol{X}$ ($\boldsymbol{t}_1 \propto \boldsymbol{X}\boldsymbol{w}_1$) as well as the inner estimation of the LV of the block $\boldsymbol{Y}$. These double functions are justified by the inner relation $\boldsymbol{t}_1 \propto \boldsymbol{u}_1$ This is a hidden step in PLS-R algorithm, which, in a two-block PLS-PM context, can be interpreted as: the outer estimate of the LV in a block is the inner estimate of the LV in the other block.

Hence, keeping on using PLS-PM notations, in NM-PLSR we obtain quantified variables maximizing, under suitable constraints, their correlation with the inner estimate of corresponding LV. So, from the algorithmic point of view, the non-metric extension of PLS-R can be easily applied also to PLS-PM, by adding to the PLS-PM loop a quantification step in which any MV is quantified as a function of the inner estimate of the corresponding LV.

6.4 The Non-Metric PLS Path Modeling algorithm

Non-Metric PLS-PM loop differs from the standard PLS-PM loop in the fact that it starts by initializing the inner estimate of each LV, used to obtain a first scaling of the MVs. In fact, each raw MV $\boldsymbol{x}_{pq}^{*}$ is quantified so as to be maximally correlated to the corresponding LV inner estimate $\boldsymbol{v}_q$.

Quantified MVs $\widehat{\boldsymbol{x}}_{pq}$ maximizing $\text{cor}(\boldsymbol{v}_q, \widehat{\boldsymbol{x}}_{pq})$ are obtained by means of the quantification functions $\widetilde{\mathcal{Q}}(\boldsymbol{x}_{pq}^{*}, \boldsymbol{v}_q)$, $\widetilde{\widetilde{\mathcal{Q}}}(\boldsymbol{x}_{pq}^{*}, \boldsymbol{v}_q)$ and $\dot{\mathcal{Q}}(\boldsymbol{x}_{pq}^{*}, \boldsymbol{v}_q)$: they are the normalized orthogonal projections of the inner estimation of the corresponding LV on a suitable space defined by the scaling level at which each raw variable is analyzed (see section 3.4.2 for demonstration).

NM-PLSPM algorithm supports three levels of scaling analysis. Variables quantified at a nominal level preserve grouping property. Variables quantified at a ordinal level follow the secondary Kruskal's monotonic quantification. Variables transformed at a functional level are related to the corresponding LV inner estimate by polynomial relation (for further details, see section 2.4).

If a raw MV is analyzed at a nominal scale level, the corresponding scaling is

$$\widetilde{\mathcal{Q}}(\boldsymbol{x}_{pq}^{*}, \boldsymbol{z}_q) : \widehat{\boldsymbol{x}}_{pq} \propto \widetilde{\boldsymbol{X}}_{pq}(\widetilde{\boldsymbol{X}}'_{pq}\widetilde{\boldsymbol{X}}_{pq})^{-1}\widetilde{\boldsymbol{X}}'_{pq}\boldsymbol{z}_q \tag{6.1}$$

If a raw MV is analzed at a ordinal scale level, is scaled as

$$\widetilde{\widetilde{\mathcal{Q}}}(\boldsymbol{x}^*_{pq}, \boldsymbol{z}_q) : \widehat{\boldsymbol{x}}_{pq} \propto \widetilde{\widetilde{\boldsymbol{X}}}_{pq}(\widetilde{\widetilde{\boldsymbol{X}}}'_{pq}\widetilde{\widetilde{\boldsymbol{X}}}_{pq})^{-1}\widetilde{\widetilde{\boldsymbol{X}}}'_{pq}\boldsymbol{z}_q \tag{6.2}$$

Finally, a metric MV can be analyzed at a functional level as

$$\dot{\mathcal{Q}}(\boldsymbol{x}^*_{pq}, \boldsymbol{z}_q) : \widehat{\boldsymbol{x}}_{pq} \propto \dot{\boldsymbol{X}}_{pq}(\dot{\boldsymbol{X}}'_{pq}\dot{\boldsymbol{X}}_{pq})^{-1}\dot{\boldsymbol{X}}'_{pq}\boldsymbol{z}_q \tag{6.3}$$

Matrices $\widetilde{\boldsymbol{X}}_{pq}$, $\widetilde{\widetilde{\boldsymbol{X}}}_{pq}$ and $\dot{\boldsymbol{X}}_{pq}$ are built as explained in section 3.4.2.

Once we get the quantified variables, the standard PLS-PM loop starts: LVs are first estimated by *Mode A* in the outer estimation process, and successively re-estimated in the inner estimation process.

After obtaining new inner estimates of the LVs, another iteration starts with a new quantification of the MVs, and the algorithm goes on until convergence.

A pseudo-code on NM-PLSPM algorithm is provided in algorithm 11.

6.5 The optimizing criterion of Non-Metric PLS Path Modeling

Unfortunately, NM-PLSPM algorithm suffers of the same drawbacks of the *Mode A* PLS-PM algorithm. That is, since the criterion to which it converges in unknown, we can not state that scalings provided by NM-PLSPM algorithm are mathematically optimal with respect to

Algorithm 11 Non-Metric PLS Path Modeling algorithm

Input: $\boldsymbol{X} = [\boldsymbol{X}_1, \ldots, \boldsymbol{X}_q \ldots, \boldsymbol{X}_Q], \boldsymbol{L}$

Output: $\boldsymbol{\beta}_j$, $\boldsymbol{w}_q$, $\widehat{\boldsymbol{\xi}}_q$, $\widehat{\boldsymbol{X}} = [\widehat{\boldsymbol{X}}_1, \ldots, \widehat{\boldsymbol{X}}_q \ldots, \widehat{\boldsymbol{X}}_Q]$

Step 0: Initialization

$\boldsymbol{z}_q = \boldsymbol{z}_q^{(0)}$

Step 1: Iteration

repeat

Step 1.1: Quantification step

$\widehat{\boldsymbol{x}}_p = \mathcal{Q}(\boldsymbol{z}_q^{(s)}, \boldsymbol{x}_{pq}^{*(s)})$

Step 1.2: Quantification step

$\widehat{\boldsymbol{X}}_q^{(s)} = [\widehat{\boldsymbol{x}}_{1q}^{(s)} \ldots \widehat{\boldsymbol{x}}_{Pq}^{(s)}]$

Step 1.3: Outer estimation of the LVs

$\boldsymbol{v}_q^{(s)} \propto \sum_{p=1}^{P_q} w_{pq}^{(s)} \boldsymbol{x}_{pq} = \boldsymbol{X}_q \boldsymbol{w}_q^{(s)}$

Step 1.4: Computation of the inner weights

$e_{qq'}^{(s)} = f\left(\boldsymbol{v}_q^{(s)}, \boldsymbol{v}_{q'}^{(s)}\right)$, according to the chosen scheme

Step 1.5: Inner estimationof the LVs

$\boldsymbol{z}_q^{(s)} \propto \sum_{q'=1}^{Q} c_{qq'} e_{qq'}^{(s)} \boldsymbol{v}_{q'}^{(s)}$

Step 1.6: Computation of the outer weights

$\boldsymbol{w}_q^{(s+1)} = (1/N) \boldsymbol{X}_q^{'} \boldsymbol{z}_q^{(s)}$ (Mode A) or

$\boldsymbol{w}_q^{(s+1)} = (\boldsymbol{X}_q^{'} \boldsymbol{X}_q)^{-1} \boldsymbol{X}_q^{'} \boldsymbol{z}_q^{(s)}$ (Mode B)

until convergence of $\boldsymbol{w}_q$

Step 2: Computation of the LVs

$\widehat{\boldsymbol{\xi}}_q \propto \boldsymbol{X}_q \boldsymbol{w}_q$

Step 3: Computation of the Path Coefficients

$\boldsymbol{\beta}_j = \left(\boldsymbol{\Xi}'_{\rightarrow j} \boldsymbol{\Xi}_{\rightarrow j}\right)^{-1} \boldsymbol{\Xi}'_{\rightarrow j} \widehat{\boldsymbol{\xi}}_j$

the model. However, we can bypass this problem using the *new Mode A* PLS-PM principles.

As we saw in section 1.4.5, *new Mode A* PLS-PM has been recently showed to optimize criterion 1.43 when centroid scheme is used, as well as criterion 1.44 when factorial scheme is used. However, Tenenhaus [2009] showed also that the value of the maximized criterion is always equal to

$$\lambda = \sum_q \text{cov}(\boldsymbol{v}_q, \boldsymbol{z}_q) \tag{6.4}$$

In the non-metric version of *new Mode A* PLS-PM scaling functions $\widetilde{\mathcal{Q}}(\widehat{\boldsymbol{x}}_{pq}, \boldsymbol{v}_q)$, $\widetilde{\widetilde{\mathcal{Q}}}(\widehat{\boldsymbol{x}}_{pq}, \boldsymbol{v}_q)$ and $\dot{\mathcal{Q}}(\widehat{\boldsymbol{x}}_{pq}, \boldsymbol{v}_q)$ can be showed to be optimal scaling functions, as they maximize model criterion. This result is obtained by re-writing the criterion as

$$\begin{aligned}
\lambda &= \sum_q \text{cov}(\boldsymbol{X}_q \boldsymbol{w}_q, \boldsymbol{z}_q) \\
&= \sum_q \frac{1}{N} \boldsymbol{w}_q' \boldsymbol{X}_q' \boldsymbol{z}_q \\
&= \sum_q \text{cov}^2(\boldsymbol{X}_q, \boldsymbol{z}_q) \\
&= \sum_q \sum_p^{P_q} \text{cov}^2(\boldsymbol{x}_{pq}, \boldsymbol{z}_q)
\end{aligned} \tag{6.5}$$

Wold's PLS-PM algorithm implemented using the *new Mode A*, monotonically converges to this criterion. In an optimal scaling framework, this criterion has to be maximized with respect to scaling pa-

rameters too. In order to obtain optimal quantifications with respect to criterion 6.5, we have to maximize, for fixed $\boldsymbol{z}_q$, the quantity

$$\lambda = \sum_q \sum_p^{P_q} \mathrm{cov}^2(\widehat{\boldsymbol{x}}_{pq}, \boldsymbol{z}_q)$$

with respect to $\widehat{\boldsymbol{x}}_{pq}$, normalized to unit variance and constrained to the scale level analysis for variable $\boldsymbol{x}^*_{pq}$.

Criterion 6.5 consists in a sum of criteria λ_{pq}, each of which is a function of a single scaled variable. Hence, it can be maximized by separately maximizing each criterion

$$\begin{aligned} \lambda_{pq} &= \mathrm{cov}^2(\widehat{\boldsymbol{x}}_{pq}, \boldsymbol{z}_q) \\ &= \mathrm{var}(\boldsymbol{z}_q)\mathrm{cor}^2(\widehat{\boldsymbol{x}}_{pq}, \boldsymbol{z}_q) \end{aligned} \tag{6.6}$$

with respect to $\boldsymbol{x}_{pq}$), normalized to unit variance.

Since $\mathrm{var}(\boldsymbol{z}_q)$ is a constant in each λ_{pq}, the optimization problem is solved by maximizing, for each $\widehat{\boldsymbol{x}}_{pq}$,

$$\sum_p \mathrm{cor}^2(\widehat{\boldsymbol{x}}_{pq}, \boldsymbol{z}_p) \tag{6.7}$$

Hence, each criterion λ_{pq} is optimized, under suitable scale level analysis constraints, by scaling each raw MV by means of one among the quantification functions 6.1, 6.2 and 6.2.

A procedure alternating the optimization of criterion 6.5 with respect to model parameters by means of *new Mode A* PLS loop and

with respect to scaling parameters by means of functions $\widetilde{\mathcal{Q}}(\boldsymbol{x}^*_{pq}, \boldsymbol{v}_q)$, $\widetilde{\widetilde{\mathcal{Q}}}(\boldsymbol{x}^*_{pq}, \boldsymbol{v}_q)$ and $\dot{\mathcal{Q}}(\boldsymbol{x}^*_{pq}, \boldsymbol{v}_q)$ can be used.

Since in each step of such a procedure criterion 6.5 is optimized, algorithm will converge to both optimal model parameters and optimal scaling parameters.

However, in order to avoid unuseful computational burden, a quantification step can be directly inserted in *new Mode A* PLS loop, exactly as shown in algorithm 11.

6.6 The interpretation of the weights

In MN-PLSPM the outer weights of quantified non-metric variable can be interpreted as a function of both quantified and raw MVs.

Since quantified MVs are standardized to unitary variance by construction, the outer weight vector w_{pq} for the p-th quantified MV $\widehat{\boldsymbol{x}}_{pq}$ of the q-th block is given by $\mathrm{cor}(\widehat{\boldsymbol{x}}_{pq}, \boldsymbol{z}_q)$ for less than a proportionality factor constant in each block.

When a raw MV $\boldsymbol{x}^*_{pq}$ is quantified at nominal scale level, the following relation holds:

$$\mathrm{cor}(\widehat{\boldsymbol{x}}_{pq}, \boldsymbol{z}_q) = \eta_{(\boldsymbol{z}_q|\boldsymbol{x}^*_p)}, \tag{6.8}$$

where η is the Pearson's correlation ratio, *i.e.* the part of variability of $\boldsymbol{z}_q$ explicated by the categories of $\boldsymbol{x}^*_{pq}$.
It is necessary to pay attention in interpreting these weights, because they are always not negative, since $0 \leq \eta \leq 1$. The weight referred

to a variable analyzed at a nominal scale level can be interpreted in terms of intensity, but not direction. After all, it is conceptually wrong expecting a sign in the relation between a numerical and a nominal variable, since a nominal variable neither increases, or decreases.

If a MV is quantified at an ordinal level, instead, the sign of corresponding weight can be interpreted. In fact, the following equations hold:

$$\text{cor}(\widehat{\boldsymbol{x}}_{pq}, \boldsymbol{z}_q) = \begin{cases} \sqrt{1 - STRESS1^2_{(\boldsymbol{z}_q, \boldsymbol{x}^*_{pq})}} & \text{if cor}(\widehat{\boldsymbol{x}}_{pq}, \boldsymbol{z}_q) \geq 0 \\ -\sqrt{1 - STRESS1^2_{(\boldsymbol{z}_q, \boldsymbol{x}^*_{pq})}} & \text{if cor}(\widehat{\boldsymbol{x}}_{pq}, \boldsymbol{z}_q) < 0 \end{cases}$$

From this relation it descends that if the sign it is positive there is a direct relation between the raw MV and the first LV. In the opposite case, it means that this relation is inverse. Moreover, intensity of the weight of a quantified variable measures the strength of the statistical relations of the LV with both the quantified and the raw variable: the first in terms of correlation, the second in terms of approaching to perfect monotonicity, intended as in Kruskal's secondary approach [Kruskal 1964*a*].

From these consideration, we can conclude that NM-PLSPM algorithm weights are coherent, in the sense that they reflects the statistical relation between the raw MVs and the corresponding LV inner estimate. This property makes much clearer their interpretation.

6.7 An application to macroeconomic data

The data for this example ate taken from a paper by Russet [1964]. The basic hypothesis in Russet's paper is that economic inequality leads to political instability. In particular in Russet model political instability is function of inequality of land distribution and of industrial development. Three variables are used to measure inequality of land distribution. Variable "gini" is the Gini's index of concentration, which measures the deviation of the Lorenz curve from the line of equality. Variable "farm" is the percentage of farmers that own half of the lands, starting with the smallest ones. Thus if "farm" is 90%, then 10% of the farmers own half of the land. The third indicator is "rent", which is the percentage of farm households that rent all their land. Two variables are used to measure industrial development: variable "gnpr" is the gross national product *pro capite* (in U.S. dollars) in 1955, and variable "labo" is the percentage of labor force employed in agriculture. Political stability is measured by four variables. Variable "inst" is a function of the number of the chiefs of the executive and of the number of years of independence of the country during the period 1946-1961. This index bounds between 0 (very stable) and 17 (very unstable). Variable "ecks" is the Eckstein's index, which measures the number of violent internal war incidents during the same period. Variable "death" is the number of people killed as a result of violent manifestations during the period 1950-1962. Variable "demo" classi-

fies countries in three groups: stable democracy, unstable democracy and dictatorship.

This data-set was analyzed in Gifi [1990] using program CANALS (Canonical Correlation Analysis by Alternating Least Squares). Variables were scaled in such a way to maximize the canonical correlation between the block of variables regarding the economic inequality and the block of variables regarding the political instability. However, Gifi himself noticed that partitioning data in three set of variables (agricultural inequality, industrial development and political instability) would have been a more rational approach.

6.7.1 Model estimation with standard PLS Path Modeling

Starting from this idea, Tenenhaus [1998] modeled the Russet dataset in a PLS-PM framework (see figure 6.1). He partitioned Russet data in three reflective blocks. The first block, consisting of variables "gini", "farm" and "rent" measures the latent concept "Agricultural Inequality". The second one, formed by variables "gnpr" and "labo", measures the latent concept "Industrial Development". The third block, composed by variables "inst", "ecks", "death" and "demo", expresses the latent concept "Political Instability". Relations between latent variables are modeled in the following way: Agricultural Inequality and Industrial Development predict Political Instability (see figure 6.1).

Since Gifi's analysis suggested an high degree of non-linearity of

Country	gini	farm	rent	gnpr	labo	inst	ecks	death	demo
Argentina	86.3	98.2	32.9	374	25	13.6	57	217	unstable
Australia	92.9	99.6	NA	1215	14	11.3	0	0	stable
Austria	74	97.4	10.7	532	32	12.8	4	0	unstable
Belgium	58.7	85.8	62.3	1015	10	15.5	8	1	stable
Bolivia	93.8	97.7	20	66	72	15.3	53	663	dict.
Brasil	83.7	98.5	9.1	262	61	15.5	49	1	unstable
Canada	49.7	82.9	7.2	1667	12	11.3	22	0	stable
Chile	93.8	99.7	13.4	180	30	14.2	21	2	unstable
Colombia	84.9	98.1	12.1	330	55	14.6	47	316	unstable
CostaRica	88.1	99.1	5.4	307	55	14.6	19	24	unstable
Cuba	79.2	97.8	53.8	361	42	13.6	100	2900	dict.
Denmark	45.8	79.3	3.5	913	23	14.6	0	0	stable
Domin. Rep.	79.5	98.5	20.8	205	56	11.3	6	31	dict.
Ecuador	86.4	99.3	14.6	204	53	15.1	41	18	dict.
Egypt	74	98.1	11.6	133	64	15.8	45	2	dict.
Salvador	82.8	98.8	15.1	244	63	15.1	9	2	dict.
Finland	59.9	86.3	2.4	941	46	15.6	4	0	unstable
France	58.3	86.1	26	1046	26	16.3	46	1	unstable
Guatemala	86	99.7	17	179	68	14.9	45	57	dict.
Greece	74.7	99.4	17.7	239	48	15.8	9	2	unstable
Honduras	75.7	97.4	16.7	137	66	13.6	45	111	dict.
India	52.2	86.9	53	72	71	3	83	14	stable
Irak	88.1	99.3	75	195	81	16.2	24	344	dict.
Ireland	59.8	85.9	2.5	509	40	14.2	9	0	stable
Italy	80.3	98	23.8	442	29	15.5	51	1	unstable
Japan	47	81.5	2.9	240	40	15.7	22	1	unstable
Libia	70	93 8.	5	90	75	14.8	8	0	dict.
Luxemburg	63.8	87.7	18.8	1194	23	12.8	0	0	stable
The Netherl.	60.5	86.2	53.3	708	11	13.6	2	0	stable
New Zealand	77.3	95.5	22.3	1259	16	12.8	0	0	stable
Nicaragua	75.7	96.4	NA	254	68	12.8	16	16	dict.
Norway	66.9	87.5	7.5	969	26	12.8	1	0	stable
Panama	73.7	95	12.3	350	54	15.6	29	25	dict.
Peru	87.5	96.9	NA	140	60	14.6	23	26	dict.
Philippine	56.4	88.2	37.3	201	59	14	15	292	dict.
Poland	45	77.7	0	468	57	8.5	19	5	dict.
S. Vietnam	67.1	94.6	20	133	65	10	50	1000	dict.
Spain	78	99.5	43.7	254	50	0	22	1	dict.
Sweden	57.7	87.2	18.9	1165	13	8.5	0	0	stable
Switzerland	49.8	81.5	18.9	1229	10	8.5	0	0	stable
Taiwan	65.2	94.1	40	132	50	0	3	0	dict.
UK	71	93.4	44.5	998	5	13.6	12	0	stable
USA	70.5	95.4	20.4	2343	10	12.8	22	0	stable
Uruguay	81.7	96.6	34.7	569	37	14.6	1	1	stable
Venezuela	90.9	99.3	20.6	762	42	14.9	36	111	dict.
W. Germany	67.4	93	5.7	762	14	3	4	0	unstable
Yugoslavia	43.7	79.8	0	297	67	0	9	0	dict.

Table 6.1: Russet data-set

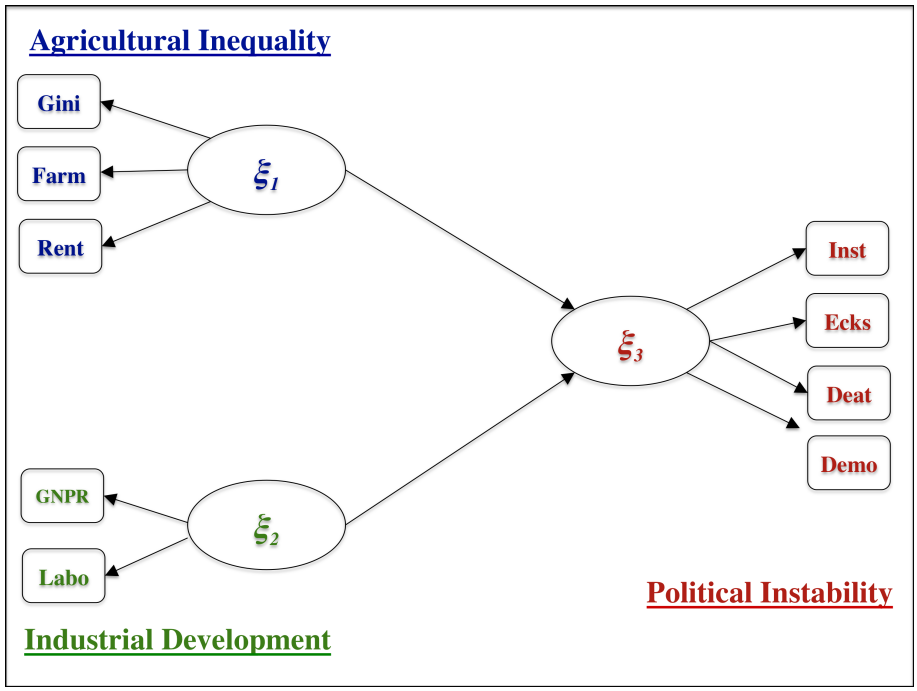

Figure 6.1: Russet data as modeled by Tenenhaus

data, Tenenhaus approximated CANALS scalings by means of monotone functional transformations. Variables "rent" "gnpr" , "labo", "ecks" and "death" were transformed as functions of respective standardized logarithms. In particular, new variables l_rent = ln(rent), l_gnpr = ln(gnpr), l_labo = ln(labo), l_ecks = ln(ecks+1), and l_death = ln(death + 1) replaced the old ones. Variable "inst" was transformed according to the exponential rule (*i.e.* as e_ins = $exp^{\text{inst}-16.3}$) and standardized. Finally, variables "gini" and "farm" were just standardized. Since variable "demo" is categorical, it was replaced by the three dummy variables "d-stb", "d-inst", and "dict" corresponding to its categories.

Tenenhaus performed a PLS-PM analysis on the model defined in figure 6.1 by using the option *centroid* for inner weight estimation and handling all the blocks as reflective. We run the same analysis by using an R code developed by the author.

Quality of Tenenhaus' model is assessed looking at table 6.2. As regard to the inner model, a good part of the variability of the latent response $\boldsymbol{\xi}_3$ ("Political Instability") is explained by the two latent predictors, with an R^2 value of 0.622. With respect to the quality of the outer model the mean Communalities of exogenous blocks are satisfying. However, the LV "Political Instability" only explain the 45.2% of its own MV variability.

Parameters estimates are represented in figure 6.2. It is possible to investigate the relations between Agricultural Inequality, Industrial Development and Political Instability through the path coefficients

LV	R^2	Mean Comm.	Mean Red.
$\boldsymbol{\xi}_1$		0.731	
$\boldsymbol{\xi}_2$		0.907	
$\boldsymbol{\xi}_3$	0.622	0.452	0.282

Table 6.2: PLS-PM analysis of Russet data as transformed by Tenenhaus: model assessment

represented in figure; obviously, the two latent predictors impact in opposite sense on the response. However, Political Instability largely depends on Industrial Development than on Agricultural Inequality. The higher the Industrial Development is, the smaller the Political Instability is.

As one can expect, variables "gini", "farm" and "l_rent" are positively correlated to the LV $\boldsymbol{\xi}_1$, which measures the Agricultural Inequality. LV Industrial Development, is positively affected by the gross national product (variable "l_gnpr") and negatively affected by the percentage of agricultural workers (variable "l_labo"). All of the MVs of the block representing Political Instability positively impact on the LV $\boldsymbol{\xi}_3$ but binary variable "d-stb", which indicates the countries with a stable democratic regime.

It is not clear if the weight of variable "demo", expressed by the three dummy "d-stb", "d-inst" and "dict", is high or low. While weights of "d-stb" and "dict" are large, the weight of "d-inst" is almost zero (see table 6.3). As matter of fact the weight of the binary variable "d-inst" is so small just because there is a strong relation between the categorical variable "demo" and the LV Political Insta-

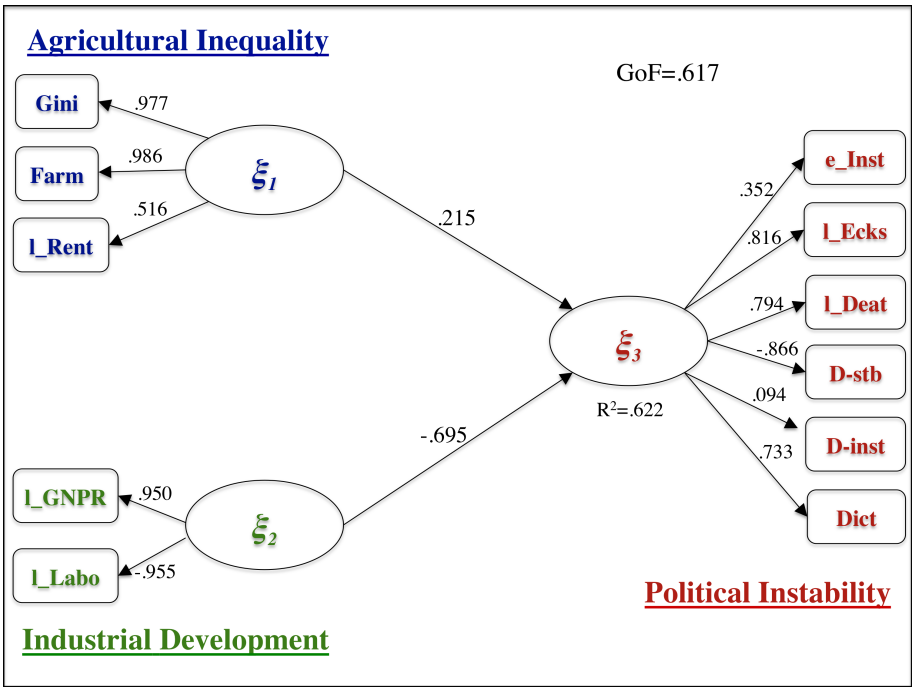

Figure 6.2: PLS-PM analysis of Russet data as transformed by Tenenhaus: model parameter estimates

LV	MV	Outer weights	Stand. load.	Comm.	Red.
$\boldsymbol{\xi}_1$	gini	0.460	0.977	0.955	
	farm	0.516	0.986	0.972	
	l_rent	0.081	0.516	0.266	
$\boldsymbol{\xi}_2$	l_gnpr	0.511	0.950	0.903	
	l_labo	-0.538	-0.955	0.912	
$\boldsymbol{\xi}_3$	e_inst	0.104	0.352	0.124	0.077
	l_ecks	0.270	0.816	0.665	0.414
	l_death	0.302	0.794	0.630	0.392
	d-stb	-0.336	-0.866	0.749	0.466
	d-inst	0.037	0.094	0.009	0.006
	dict	0.285	0.733	0.537	0.334

Table 6.3: PLS-PM analysis of Russet data as transformed by Tenenhaus: outer model results

bility. In fact, category "d-stb" is mainly associated to observation sharing the lowest values of $\boldsymbol{\xi}_3$, while category "dict" is mainly associated to observations sharing the highest values of the LV and category "d-inst" is mainly associated to observation sharing the central values of political instability score distribution. Hence, there are a strong relation between $\boldsymbol{\xi}_3$ and all of the binary variables representing the categories of MV "demo". Unfortunately, while relations between binary variable "dict" and "d-stb" and $\boldsymbol{\xi}_1$ are pretty monotone (and so they can easily approximated by a linear function), binary variable "d-inst" is linked to $\boldsymbol{\xi}_1$ by a non-monotonic relation (see figure 6.3). As a consequence, this variable is underestimated in the model.

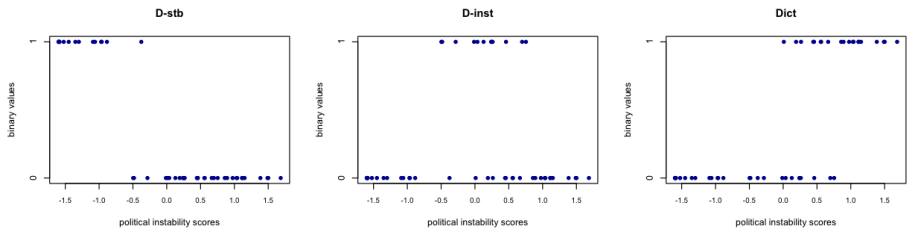

Figure 6.3: Raw values of binary variables corresponding to categories of variable "demo" plotted versus the LV Political Instability values

6.7.2 Model estimation with Non-Metric PLS Path Modeling

In order to overcame the binary coding drawbacks, we perform two Non-Metric PLS-PM analyses on Russet data-set by using an R code developed by the author (see Appendix). In the first analysis, we let metric variables as transformed by Tenenhaus while the non-metric variable "demo" will be properly quantified (see subsection 6.7.2). In the second, we provide new transformations for all of the original variables of the Russet data-set (see subsection 6.7.2).

Analyzing a variable at a nominal scaling level

In this NM-PLSPM analysis, variable "demo" is analyzed at a nominal scaling level; for all the other variables we keep Tenenhaus' transformations and we analyze them at a linear scaling level (*i.e.* we simply standardize them). The new model is represented in figure 6.4. Now

the LV political instability is expressed by just four MVs: "e_inst", "l_ecks", "l_death" and "demo". The quality of this model is summa-

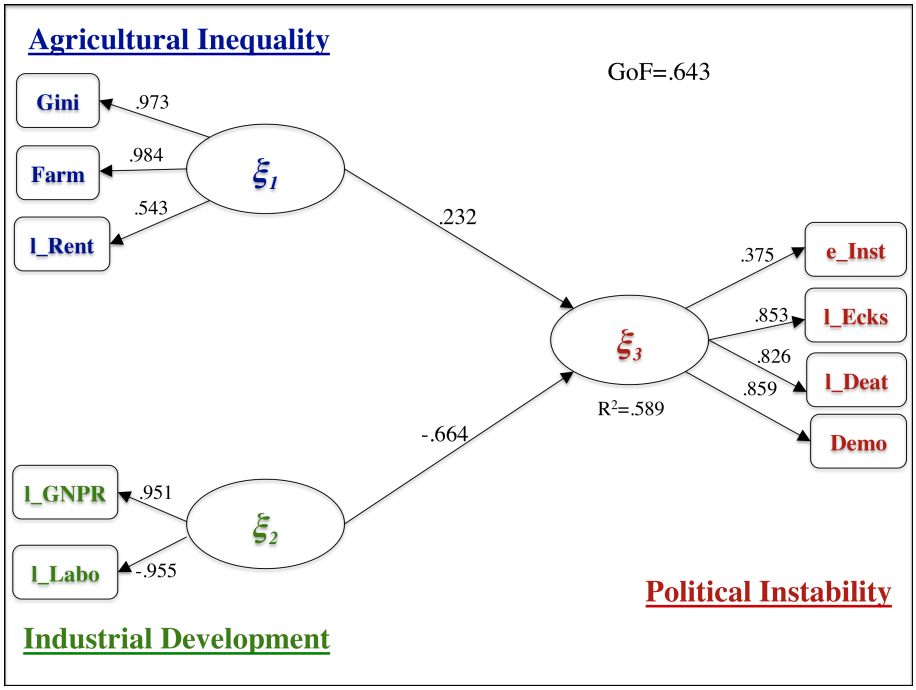

Figure 6.4: NM-PLSPM analysis of Russet data as transformed by Tenenhaus (variable "demo" is analyzed at a nominal scaling level): model parameter estimates

rized in table 6.4. With respect to the previous, this model loses in predictive capability of the latent response, while gains in explicative capability of the MV underlying the concept of Political Instability. The mean Communalities of the other two blocks remain about the

LV	R^2	Mean Comm.	Mean Red.
$\boldsymbol{\xi}_1$		0.737	
$\boldsymbol{\xi}_2$		0.908	
$\boldsymbol{\xi}_3$	0.589	0.572	0.337

Table 6.4: NM-PLSPM analysis of Russet data as transformed by Tenenhaus (variable "demo" is analyzed at a nominal scaling level): model assessment

same. However, the global model fit improves, as GoF passes from 0.617 to 0.643.

The non-metric analysis makes it clear that MV "demo" is the most important in the construction of the LV Political Instability (see table 6.5). According to these results we can conclude that the categories of the MV "demo" are very discriminant with respect to the Political Instability scores. In fact, the weight of a MV quantified at a nominal scaling level reflects the variability of the corresponding LV explained by the categories of the MV (see section 6.6).

Exploring non linearity by means of monotone transformations

Tenenhaus himself pinpointed that approximating CANALS transformations is not the better choice, as they are optimized for canonical correlation, while transformations optimized for PLS-PM would be preferable. In order to have monotone quantification that are coherent with the model, we perform a second NM-PLS path model, where

LV	MV	Outer weights	Stand. load.	Comm.	Red.
$\boldsymbol{\xi}_1$	gini	0.455	0.973	0.947	
	farm	0.502	0.984	0.968	
	l_rent	0.117	0.543	0.294	
$\boldsymbol{\xi}_2$	l_gnpr	0.514	0.951	0.904	
	l_labo	-0.536	-0.955	0.911	
$\boldsymbol{\xi}_3$	e_inst	0.127	0.375	0.140	0.083
	l_ecks	0.329	0.853	0.728	0.429
	l_death	0.370	0.826	0.682	0.402
	demo	0.427	0.859	0.739	0.435

Table 6.5: NM-PLSPM analysis of Russet data as transformed by Tenenhaus (variable "demo" is analyzed at a nominal scaling level): outer model results

all quantitative MVs are analyzed at an ordinal scaling level, and the MV "demo" is analyzed at a nominal scaling level.

The resulting quantifications (see figure 6.6) yield a sensibly better model ($GOF = 0.794$). The interpretation of the inner relation does not change: the impact on Political Instability of Industrial Developement ($\widehat{\beta} = -0.716$) is higher than the one of Agricultural Inequality ($\widehat{\beta} = 0.291$). However, the multiple determination index of the regression sensibly increases, as well the mean Redundancy of the MVs connected to $\boldsymbol{\xi}_3$ (see figure 6.5).

As regards the outer model, we notice a substantial improvement of the mean Communality of the endogenous block; also the mean Communality of block referring to LV $\boldsymbol{\xi}_2$ increases, while the capability of $\boldsymbol{\xi}_1$ in explaining its own MVs remains substantially stable (see table

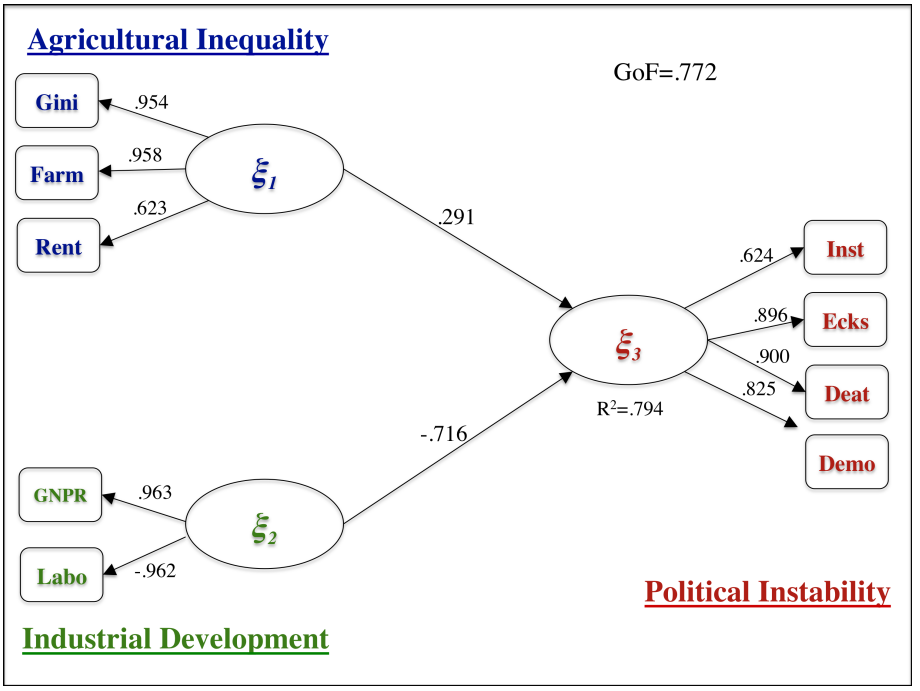

Figure 6.5: Non-Metric PLS-PM analysis of Russet data (all of the manifest variables are properly quantified): model parameter estimates

LV	R^2	Mean Comm.	Mean Red.
$\boldsymbol{\xi}_1$		0.739	
$\boldsymbol{\xi}_2$		0.927	
$\boldsymbol{\xi}_3$	0.794	0.671	0.532

Table 6.6: Non-Metric PLS-PM analysis of Russet data (all of the manifest variables are properly quantified): model assessment

LV	MV	Outer weights	Stand. load.	Comm.	Red.
$\boldsymbol{\xi}_1$	gini	0.425	0.954	0.910	
	farm	0.454	0.958	0.917	
	rent	0.256	0.623	0.389	
$\boldsymbol{\xi}_2$	gnpr	0.523	0.963	0.928	
	labo	-0.516	-0.962	0.926	
$\boldsymbol{\xi}_3$	inst	0.201	0.624	0.390	0.310
	ecks	0.310	0.896	0.802	0.637
	death	0.358	0.900	0.810	0.643
	demo	0.332	0.825	0.680	0.540

Table 6.7: Non-Metric PLS-PM analysis of Russet data (all of the manifest variables are properly quantified): outer model results

6.6). Moreover, we observe an improvement of the loading estimates of the worst modeled variables in previous analysis; we refer in particular to variables "rent" and "inst". This improvement is reflected also in the corresponding outer weights, which in standard PLS-PM express the degree of linearity of the relation between each MV and the corresponding LV, while in NM-PLSPM indicate the degree of monotonicity of this relation (see table 6.7).

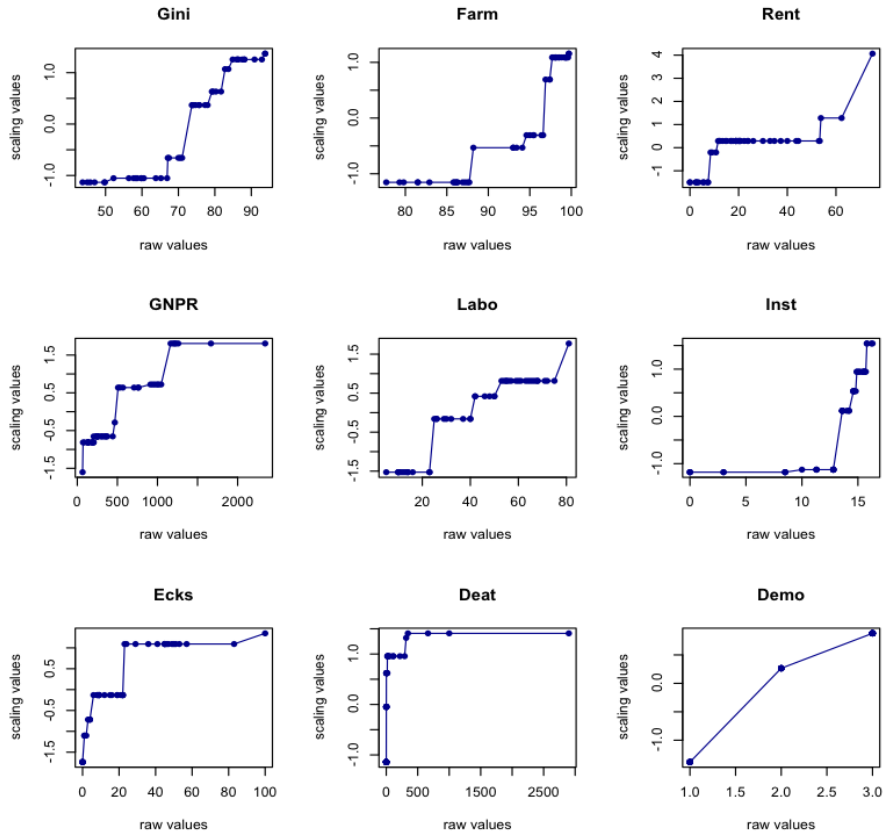

Figure 6.6: Non-Metric PLS-PM analysis of Russet data (all of the manifest variables are properly quantified): Raw values vs Scaling values

6.7.3 Conclusion

Previous analyses showed that NM-PLSPM is a valid tool to obtain coherent models when we observe variables measured on a variety of measurement scale, as well as when we want to discard linearity hypothesis in relations between the MVs and the corresponding LV. In fact, a milder hypothesis of monotonicity can be adopted in a non-metric approach. In general, we can state that NM-PLS Path Models provides better models, since MV are transformed in such a way to make relations between manifest and latent variables linear.

Conclusion

In 1966 Herman Wold proposed the estimation of principal components and related models by means of a Non-linear Iterative Partial Least Squares procedure. Ten years later, for the first time an iterative algorithm based on an Alternating Least Squares (ALS) procedure was proposed by Jan de Leeuw, Yoshio Takane and Forrest W. Young for implementing optimal scaling in additive structure analysis. They themselves noticed, referring to ALS, that "*this type of procedure is philosophically much like the NILES/NIPALS procedure developed by Wold and his associates with the distinction that Wold is usually concerned with optimizing only model parameters*" [de Leeuw et al. 1976].

It is surprising how along the last 33 years these two procedures have been developed in a parallel way, never crossing each other.

ALS have become the most used procedure for optimal scaling in joint non-parametric multivariate analysis of non-metric and metric data. A whole system of non-linear multivariate analysis, working on ALS principles, was developed by the data theory group of Leiden University [Gifi 1990].

NIPALS algorithm has been developed in order to implement Principal Component Analysis, Canonical Correlation Analysis, Redundancy Analysis, Multiple Factorial Analysis, Canonical Correlation Analysis and Generalized Canonical Correlation Analysis. PLS Regression and PLS Path Modeling, instead, are new methods, developed to perform respectively Regularized Regression and Structural Equation Models in a soft modeling framework.

Nowadays, "*among the open issue that currently represent the most important and promising research challenges in PLS-PM,*" there is the "*specific treatment of categorical (nominal and ordinal) variable and specific treatment of non linearity*" [Esposito Vinzi et al. 2009].

In this work we found out how NIPALS based algorithms, properly adjusted, can work as optimal scaling algorithms. This new feature of PLS, which had been until now totally unexplored, allowed us to device a new suite of PLS methods: the Non-Metric PLS (NM-PLS) methods.

NM-PLS methods can be used with different aims:

- to analyze at the same time variables observed on different measurement scales;
- to investigate non linearity;
- to discard the hard assumption of linearity in favor of a milder assumption of monotonicity.

In particular, these methods generalize standard NIPALS, PLS Regression and PLS Path Modeling in order to handle variables observed

on a variety of measurement scales, as well as to cope with non linearity problems.

Three new algorithms have been proposed to implement NM-PLS methods: the Non-Metric NIPALS algorithm, the Non-Metric PLS Regression algorithm, and the Non-Metric PLS Path Modeling algorithm.

All these algorithms provide at the same time specific PLS model parameters as well as scaling values for variables to be scaled.

Scaling values provided by these algorithms have been proved to be optimal, in the sense that they optimize the same criterion of the model in which they are involved. Moreover, they are suitable, since they respect the constraints depending on which among the properties of the original measurement scale we want to preserve.

Further studies on stability of results have to be done in future research, as well as to investigate if NM-PLS algorithms converge to global or local optima. Moreover, future research challenges in NM-PLS methods involve the extension of these methods to:

- optimal quantifications for H-dimensional models;
- monotone polynomials and splines transformations;
- *Mode B* PLS-PM.

Appendix

A.1 R code for Non-Metric PLS Regression

```
myPLSQQ <- function(Y=NA,Yc=NA,X=NA,Xc=NA,ncomp)
 {
  if (is.na(Y)==F)
   {
    n<-nrow(Y)
    rownamesY<-rownames(Y)
   }
   else
   {
    n<-nrow(Yc)
    rownamesY<-rownames(Yc)
   }
   ncolX<-0
   if (is.na(X)==F)
   {
    ncolX<-ncol(as.matrix(X))
   }
   ncolXc<-0
   if (is.na(Xc)==F)
   {
    ncolXc<-ncol(as.matrix(Xc))
```

```
}
p<-numeric()
if (is.na(Xc)==F && is.na(X)==F)
{
 p<-ncolX+ncolXc
}
if (is.na(Xc)==F && is.na(X)==T)
{
 p<-ncolXc
}
if (is.na(X)==F && is.na(Xc)==T)
{
 p<-ncolX
}
ncolY<-0
if (is.na(Y)==F)
{
 ncolY<-ncol(as.matrix(Y))
}
ncolYc<-0
if (is.na(Yc)==F)
{
 ncolYc<-ncol(as.matrix(Yc))
}
q<-numeric()
if (is.na(Yc)==F && is.na(Y)==F)
{
 q<-ncolY+ncolYc
}
if (is.na(Yc)==F && is.na(Y)==T)
{
 q<-ncolYc
}
if (is.na(Y)==F && is.na(Yc)==T)
{
 q<-ncolY
}
a<-ncomp
Q<-matrix(,n,ncolXc)
```

```
Qy<--matrix(,n,ncolYc)
W<-matrix(,p,a)
rownames(W)<-c(colnames(X),colnames(as.matrix(Xc)))
U<-matrix(,n,a)
U<-matrix(c(rep(c(1,rep(0,(n-1))),a)),n,a)
#if (is.na(Y)==F) { U[,1]<-Y[,1]}
#else {U[,1]<-c(1,rep(0,(n-1)))}
rownames(U)<-rownamesY
T<-matrix(,n,a)
rownames(T)<-rownamesY
C<-matrix(,q,a)
rownames(C)<-c(colnames(Y),colnames(Yc))
P<-matrix(,p,a)
rownames(P)<-c(colnames(X),colnames(Xc))
W_star<-matrix(,p,a)
b<-matrix(,a,1)
B<-matrix(,p,q)
Pcorr<-matrix(,p,a)
rownames(Pcorr)<-c(colnames(X),colnames(Xc))
Ccorr<-matrix(,q,a)
rownames(Ccorr)<-c(colnames(Y),colnames(Yc))
Tcorr<-matrix(,n,a)
rownames(Tcorr)<-rownamesY
Xi<-X
Yi<-Y
Xarray<-array(,c(n,p,a))
Yarray<-array(,c(n,q,a))
for (i in 0:(a-1))
{
 ncicli<-0
 repeat
 {
  Ustart<-U[,i+1]
  if (i==0)
  {
  if (is.na(Xc)==F)
  {
  Q<-dummy.G(U[,i+1],Xc)$Quant
  Q<-myScale(Q)
```

```
   colnames(Q)<-colnames(Xc)
   if (is.na(X)==F)
   {
    Xi<-cbind(Xi[,1:ncolX],Q)
   }
   else
   {
    Xi<-Q
   }
  }
 }
 W[,i+1]<-as.matrix((t(Xi)%*%U[,i+1])/as.numeric(t(U[,i+1])%*%U[,i+1]))
 W[,i+1]<-W[,i+1]/sqrt(as.numeric(t(W[,i+1])%*%W[,i+1]))
 T[,i+1]<-(Xi%*%W[,i+1])/as.numeric(t(W[,i+1])%*%W[,i+1])
 if (i==0)
 {
  if (is.na(Yc)==F)
  {
   Qy<-dummy.G(T[,i+1],Yc)$Quant
   Qy<-myScale(Qy)
   if (is.na(Y)==F)
   {
    Yi<-cbind(Yi[,1:ncolY],Qy)
   }
   else
   {
    Yi<-Qy
   }
  }
 }
 C[,i+1]<-(t(Yi)%*%T[,i+1])/as.numeric(t(T[,i+1])%*%T[,i+1])
 C[,i+1]<-C[,i+1]/sqrt(as.numeric(t(C[,i+1])%*%C[,i+1]))
 U[,i+1]<-(Yi%*%C[,i+1])/as.numeric(t(C[,i+1])%*%C[,i+1])
 conv<-max(abs(Ustart-U[,i+1]))
 print("conv");print(conv);
 ncicli<-ncicli+1
 if (conv<0.0000001 | ncicli>149) {break}
}
print("numero cicli"); print(ncicli);
```

```
 P[,i+1]<-t(Xi)%*%T[,i+1]/as.numeric(t(T[,i+1])%*%T[,i+1])
 Xi<-Xi-(T[,i+1]%*%t(P[,i+1]))
 Xarray[,,i+1]<-Xi
 b[i+1,1]<-(t(U[,i+1])%*%T[,i+1])/(t(T[,i+1])%*%T[,i+1])
 print("inner coefficient: ");print(b[i+1,1])
 Yi<-Yi-(as.numeric(b[i+1,1])*T[,i+1]%*%t(C[,i+1]))
 Yarray[,,i+1]<-Yi
}
W_star<-W
rownames(Q)<-rownamesY
colnames(Q)<-colnames(Xc)
rownames(B)<-c(colnames(X),colnames(Xc))
colnames(B)<-c(colnames(Y),colnames(Yc))
rownames(W_star)<-c(colnames(X),colnames(Xc))
if (is.na(Y)==F)
{
 newY<-cbind(Y,Qy)
}
else
{
 newY<-Qy
}
if (is.na(X)==F)
{
 newX<-cbind(X,Q)
}
else
{
 newX<-Q
}
R2X<-1-(sum(apply(as.matrix(Xarray[,,a]),2,var))/sum(apply(newX,2,var)))
R2Y<-1-(sum(apply(as.matrix(Yarray[,,a]),2,var))/sum(apply(newY,2,var)))
if (a>1)
{
 W_star<-W%*%solve(t(P)%*%W)
 Pcorr<-P%*%(diag(apply(T,2,sd)))
 Ccorr<-C%*%(diag(apply(T,2,sd)))
 diag_matr<-diag(1/(apply(T,2,sd)*sqrt(n-1)))
 Tcorr<-T%*%(diag_matr)
```

```
IDYarray<-array(,c(n,2,q))
lista_Ymedie<-list()
for (j in 1:q)
{
 IDYarray[,1,j]<-(T%*%(C[j,]))/sqrt(1+(Ccorr[j,2]/Ccorr[j,1])^2)
 IDYarray[,2,j]<-IDYarray[,1,j]*(Ccorr[j,2]/Ccorr[j,1])
}
if (is.na(Yc)==F)
{
 for (k in 1:ncolYc)
 {
  matrice_Ymedie<-matrix(,max(as.matrix(Yc)[,k]),2)
  matrice_Ymedie[,1]<-as.vector(tapply(IDYarray[,1,k+ncolY],Yc[,k],mean,na.rm=T))
  matrice_Ymedie[,2]<-as.vector(tapply(IDYarray[,2,k+ncolY],Yc[,k],mean,na.rm=T))
  lista_Ymedie[[k]]<-matrice_Ymedie
 }
}
IDarray<-array(,c(n,2,p))
lista_medie<-list()
for (j in 1:p)
{
 IDarray[,1,j]<-(T%*%(P[j,]))/sqrt(1+(Pcorr[j,2]/Pcorr[j,1])^2)
 IDarray[,2,j]<-IDarray[,1,j]*(Pcorr[j,2]/Pcorr[j,1])
}
if (is.na(Xc)==F)
{
 for (k in 1:ncolXc)
 {
  matrice_medie<-matrix(,max(as.matrix(Xc)[,k]),2)
  matrice_medie[,1]<-as.vector(tapply(IDarray[,1,k+ncolX],Xc[,k],mean,na.rm=T))
  matrice_medie[,2]<-as.vector(tapply(IDarray[,2,k+ncolX],Xc[,k],mean,na.rm=T))
  lista_medie[[k]]<-matrice_medie
 }
}
B<-W_star%*%diag(b[,1])%*%t(C)
VIP<-matrix(,p,1)
rownames(VIP)<-rownames(B)
for (j in 1:p)
{
```

```
   SumRdW2<-0
   SumRd<-0
   for (h in 1:a)
   {
    Rd<-0
    for (k in 1:q)
    {
     Rd<-Rd+(cor(newY[,k],T[,h])^2)
    }
    SumRdW2<-SumRdW2+((Rd)*(as.numeric(W[j,h])^2))
    SumRd<-SumRd+(Rd)
   }
   VIP[j,1]<-sqrt(p*SumRdW2/SumRd)
  }
  list(Q=Q,Qy=Qy,U=U,T=T,C=C,P=P,W=W,b=b,B=B,W_star=W_star,Pcorr=Pcorr,Ccorr=Ccorr,
       Tcorr=Tcorr, Xarray=Xarray, Yarray=Yarray, IDarray=IDarray, lista_medie=lista_medie,
       IDYarray=IDYarray, lista_Ymedie=lista_Ymedie,VIP=VIP,R2X=R2X,R2Y=R2Y)
 }
 else
 {
  B<-as.numeric(b[1,1])*W%*%t(C)
  list(Q=Q,Qy=Qy,U=U,T=T,C=C,P=P,W=W,W_star=W_star,b=b,B=B, Xarray=Xarray,
       Yarray=Yarray, R2X=R2X, R2Y=R2Y)
 }
}
```

A.2 R code for Non-Metric PLS-Path Modeling

```
myPLSPM<-function(X, p_blocchi, path, scaling=NA)
 {
  if (is.na(scaling)==T)
  {
   scaling<-vector("list", length(p_blocchi))
   for (i in 1:length(scaling))
   {
    scaling[[i]]<-c(rep("NUM",p_blocchi[i]))
   }
  }
  X <- as.matrix(X)
  path <- as.matrix(path)
  link <- t(path)+path
  N <- nrow(X)
  P<- ncol(X)
  blocchi<-list()
  mean_X <-list()
  var_X <- list()
  correzione<-(sqrt((N-1)/N))
  QQ <- list()
  p_blocchi<-c(1,p_blocchi)
  for (q in 1:(length(p_blocchi)-1))
  {
   blocchi[[q]]<-as.matrix(X[,(sum(p_blocchi[1:q])):(sum(p_blocchi[1:q])+p_blocchi[q+1]-1)])
   QQ[[q]] <- blocchi[[q]]
  }
  p_blocchi<-p_blocchi[2:length(p_blocchi)]
  nbloc<-length(p_blocchi)
  w <- vector("list", nbloc)
  z <- vector("list", nbloc)
  for (q in 1:nbloc)
  {
   z[[q]]<-scale(svd(scale(blocchi[[q]]))$u[,1])*correzione
   w[[q]]<-c(rep(1,p_blocchi[q]))
  }
```

```
y <- vector("list", nbloc)
e <- matrix(,nbloc,nbloc)
converg<-numeric()
ncicli<-0
z_temp<-matrix(0,N,1)
#####################################################################
#                        iterative cycle                            #
#####################################################################
repeat
{
 ncicli<-ncicli+1
 w_old <- w[[nbloc]]
 # --- MV quantification ["QQ"] ---- #
 for (q in 1:(nbloc))
 {
  for (p in 1:(p_blocchi[q]))
  {
   if (scaling[[q]][p]=="NOM")
   {
    QQ[[q]][,p]<-dummy.G(z[[q]],(blocchi[[q]][,p]))$Quant
    QQ[[q]][,p]<-scale(QQ[[q]][,p])*correzione
   }
   if (scaling[[q]][p]=="NUM")
   {
    QQ[[q]][,p]<-scale(QQ[[q]][,p])*correzione
   }
   if (scaling[[q]][p]=="ORD")
   {
    eta2_temp<- (dummy.ord(z[[q]],(blocchi[[q]][,p]))$eta2)
    if (  eta2_temp < (dummy.ord_decr(z[[q]],(blocchi[[q]][,p]))$eta2))
    {
     QQ[[q]][,p] <- -dummy.ord_decr(z[[q]],(blocchi[[q]][,p]))$Quant
    }
    else { QQ[[q]][,p] <- (dummy.ord(  z[[q]],(blocchi[[q]][,p]))$Quant)}
    QQ[[q]][,p]<-scale(QQ[[q]][,p])*correzione
   }
   if (scaling[[q]][p]=="RAW")
   {
    QQ[[q]][,p]<-QQ[[q]][,p]
```

```
   }
  }
 }
 # --- updating the weights ["w"]: REFECTIVE WAY ---- #
 for (q in 1:nbloc)
 {
  w[[q]]<-(1/N)*(t(QQ[[q]]) %*% z[[q]])
 }
 # --- updating the weights ["w"]:FORMATIVE WAY ---- #
 #for (q in 1:nbloc)
 #{
 # w[[q]]<-solve(t(QQ[[q]]) %*% QQ[[q]]) %*% t(QQ[[q]]) %*% z[[q]]
 #}
 # --- outer estimations ["y"] ---- #
 for (q in 1:nbloc)
 {
  y[[q]] <- QQ[[q]] %*% w[[q]]
  y[[q]] <- scale(y[[q]])*correzione
 }
 # --- updating the weights ["e"] ---- #
 for (q in 1:nbloc)
 {
 z[[q]] <- z_temp
 for (k in 1:nbloc)
 {
  e[q,k]<-cor(y[[q]],y[[k]])
  ##############################################################################
  #   if centroid approach, ok; in factorial approach, delete next 2 raws  #
  ##############################################################################
  if (e[q,k]>0) {e[q,k]<-1}
  else {e[q,k]<- -1}
  z[[q]]<-(z[[q]])+(link[q,k]*e[q,k]*y[[k]])
 }
 z[[q]]<-scale(z[[q]])*correzione
}
converg <- sum((w_old-w[[nbloc]])^2)
print("converg")
print(converg)
print("ncicli")
```

```
 print(ncicli)
 if (converg<0.0000001 | ncicli>51)  {break}
}
#################################################################
#   computation of the LV scores using the outer weigts w     #
#################################################################
VL <- list()
sqm_VL <- array(, nbloc)
w_tilde <- list()
abs_w_tilde<-list()
VLS <- list()
somma_w_tilde<-array(,nbloc)
w_tilde_normal <- list()
for (q in 1:nbloc)
{
 VL[[q]] <- QQ[[q]] %*% w[[q]]
 sqm_VL[q] <- sd(VL[[q]])*sqrt((N-1)/N)
 w_tilde[[q]] <- w[[q]]/as.numeric(sqm_VL[q])
 VLS[[q]] <- QQ[[q]]  %*% w_tilde[[q]]
 abs_w_tilde[[q]] <- abs(w_tilde[[q]])
 somma_w_tilde[[q]] <- sum(abs_w_tilde[[q]])
 w_tilde_normal[[q]] <- w_tilde[[q]]/somma_w_tilde[[q]]
}
# ----- the LVs are standardized  ----- #
#############################################################################
# computation of the correlation between each LV and the corresponding MVs #
#############################################################################
CORR_VL <- list()
COMM_vm <- list()
COMM <- list()
for (q in 1:nbloc)
{
 CORR_VL[[q]] <-cor(VLS[[q]],QQ[[q]])
 # ----- computation of the Communality and Redundancy indexes  ------ #
 COMM_vm[[q]] <- CORR_VL[[q]]^2
 COMM[[q]] <- sum(COMM_vm[[q]])/p_blocchi[[q]]
}
#################################################################
#                     Average Communality                       #
```

```
#################################################################
COMM_M <-0
for (i in 1:nbloc)
{
 if (p_blocchi[i]>1)
 {
  COMM_M<-COMM_M+(p_blocchi[[i]]*COMM[[i]])
 }
}
COMM_M<-COMM_M/sum(  p_blocchi[which(p_blocchi>1)] )
#################################################################
#      computation of the patameters of the inner model       #
#################################################################
n_eso<-0
repeat
{
 n_eso<-n_eso+1
 if (path[n_eso,1]==1) {break}
}
n_eso<-n_eso-1
n_endo<-nbloc-n_eso
print(n_endo)
pred<-vector("list",n_endo)
inn_regr<-vector("list",n_endo)
R2<-array(,n_endo)
RED_blocco<-array(,n_endo)
RED_vm<-vector("list", n_endo)
for (i in 1:n_endo)
{
 pred[[i]]<-matrix(,N,sum(path[n_eso+i,]))
 count<-0
 for (j in 1:ncol(pred[[i]]))
 {
  repeat
  {
   count<-count+1
   if (sum(path[n_eso+i,1:count])==j) {break}
  }
  pred[[i]][,j]<-VLS[[count]]
```

```
  }
  inn_regr[[i]]<-lm(VLS[[n_eso+i]]~pred[[i]])
  R2[i]<-(var(VLS[[n_eso+i]])-(var(residuals(inn_regr[[i]])))/var(VLS[[n_eso+i]])
  RED_blocco[i]<-R2[i]*COMM[[n_eso+i]]
  RED_vm[[i]]<-R2[i]*COMM_vm[[n_eso+i]]
 }
 R2_M<-mean(R2)
 GOF<-sqrt(R2_M*COMM_M)
 list(QQ=QQ, w=w,pred=pred,ncicli=ncicli,VLS=VLS,VL=VL,CORR_VL=CORR_VL,w_tilde=w_tilde,
      w_tilde_normal=w_tilde_normal,COMM=COMM,COMM_M=COMM_M,COMM_vm=COMM_vm,blocchi=blocchi,
      N=N,inn_regr=inn_regr,GOF=GOF,R2=R2,R2_M=R2_M,RED_blocco=RED_blocco,RED_vm=RED_vm)
}
```

Bibliography

Amato, V. [1977], *Statistica*, Cacucci Editore. vol. 1.

Andrich, D. [1978], 'A rating formulation for ordered response categories', *Psychometrika* **43**, 561–573.

Baffi, G., Martin, E. & Morris, A. [1999a], 'Non-linear projection to latent structures revisited: the quadratic PLS algorithm', *Computers and Chemical Engineering* **23**, 395–411.

Ball, R. J. [1963], 'The significance of simultaneous methods of parameter estimation in econometric models', *Applied Statistics* **12**, 14–25.

Barker, M. & Rayens, W. [2003], 'Partial least squares for discrimination', *Journal of chemometrics* **17**, 166–173.

Barvinok, A. [2003], *A course in convexity*, American Mathematical Society.

Bastien, P., Esposito Vinzi, V. & Tenenhaus, M. [2005], 'PLS generalised linear regression', *Computational Statistics & Data Analysis* **48 (1)**, 17–46.

Benzécri, J. [1973], *L'analyse des Données*, Dunod.

Berglund, A., Kettaneh, N., Uppgard, L., Wold, S., Bendwell, N. & Cameron, D. R. [2001], 'The GIFI approach to non-linear PLS modeling', *Journal of Chemometrics* **15**, 321–336.

Berglund, A. & Wold, S. [1997], 'INLR, implicit non-linear latent variable regression', *Journal of Chemometrics* **11**, 141–156.

Betzin, J. & Henseler, J. [2005], Looking at the antecedents of perceived switching costs. a PLS path modeling approach with categorical indicators, *in* T. Aluja, J. Casanovas, V. Esposito Vinzi, A. Morineau & M. Tenenhaus, eds, 'Proceedings of the PLS'05 International Symposium', SPAD.

Bjørnstad, A., Westad, F. & Martens, H. [2004], 'Analysis of genetic marker-phenotype relationships by jack-knifed partial least squares regression (PLSR)', *Hereditas* **141**, 149–165.

Bock, R. [1960], Methods and application of optimal scaling, Technical report, University of North Caroline Psychometric Laboratory Research Memorandum No. 25.

Bollen, K. A. [1989], *Structural equations with latent variables*, Wiley, New York.

Bollen, K. A. & Ting, K. F. [1993], 'Confirmatory Tetrad Analysis', *Sociological Methodology* **23**, 147–175.

Bookstein, F. L. [1982], Soft modeling: The geometrical meaning of soft modeling with some generalization, *in* K. G. Jöreskog & H. Wold, eds, 'Systems under Indirect Observation', Vol. Part II, North-Holland, Amsterdam, pp. 55–74.

Bouroche, J., Saporta, G. & Tenenhaus, M. [1975], Generalized canonical analysis of qualitative data, *in* 'Paper presented at the U.S.-Japan Seminar on theory, Methods and applications of Multidimensional Scaling and Related techniques', University of California, San Diego, CA.

Breiman, L. & Friedman, J. [1985], 'Estimating optimal transformations for multiple regression and correlation', *Journal of the American Statistical Association* **80 (391)**, 580–598.

Burt, C. [1950], 'The factorial analysis of qualitative data', *British Journal of Psychology* **3**, 166–185.

Carroll, J. D. [1968], Generalization of canonical analysis to three or more sets of variables, *in* 'Proceedings of the 76th Convention of the American Psychological Association', Vol. 3, pp. 227–228.

Chin, W. W. [1998], The partial least squares approach for structural equation modeling, *in* G. A. Marcoulides, ed., 'Modern Methods for Business Research', Lawrence Erlbaum Associates, London, pp. 295–236.

Couch, A. & Keniston, K. [1960], 'Yea-sayers and nay-sayers: Agreeing response set as a personality variable', *Journal of Abnormal and Social Psychology* **60**, 151–174.

D'Ambra, L. & Lauro, C. N. [1982], 'Analisi in componenti principali in rapporto a un sottospazio di riferimento', *Rivista di Statistica Applicata* **15**(1), 51–67.

De Jong, S. [1995], 'PLS shrinks', *Journal of Chemometrics* **9 (4)**, 323–326.

de Leeuw, J. & Van Rijckevorsel, J. [1980], Homals & Princals, some generalizations of components analysis, *in* D. E. *et al.*, ed., 'Data Analysis and Informatics', Amsterdam: North-Holland.

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

Durand, J. [2001], 'Local polynomial additive regression through PLS and splines: PLSS', *Chemometrics and Intelligent Laboratory Systems* **58**, 235–246.

Durand, J.-F. & Sabatier, R. [1997], 'Additive splines for partial least squares regression', *Journal of the American Statistical Association* **92 (440)**, 1546–1554.

Efron, B. [1982], *The jackknife, the bootstrap and other resampling plans*, SIAM, USA: Philadelphia.

Escofier, B. & Pagés, J. [1994], 'Multiple factor analysis (AFMULT package)', *Computational Statistics and Data Analysis* **18**, 121–140.

Esposito Vinzi, V. & Russolillo, G. [2010], Partial least squares path modeling and regression, *in* E. Wegman, Y. Said & D. Scott, eds, 'Wiley Interdisciplinary Reviews: Computational Statistics', John Wiley and Sons, p. to appear.

Esposito Vinzi, V., Trinchera, L. & Amato, S. [2009], PLS path modeling: Recent developments and open issues for model assessment and improvement, *in* V. Esposito Vinzi, W. Chin, J. Henseler & H. Wang,

eds, 'Handbook of Partial Least Squares (PLS): Concepts, Methods and Applications', Springer, Berlin, Heidelberg, New York.

Ferrando, P. J. [2008], 'The impact of social desirability bias on the epq-r item scores: An item response theory analysis', *Personality and Individual Differences* **44 (8)**, 1784–1794.

Fisher, R. A. [1940], 'The precision of discriminant functions', *Annals of Eugenics* **10**, 422–429.

Fornell, C. & Bookstein, F. L. [1982], 'Two structural equation models: LISREL and PLS appliead to consumer exit-voice theory', *Journal of Marketing Research* **XIX**, 440–452.

Fornell, C. & Larcker, D. F. [1981], 'Evaluating structural equation models with unobservable variables and measurement error', *Journal of Marketing Reseach* **18**, 39–50.

Fort, G. & Lambert-Lacroix, S. [2005], 'Classification using partial least squares with penalized logistic regression', *Bioinformatics* **7 (21)**, 1104–1111.

Frank, I. [1990], 'A nonlinear PLS model', *Chemometrics and Intelligent Laboratory Systems* **8**, 109–119.

Frank, I. & Friedman, J. [1993], 'A statistical view of some chemometrics regression tools', *Technometrics* **35 (2)**, 109–135.

Frazer, R., Duncan, W. & Collar, A. [1938], *Elementary Matrices and Some Applications to Dynamics and Differential Equations*, Cambridge University Press. 1963 printing.

Garthwaite, P. [1994], 'An interpretation of partial least squares', *Journal of the American Statistical Association* **89 (425)**, 122–127.

Geisser, S. [1974], 'A predictive approach to the random effect model', *Biometrika* **61**, 101–107.

Gifi, A. [1990], *Nonlinear Multivariate Analysis*, Chichester, UK: Wiley.

Glang, M. [1988], Maximierung der Summe erklärter Varianzen in linear-rekursiven Strukturgleichungsmodellen mit multiple Indikatoren: Eine Alternative zum Schätzmodus B des Partial-Least-Squares-Verfahren, Phd thesis, Universität Hamburg, Hamburg, Germany.

Golub, G. & Loan, C. V. [1996], *Matrix Computations (Johns Hopkins Studies in Mathematical Sciences), Third Edition*, The Johns Hopkins University Press.

Greenacre, M. [1984], *Theory and Applications of Correspondence Analysis*, London: Academic Press.

Greenacre, M. [2007], *Correspondence Analysis in practice*, London: Chapman & Hall/CRC.

Greenleaf, E. A. [1992], 'Improving rating scale measures by detecting and correcting bias components in some response styles', *Journal of Marketing Research* **29**, 176–188.

Gudergan, S. P., Ringle, C. M., Wende, S. & Will, A. [2008], 'Confirmatory tetrad analysis in PLS path modeling', *Journal of Business Research* **61 (12)**, 1238–1249.

Guttman, L. [1941], The quantification of a class of attributes: A theory and method of scale costruction, *in* P. H. et al., ed., 'The prediction of personal adjustment', Social Science Research Council, pp. 319–348.

Guttman, L. [1950], The principal components of scale analysis, *in* S. Stouffer, L. Guttman, E. Suchman, P. Lazarsfeld, S. Star & J. Clausen, eds, 'Measurement and Prediction', Wiley (New York), pp. 312–361.

Guttman, L. [1968], 'A general nonmetric technique for finding the smallest coordinate space for configuration of points', *Psychometrika* **33**, 469–506.

Haario, H. & Taavitsainen, V. [1994], 'Nonlinear data analysis ii. examples on new link functions and optimisation aspects', *Chemometrics and Intelligent Laboratory Systems* **23**, 51–64.

Hanafi, M. [2007], 'PLS path modeling: computation of latent variables with the estimation mode B', *Computational Statistics* **22**, 275–292.

Hasegawa, K., Miyashita, Y. & Funatsu, K. [1997], 'GA strategy for variable selection in QSAR studies: GA-based PLS analysis of calcium channel antagonists', *Journal of Chemical Information and Computer Sciences* **37 (2)**, 306–310.

Hayashi, C. [1950], 'On the quantification of qualitative data from the mathematico-statistical point of view', *Annals of the Institute of Statistical Mathematics* **2**, 35–47.

Hayashi, C. [1952], 'On the prediction of phenomena from qualitative data and the quantification of qualitative data from the mathematico-

statistical point of view', *Annals of the Institute of Statistical Mathematics* **3**, 69–98.

Hayashi, C. [1954], 'Multidimensional quantification–with applications to the analysis of social phenomena', *Annals of the Institute of Statistical Mathematics* **5**, 120–143.

Hirschfield, H. [1935], 'A connection between correlation and contingency', *Cambridge Philosophical Society Preceedings* **31**, 520–524.

Hoerl, A. E. & Kennard, R. W. [1970], 'Ridge regression: biased estimation of non-orthogonal components', *Technometrics* **12**, 55–67.

Horst, P. [1935], 'Measuring complex attitude', *Journal of Social Psychology* **6**, 369–374.

Horst, P. [1961], 'Relations among m sets of measures', *Psychometrika* **26**, 129–149.

Höskuldsson, A. [1988], 'PLS regression methods', *Journal of Chemometrics* **2**, 211–228.

Höskuldsson, A. [1992], 'Quadratic PLS regression', *Journal of Chemometrics* **6**, 307–334.

Hotelling, H. [1933], 'Analysis of a complex of statistical variables into components', *Journal of Educational Psychology* **24**.

Hotelling, H. [1936], 'Relations between two sets of variates', *Biometrika* **28**, 321–377.

Israels, A. Z. [1984], 'Redundancy analysis for qualitative variables', *Psychometrika* **49 (3)**, 331–346.

Jakobowicz, E. & Darquenne, C. [2007], 'A modified PLS path modeling algorithm handling reflective categorical variables and a new model building strategy', *Computational Statistics and Data Analysis* **51 (8)**, 3666–3678.

Jolliffe, I. [1982], 'A note on the use of principal components in regression', *Journal of the Royal Statistical Society, Series C (Applied Statistics)* **31 (3)**, 300–303.

Jöreskog, K. [1970], 'A general method for analysis of covariance structure', *Biometrika* **57**, 239–251.

Kelley, H. [1962], Method of gradients, *in* G. L. Ed., ed., 'Optimization techniques with application to aerospace systems', New York: Academic Press, pp. 205–254.

Kettenring, J. R. [1971], 'Canonical analysis of several sets of variables', *Biometrika* **58**, 433–451.

Knowles, E. S. & Nathan, K. [1997], 'Acquiescent responding in self-reports: Social concern or cognitive style?', *Journal of Research in Personality* **31**, 293–301.

Krämer, N. [2007], Analysis of high-dimensional data with partial least squares and boosting, Phd thesis, Technischen Universität Berlin, Berlin, Germany.

Kruskal, J. [1964*a*], 'Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis', *Psychometrika* **29 (1)**, 1–27.

Kruskal, J. [1964*b*], 'Nonmetric multidimensional scaling: A numerical method', *Psychometrika* **29 (2)**, 115–129.

Kruskal, J. & Shepard, R. [1974], 'A nonmetric variety of linear factor analysis', *Psychometrika* **39 (2)**, 123–157.

Kuhfeld, W. F. [1993], Conjoint analysis examples, SAS technical report r-109, Gary, NC: SAS Institute Inc.

Lauro, N. & D'Ambra, L. [1984], L'analyse non symmétrique des correspondances, *in* E. Diday & al., eds, 'Data Analysis and Informatics, III', North-Holland.

Likert, R. [1932], 'A technique for the measurement of attitudes', *Archives of Psychology* **22 (140)**, 1–55.

Lohmöller, J. [1987], LVPLS program manual, version 1.8, Technical report, Zentralarchiv für Empirische Sozialforschung, Universität Zu Köln, Köln.

Lohmöller, J. [1989], *Latent variable path modeling with partial least squares*, Physica-Verlag, Heildelberg.

Lovaglio, P. G. [2001], Modelli con variabili latenti ed indicatori di tipo misto, Phd thesis, Università di Milano Bicocca, Milano, Italy.

Lovaglio, P. G. [2002], 'La stima di Variabili Latenti da variabili osservate miste', *Statistica* **LXII (2)**, 203–213.

Lyttkens, E., Areskoug, B. & Wold, H. [1975], The convergence of NIPALS estimation procedures for six path models with one or two latent variables, Technical report, University of Göteborg.

Martens, H. & Naes, T. [1989], *Multivariate calibration*, Wiley, Chichester, U.K.

Mathes, H. [1993], Global optimisation criteria of the pls-algorithm in recursive path models with latent variables, *in* K. Haagen, D. Bartholomew & M. Deister, eds, 'Statistical modelling and latent variables', Elsevier Science: Amsterdam.

Michalidis, G. & de Leeuw, J. [1998], 'The Gifi system for descriptive multivariate analysis', *Statistical Science* **13 (4)**, 307–336.

Murtagh, F. [2005], *Correspondence Analysis and Data Coding with Java and R*, London: Chapman & Hall/CRC.

Nishisato, S. [1980], *Analysis of categorical data: Dual Scaling and its applications*, Toronto: University of Toronto Press.

Nishisato, S. [2007], *Multivariate Nonlinear Descriptive Analysis*, Chichester, UK: Wiley.

Nishisato, S. & Arri, P. [1975], 'Nonlinear programming approach to optimal scaling of partially ordered categories', *Psychometrika* **40**, 525–548.

Qin, S. & McAvoy, T. [1992], 'Non linear PLS modelling using neural networks', *Computers and Chemical Engineering* **16**, 379–391.

Richardson, M. & Kuder, G. F. [1933], 'Making a rating scale that measures', *Personnel Journal* **12**, 36–40.

Rosipal, R. & Krämer, N. [2006], Overview and recent advances in partial least squares, *in* C. Saunders, M. Grobelnik, S. Gunn & J. Shawe-Taylor, eds, 'Subspace, Latent Structure and Feature Selection Techniques', Springer, pp. 34–51.

Rosipal, R. & Trejo, L. J. [2001], 'Kernel partial least squares regression in reproducing kernel hilbert space', *Journal of Machine Learning Research* **2**, 97–123.

Rothman, L. & Parker, M. J. [2009], *Just-About-Right (JAR) Scales: Design, Usage, Benefits and Risks*, ASTM International.

Russet, B. M. [1964], 'Inequality and instability', *Word politics* **21**, 422–454.

Russolillo, G. & Lauro, N. C. [2010], A proposal for handling categorical predictors in PLS regression framework, *in* B. Fichet, D. Piccolo, R. Verde & M. Vichi, eds, 'Post Conference Proceedings of the First joint meeting of the Société Francophone de Classification and the Classification and Data Analysis Group of the Italian Society of Statistics', Springer.

Russolillo, G. & Trinchera, L. [2009a], PLS algorithms for non-metric data, *in* V. Esposito Vinzi, M. Tenenhaus & R. Guan, eds, 'Proceedings of the 6th International Conference on Partial Least Squares and Related Methods', PHEI, pp. 60–65.

Russolillo, G., Trinchera, L. & Esposito Vinzi, V. [2009b], A non linear regularized component based approach to structural equation modeling, *in* 'Book of short papers - Convegno Intermedio SIS-2009 Statistical Methods for the analysis of large data-sets', CLEUP.

Sampson, P., Streissguth, A., Barr, H. & Bookstein, F. [1989], 'Neurobehavioral effects of prenatal alcohol: Part ii. partial least squares analysis', *Neurotoxicology and tetralogy* **11**, 477–491.

Sánchez, G. [2009], PATHMOX Approach: Segmentation Trees in Partial Least Squares Path Modeling, Phd thesis, Universitat Politècnica de Catalunya, Barcelona, Spain.

Saporta, G. [1980], About some remarkable properties of generalized canonical analysis, *in* 'Paper presented at the second European meeting of the Psychometric Society', Groningen, The Netherlands.

Schulbach, K. F., Rouseff, R. L. & Sims, C. A. [2006], 'Relating descriptive sensory analysis to gas chromatography/olfactometry ratings of fresh strawberries using partial least squares regression', *Journal of Food Science* **69 (7)**, 273–277.

Shepard, R. [1966], 'Metric structures in ordinal data', *Journal of Mathematical Psychology* **3**, 287–315.

Stevens, S. S. [1946], 'On the theory of scales of measurement', *Science* **103 (2684)**, 677–680.

Stone, M. [1974], 'Cross-validatory choice and assessment of statistical predictions', *Journal of the Royal Statistical Society* **36**, 111–147.

Taavitsainen, V. & Korhonen, P. [1992], 'Nonlinear data analysis with latent variables', *Chemometrics and Intelligent Laboratory Systems* **14**, 185–194.

Takane, Y., Young, F. & de Leeuw, J. [1977], 'Nonmetric individual differences multidimensional scaling: an alternating least square method with optimal scaling features', *Psychometrika* **42**, 7–67.

Takane, Y., Young, F. & de Leeuw, J. [1979], 'Nonmetric common factor analysis: An alternating least squares method with optimal scaling features', *Behaviormetrika* **6**, 45–56.

Tanaka, Y. [1979], 'Review of the methods of quantification', *Environmental Health Perspectives* **32**, 113–123.

Tenenhaus, A., Giron, A., Viennet, E., Bera, M., Saporta, G. & Fertil, B. [2007], 'Kernel logistic PLS: A tool for supervised nonlinear dimensionality reduction and binary classification', *Computational Statistics & Data Analysis* **51 (9)**, 4083–4100.

Tenenhaus, M. [1977], 'Analyse en composantes principales d'un ensemble de variables nominales ou numéric', *Revue de statitique appliquée* **25**, 39–56.

Tenenhaus, M. [1998], *La Régression PLS: théorie et pratique*, Technip, Paris.

Tenenhaus, M. [2009], A criterion based PLS approach to structural equation modelling, *in* 'Programme and Abstract of the 6th International Conference n Partial Least Squares Methods', p. 3.

Tenenhaus, M., Amato, S. & Esposito Vinzi, V. [2004], A global goodness-of-fit index for PLS structural equation modelling, *in* 'Proceedings of the XLII SIS Scientific Meeting', Vol. Contributed Papers, CLEUP, Padova, pp. 739–742.

Tenenhaus, M., Esposito Vinzi, V., Chatelin, Y. & Lauro, C. [2005], 'PLS path modeling', *Computational Statistics and Data Analysis* **48**, 159–205.

Tenenhaus, M. & Hanafi, M. [2009], A bridge between PLS path modeling and multi-block data analysis, *in* K. Haagen, D. Bartholomew & M. Deister, eds, 'Handbook of Partial Least Squares (PLS): Concepts, Methods and Applications', Elsevier Science: Amsterdam.

Tenenhaus, M. & Young, F. [1985], 'An analysis of synthesis of multiple coreespondence analysis, optimal scaling, dual scaling, homogeneity analysis and other methods for quantifying categorical multivariate data', *Psychometrika* **50 (1)**, 91–119.

Thurstone, L. [1947], *Multiple factors analysis*, Chicago, ILL: The University of Chicago Press.

Thurstone, L. L. [1931], *The theory of multiple factors*, Edwards Brothers, Ann Arbor, MI.

Tishler, A., Dvir, D., Shenhar, A. & Lipovetsky, S. [1996], 'Identifying critical success factors in defense development projects: a multivariate analysis', *Technological forecasting and social change* **51**, 151–171.

Tucker, L. [1958], 'An inter-battery method of factor analysis', *Journal of Chemometrics* **23**, 111–136.

Van de Wollemberg, A. L. [1977], 'Redundancy analysis. an alternative for canonical correlation analysis', *Psychometrika* **42**, 207–219.

Velleman, P. F. & Wilkinson, L. [1993], 'Nominal, ordinal, interval, and ratio typologies are misleading', *The American Statistician* **47 (1)**, 65–72.

Vivien, M. [2002], Approaches PLS linéaires et non-linéaires pour la modélisation de multi-tableaux: théorie et applications, Phd thesis, Université de Montpellier I, Montpellier, France.

Wegelin, J. [2000], A survey on partial least squares (PLS) methods, with emphasis on the two-block case, Technical report n.371, Departement of Statistics, University of Washington, USA: Seattle.

Wold, H. [1966a], *Non linear Estimation by Iterative Least Squares procedure*, Research paper in Statistics: Festschift for J. Neyman, F. David.

Wold, H. [1966b], Estimation of principal component and related models by iterative least squares, *in* P. R. Krishnaiah, ed., 'Multivariate Analysis', Academic Press, New York, pp. 391–420.

Wold, H. [1975a], Path models with latent variables: The non-linear iterative partial least squares (NIPALS) approach, *in* H. M. Blalock, A. Aganbegian, F. M. Borodkin, R. Boudon & V. Capecchi, eds, 'Quantitative Sociology: Intentional Perspective on Mathematical and Statistical Modeling', Accademic Press, New York, pp. 307–357.

Wold, H. [1975b], Soft modeling by latent variables: The nonlinear iterative partial least squares approach, *in* 'Gani, J. (Ed.), Perspectives in probability and statistics, papers in honor of M.S. Bartlett', Academic Press, London, pp. 117–142.

Wold, H. [1982], Soft modeling: the basic design and some extensions, *in* K. G. Jöreskog & H. Wold, eds, 'Systems under Indirect Observation', Vol. Part II, North-Holland, Amsterdam, pp. 1–54.

Wold, S. [1992], 'Nonlinear partial least squares modelling ii. spline inner relation', *Chemometrics and Intelligent Laboratory Systems* **14**, 71–84.

Wold, S. [2009], PLS for multivariate linear modeling, *in* H. van de Waterbeemd, ed., 'QSAR: Chemometric Methods in Molecular Design. Methods and Principles in Medicinal Chemistry', Weinheim, Germany: Verlag-Chemie.

Wold, S., Kettaneh-Wold, N. & Skagerberg, B. [1989], 'Nonlinear PLS modeling.', *Chemometrics and Intelligent Laboratory Systems* **7**, 53–65.

Wold, S., Martens, H. & Wold, H. [1983], The multivariate calibration problem in chemistry solved by the PLS method, *in* A. Ruhe & B. Kagstrom, eds, 'Proceedings of the Conference on Matrix Pencils. Lectures Notes in Mathematics', Springer, Heidelberg.

Wold, S., Ruhe, A., Wold, H. & Dunn, W. [1984], 'The collinearity problem in linear regression. the PLS approach to generalised inverses', *Journal of Scientific Statistical Computing, SIAM* **5**, 735–743.

Wold, S., Sjöström, M. & Eriksson, L. [2001], 'PLS regression: a basic tool of chemometrics', *Chemometrics and Intelligent Laboratory Systems* **58**, 109–130.

Xiong, R. & Meullenet, J.-F. [2008], 'A PLS dummy variable approach to assess the impact of jar attributes on liking', *Food Quality and Preference* **17 (3-4)**, 188–198.

Young, F. [1975], 'Methods for describing ordinal data with cardinal models', *Journal of Mathematical Psychology* **12**, 416–436.

Young, F. [1981], 'Quantitative analysis of qualitative data', *Psychometrika* **44 (4)**, 357–388.

Young, F., de Leeuw, J. & Takane, Y. [1976], 'Regression with qualitative and quantitative variables: An alternating least squares method with optimal scaling features', *Psychometrika* **41**, 505–529.

Young, F., Takane, Y. & de Leeuw, J. [1978], 'The principal components of mixed measurement level multivariate data: An alternating least squares method with optimal scaling features', *Psychometrika* **43 (2)**, 279–281.