Electronic Journal of Applied Statistical Analysis
EJASA, Electron. J. App. Stat. Anal.
`http://siba-ese.unisalento.it/index.php/ejasa/index`
e-ISSN: 2070-5948
DOI: 10.1285/i20705948v18n1p1

Finite Iterative Method Based Algorithm To Estimate Latent Variables For Reflective Blocks In Partial Least Squares Structural Equation Modeling

By EL Hadri, Ebnou Abdem

15 March 2025

Electronic Journal of Applied Statistical Analysis
Vol. 18, Issue 01, march 2025, 1-25
DOI: 10.1285/i20705948v18n1p1

Finite Iterative Method Based Algorithm To Estimate Latent Variables For Reflective Blocks In Partial Least Squares Structural Equation Modeling

Zouhair EL Hadri[a] and Seyid Abdellahi Ebnou Abdem*[a]

[a] *Laboratory of Mathematics, Statistics, and Applications, Faculty of Sciences, Mohammed V University in Rabat. Morocco*

15 March 2025

Partial Least Squares Structural Equation Modeling (PLS-SEM) is a powerful statistical approach that has become a mainstream method in many application areas. It offers flexibility in handling formative and reflective measurement blocks, enabling researchers to model relationships among observed and latent variables. The crucial step in this approach is the PLS-SEM algorithm, which involves computing the scores of latent variables by alternating between inner and outer estimation. The aims of the present paper are twofold. The first contribution shows that the computations used in the outer estimation are inappropriate for reflective blocks. The second contribution involves introducing an alternative algorithm to overcome this drawback by using a new strategy based on considering the true structure of reflective blocks. Numerical studies and empirical simulations are provided to illustrate the advantages of the proposed algorithm compared to the classical one.

keywords: Partial Least Squares Structural Equation Modeling, Reflective Path Analysis Model, Finite Iterative Method, Unweighted Least Squares.

*Corresponding authors: seyidebnou@gmail.com

ISSN: 2070-5948
http://siba-ese.unisalento.it/index.php/ejasa/index

Introduction

Structural Equation Modeling (SEM) is a set of statistical techniques commonly used in order to model complex systems of human behavior while allowing for the use of latent variables and variables measured with error. SEM helps the researcher to test or validate a theoretical model for theory testing and extension.

Two major approaches are most widely used to estimate SEM: Covariance-based (CB-SEM) and Variance-based approaches Bollen (1989). The first approach called Linear Structural RELations (LISREL) Bollen (1989) minimizes the discrepancy between the empirical and model-implied variance-covariance matrix of the observable variables to obtain the model parameter estimates. The second approach, called Partial Least Square Structural Equation Modeling (PLS-SEM) Benitez et al. (2020) creates linear combinations of the observed variables as stand-ins for the theoretical concepts and subsequently estimates the model parameters Russolillo (2009); Benitez et al. (2020).

Originally, PLS-SEM was founded in the first half of the 20^{th} century by Herman Wold Takane and Hwang (2018); Vinzi (2010); Hair Jr et al. (2021); Hwang et al. (2020); Streukens et al. (2017); Hanafi et al. (2022); Wold (1966). Currently, PLS-SEM is a prevailing approach that is widely used in many fields such as Business administration research Hair et al. (2012a), marketing Hair et al. (2012b), operations management Peng and Lai (2012), finance Avkiran and Ringle (2018), economics Sanchez (2013), and others Fabbricatore et al. (2023); Mayrink et al. (2021); Hwang et al. (2015); Cho and Choi (2020); Takane and Hwang (2018); Vinzi (2010); Hair Jr et al. (2021); Hwang et al. (2020); Streukens et al. (2017). PLS-SEM is implemented in many software, such as LVPLS Lohmöller (2013), PLS-Graph (Chin 2003) Chin (2001), SmartPLS Ringle et al. (2015), PLS-SEM in XLSTAT XLSTAT (2019) and `plspm` package in R Sanchez (2013).

The analysis in PLS-SEM framework begins by drawing the conceptual representation of the whole model. An example is given in figure 1.

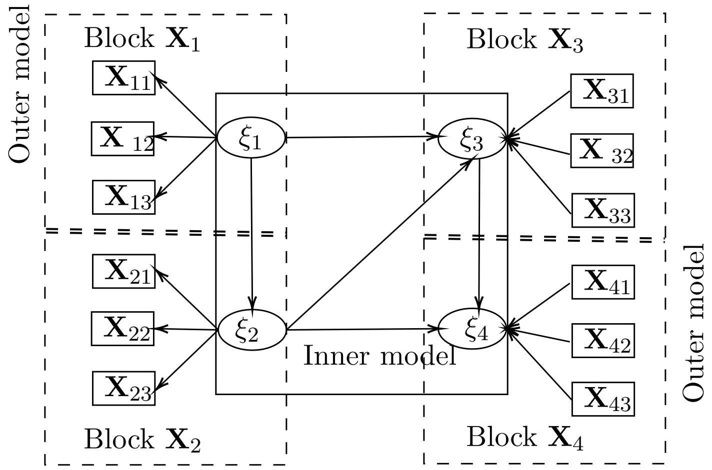

Figure 1: Schematic representation of a PLS-SEM model with four blocks.

In the sequel, we note respectively by n, K, and p_k, the number of observations, the number of blocks, and the number of observed variables called Manifest Variables (MVs) in the k^{th} block with $1 \leq k \leq K$.

The whole model contains two sub-models; (i) the *structural model* (or the inner model), (ii) and the *measurement model* (or the outer model). On one side, the structural model describes the relations between unobserved variables called Latent Valriables (LVs) denoted by ξ_k, $(1 \leq k \leq K)$ (see figure 2 (a)). These relations are assumed to follow linear regression of the form:

$$\xi_k = \sum_{j=1, j\neq k}^{K} \beta_{kj}\xi_j + \zeta_k \tag{1}$$

where ζ_k called *disturbance* is a centered random variable not correlated with all LVs appearing on the right side of (1). And β_{kj} called *path coefficient* is a quantity that measures the effect of the LV ξ_j on the LV ξ_k. Note that some path coefficients β_{kj} are structurally null and correspond to LVs ξ_j that are not connected to ξ_k.

On the other side, the measurement models describe the relations between each LV and its corresponding MVs. Two kinds of measurement models are widely used in practical studies, models with reflective blocks and models with formative blocks.

If the k^{th} block is reflective, the LV is considered as the cause of its own MVs, see figure 2 (b). In other words, for each j, $(1 \leq j \leq p_k)$, ξ_k gives rise to each $\mathbf{X}_{kj}$ as follows:

$$\mathbf{X}_{kj} = \lambda_{kj}\xi_k + \epsilon_{kj} \tag{2}$$

where ϵ_{kj} is a centred random variable not correlated with the LV ξ_k.

Contrariwise, if the k^{th} block is formative, the LV is caused by its MVs, see figure 2 (c) i.e. $\mathbf{X}_{k1}, \cdots, \mathbf{X}_{kp_k}$ gives rise to ξ_k as follows:

$$\xi_k = \sum_{j=1}^{p_k} \lambda_{kj}\mathbf{X}_{kj} + \epsilon_k \tag{3}$$

where ϵ_k is a centred random variable not correlated with all MVs $\mathbf{X}_{k1}, \cdots, \mathbf{X}_{kp_k}$.

Parameters λ_{kj}, $(1 \leq k \leq K)$ and $(1 \leq j \leq p_k)$ figuring in (2) and (3) are called *loadings*.

As aforementioned, all LVs in PLS-SEM are standardized. In the present paper, we assume that all MVs are also standardized. That is:

$$\parallel \xi_k \parallel = \parallel \mathbf{X}_{kj} \parallel = \sqrt{n} \quad \forall\, 1 \leq k \leq K \quad \text{and} \quad \forall\, 1 \leq j \leq p_k \tag{4}$$

The main objective of the analysis is to estimate the parameters of the whole model, which are the path coefficients β_{kj} and the loadings λ_{kj}. PLS-SEM consists of two main stages. The first stage computes the scores of each LV as a weighted composite of its MVs while the second stage uses these scores to estimate β_{kj} and λ_{kj}. The present paper focuses on the first stage.

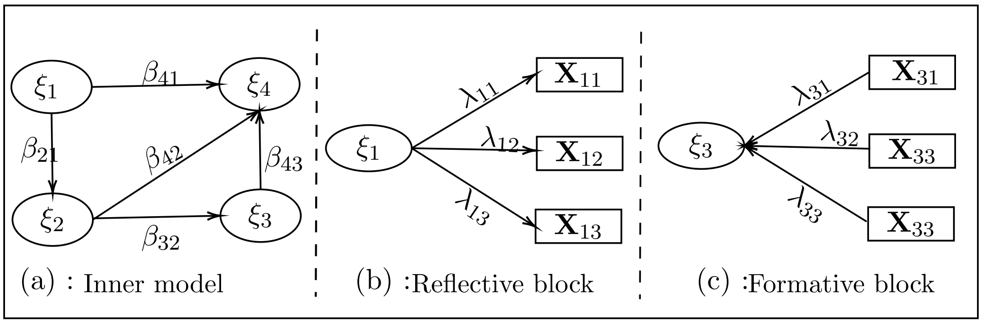

Figure 2: Structural model, formative block, and reflective block.

In order to assess the quality of the obtained scores, the communality index is widely used. It is defined as follows:

$$Com = \frac{1}{K}\sum_{k=1}^{K}\frac{1}{p_k}\sum_{j=1}^{p_k} Cor^2(\xi_k, \mathbf{X}_{kj}) \tag{5}$$

This index is obtained as the mean of the squared correlations linking each MV to the corresponding LV. It is the average proportion of variance of MVs reproduced by the LVs Tenenhaus et al. (2005). Values closer to 1 suggest that each LV is more related to its corresponding MVs.

Formally, the scores of LVs are computed as weighted sums of the MVs by alternating between two types of estimations. The first is called *internal estimation* and concerns the structural model. The internal score of ξ_k is computed as a linear combination of all scores of LVs connected to ξ_k. The user specifies the scheme and the procedure to compute the coefficients of this combination. Two schemes are generally used in practice, which are the *centroid scheme* and the *factorial scheme*. The factorial scheme considers the degree of connection between blocks by the correlation coefficient between the considered blocks. Meanwhile, in the centroid scheme, the blocks are treated fairly. Furthermore, two procedures are used to design the scores to be integrated into the computation of a new score. The first, called *Lohmöller's procedure* uses only scores computed at the previous iteration. The second, called *Hanafi-Wold's procedure* uses scores computed at the previous and actual iterations.

The second estimation, called *external estimation* concerns the measurement model. Basically, the external score of ξ_k is computed as a linear combination of all columns of the matrix $\mathbf{X}_k$ by specifying the mode to be considered to compute the *vector of weights* of this combination. Two modes are widely used, which are Mode A and Mode B. The choice of the mode is subject to theoretical reasoning, and in most cases, Mode A is more commonly used Sanchez (2013); Henseler (2010); Pelagatti et al. (2012); Eboli et al. (2018); Crocetta et al. (2021).

In Mode A, the corresponding block is reflective and it is considered as p_k simple regression models. As a consequence, the vector of weights is computed as being the vector of regression coefficients in the simple regressions of each MV on the associated

LV. Besides, in Mode B, the associated block is formative and it is considered as a multiple regression model. As a matter of fact, the vector of weights is computed as being the vector of regression coefficients in the multiple regression of the LV on its own MVs.

As will be illustrated in section 1.3, the computation used in Mode B is completely justified. However, the calculation used in Mode A is not consistent since it does not reflect the true structure of the block. An immediate consequence is that the obtained scores may lead to communality index which is not optimal.

The present paper aims to overcome this drawback by providing a new strategy to compute the vector of weights. The main idea is to consider the true structure of reflective block which is a Path Analysis Model (PAM). The direct consequence is that the obtained scores of LVs are improved. For instance, if we consider the model described in Hanafi et al. (2022) and the corresponding dataset 5, the communality index associated with the scores obtained by using simple regressions in Mode A is 0,7604 whereas this same index is 0,8472 for the scores obtained by the proposed computations.

The paper is structured as follows. Section 1 presents PLS-SEM algorithm. It recalls the outer and inner estimation of scores of LVs. In section 2, we first present briefly recursive Path Analysis models and how the parameters are estimated. Then we focus on reflective and formative models. Section 3 introduces a new algorithm to estimate the scores of latent variables. We begin by defining the augmented blocks then we provide a new strategy to compute the vectors of weights. In section 4, the introduced algorithm is numerically compared to the classical algorithm. Finally, section 5 concludes with a summary and perspectives.

1 PLSPM algorithm

The present section presents the external and internal estimation of scores of LVs, and the PLSPM algorithm. Let $1 \leq k, l \leq K$ and $s = 0, 1, \cdots$.

1.1 External estimation

Let $\mathbf{z}_k^{(s)}$ be the internal estimation of the score associated wih the k^{th} LV, at iteration s. The corresponding external estimation at iteration $s+1$ denoted by $\mathbf{y}_k^{(s+1)}$ is defined by :

$$\mathbf{y}_k^{(s+1)} = \mathbf{X}_k \mathbf{w}_k^{(s+1)} \tag{6}$$

where

$$\mathbf{w}_k^{(s+1)} = \sqrt{n} \frac{\tilde{\mathbf{w}}_k^{(s+1)}}{\left\| \mathbf{X}_k \tilde{\mathbf{w}}_k^{(s+1)} \right\|} \tag{7}$$

so that $\mathbf{y}_k^{(s+1)}$ have unit variance (see (4)) is the vector of weights and $\tilde{\mathbf{w}}_k^{(s+1)}$ is computed according to the specified mode as follows :

$$\tilde{\mathbf{w}}_k^{(s+1)} = \begin{cases} \mathbf{X}_k'\mathbf{z}_k^{(s)} & \textit{Mode A} \\ (\mathbf{X}_k'\mathbf{X}_k)^{-1}\mathbf{X}_k'\mathbf{z}_k^{(s)} & \textit{Mode B} \end{cases} \tag{8}$$

Recall that Mode A (respectively Mode B) is preferred for reflective blocks (respectively for formative blocks).

1.2 Internal estimation

The internal estimation of the score associated with the k^{th} LV at iteration $(s+1)$ is computed as follows :

$$\mathbf{z}_k^{(s+1)} = \begin{cases} \sum_{l=1}^{K} c_{kl}\theta_{kl}^{(s)}\mathbf{y}_l^{(s)} & \textit{Lohmöller's procedure} \\ \sum_{l=1}^{k-1} c_{kl}\theta_{kl}^{(s)}\mathbf{y}_l^{(s+1)} + \sum_{l=k+1}^{K} c_{kl}\theta_{kl}^{(s)}\mathbf{y}_l^{(s)} & \textit{Wold's procedure} \end{cases} \tag{9}$$

where

$$\theta_{kl}^{(s)} = \begin{cases} \text{sign}\left[cor\left(\mathbf{y}_k^{(s)}, \mathbf{y}_l^{(s)}\right)\right] & \textit{Centroid scheme} \\ cor\left(\mathbf{y}_k^{(s)}, \mathbf{y}_l^{(s)}\right) & \textit{Factorial scheme} \end{cases} \tag{10}$$

with $c_{kl} = 1$ if the LVs ξ_k and ξ_l are linked, and $c_{kl} = 0$ otherwise. $\mathbf{y}_k^{(s)}$ and $\mathbf{y}_l^{(s)}$ are the external estimations of the scores associated with ξ_k and ξ_l respectively as defined in (6). Lohmöller's procedure computes the score $\mathbf{z}_k^{(s+1)}$ at iteration $s+1$ as a function of all adjacent scores $\mathbf{z}_j^{(s)}$, $(1 \leq j \leq K$ *and* ξ_j *connected to* $\xi_k)$ obtained in the previous iteration (s) Hanafi (2007); Tenenhaus et al. (2005); Henseler (2010) while Wold's procedure uses the most recent available information, i.e. $\mathbf{z}_j^{(s+1)}$, $(1 \leq j \leq k-1)$ and $\mathbf{z}_j^{(s)}$, $(k+1 \leq j \leq K)$ Henseler (2010); Hanafi et al. (2021). Wold's procedure is characterized by the advantage of being monotonically convergent. In contrast, Lohmöller's procedure does not always converge monotonically, but is implemented in most PLS software. The present paper focuses on Lohmöller's procedure.

1.3 PLSPM algorithm

The alternation between external and internal estimation defines the so called PLSPM algorithm which can be summarized in figure 3 below.

PLSPM algorithm is initialized by arbitrary vectors of weights $\mathbf{w}_1^{(0)}, \cdots, \mathbf{w}_K^{(0)}$ and iterated over s; $(s = 1, 2, \ldots)$ until the quantity $\sum_{k=1}^{K} \left\|\mathbf{y}_k^{(s+1)} - \mathbf{y}_k^{(s)}\right\|^2$ is less than a fixed small threshold.

Equations in (8) are the main motivation of the present paper. In deed, when mode B is considered for the k^{th} block, this latter is a multiple regression model and the second

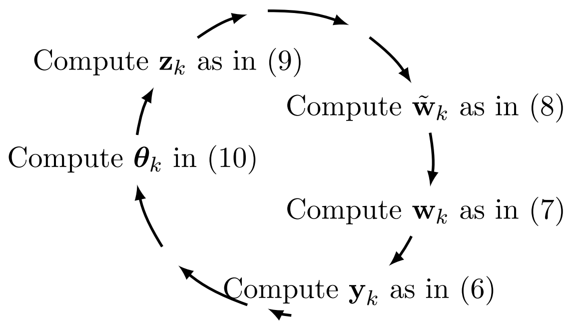

Figure 3: PLSPM algorithm to compute the scores of LVs.

equation in (8) is coherent with this structure: $\tilde{\mathbf{w}}_k^{(s+1)}$ is estimated by the vector of the coefficients of the multiple regression of $\mathbf{z}_k^{(s)}$ on $\mathbf{X}_{k1}, \cdots$ and $\mathbf{X}_{kp_k}$.

Contrariwise, when mode A is chosen for the k^{th} block, this latter is not a set of simple regression models because it does not take into account the correlations between all MVs of the block $\mathbf{X}_k$. Consequently, it $\tilde{w}_{kj}^{(s+1)}$ cannot be estimated by the regression coefficient on the simple regression of $\mathbf{X}_{kj}$ on $\mathbf{z}_k^{(s)}$. i.e. the first equation in (8) is not realistic.

The main contribution of the present paper is to introduce an alternative method to estimate the weight vectors $\tilde{\mathbf{w}}_k^{(s+1)}$ given in (8) when Mode A is used. Basically, the k^{th} block contains one explanatory variable ($\mathbf{z}_k^{(s)}$) and p_k explained variables ($\mathbf{X}_{k1}, \cdots$ and $\mathbf{X}_{kp_k}$). As a consequence, this block is a reflective path analysis model. Consequently, the associated parameters should be estimated by the appropriate strategy. In this regard, El Hadri & al El Hadri et al. (2023); Sahli et al. (2024) recently proposed a new method to estimate such type of models. It consists of an Alternating Least Square procedure based on the Finite Iterative Method El Hadri and Hanafi (2015).

2 Estimation of Reflective Path Analysis Model

2.1 Recursive Path Analysis Model

Definition 1. *A model regrouping a set of observed variables for which at least one of them is an explained variable is called Path Analysis Model (PAM)Kline (2023).*

In this section, the following notations are adopted.

- A variable that is always explanatory is called an *exogenous* variable and denoted by ξ. Otherwise, it is called an *endogenous* variable and denoted by η.
- The direct effect of an exogenous variable on an endogenous variable is denoted by γ.

- The direct effect of an endogenous variable on another endogenous variable is denoted by β.
- The disturbance term associated with an endogenous variable is denoted by ζ.

A PAM is represented algebraically by a system of a set of multiple regressions as follows:

$$\begin{cases} \eta_1 = \gamma_{11}\xi_1 + \ldots + \gamma_{1q}\xi_q + \zeta_1 \\ \vdots \\ \eta_p = \beta_{p1}\eta_1 + \ldots + \beta_{p,p-1}\eta_{p-1} + \gamma_{p1}\xi_1 + \ldots + \gamma_{pq}\xi_q + \zeta_p \end{cases} \tag{11}$$

where q and p are respectively the number of exogenous and endogenous variables. System (11) can be formulated as the following compact form:

$$\boldsymbol{\eta} = \boldsymbol{\Gamma\xi} + \mathbf{B}\boldsymbol{\eta} + \boldsymbol{\zeta} \tag{12}$$

Here, $\boldsymbol{\eta}$ and $\boldsymbol{\xi}$ are respectively the vectors of the endogenous and exogenous variables, and ζ is the vector of disturbances. In addition, $\mathbf{B} = [\beta_{ij}]_{1\leq i,j\leq p}$ denotes the $(p \times p)$ matrix of parameters relating endogenous variables. And $\boldsymbol{\Gamma} = [\gamma_{ij}]_{1\leq i\leq p; 1\leq j\leq q}$ is the $(p \times q)$ matrix of parameters relating endogenous variables to exogenous variables.

Definition 2. *A Path Analysis Model is called a Recursive Path Analysis Model if* $\mathbf{B}$ *is a lower triangular matrix.*

2.2 Estimation of Recursive Path Analysis Model

The estimation of a Recursive PAM consists of finding values for the unknown parameters by minimizing a given criterion. These parameters are the elements of the vector $\boldsymbol{\rho}$ defined as follows:

Definition 3. *The vector denoted by* $\boldsymbol{\rho}$ *whose elements are the non null elements of matrices* $\boldsymbol{\Phi}$*,* $\boldsymbol{\Gamma}$ *and* $\mathbf{B}$ *is called the vector of parameters:*

$$\boldsymbol{\rho} = Vecs(\boldsymbol{\Phi}, \boldsymbol{\Gamma}, \mathbf{B}) \tag{13}$$

where $\boldsymbol{\Phi} = [\phi_{ij}]_{1\leq i,j\leq q}$ denotes the $(q \times q)$ matrix among exogenous variables. Meanwhile, the Unweighted Least Squares is widely used as a criterion to be minimized. It is defined as Bollen (1989); Kline (2023); El Hadri and Hanafi (2015):

$$F(\boldsymbol{\rho}) = \frac{1}{2}tr\left[\left(\widehat{\mathbf{R}}(\boldsymbol{\rho}) - \mathbf{R}\right)^2\right] = \frac{1}{2} \parallel \widehat{\mathbf{R}}(\boldsymbol{\rho}) - \mathbf{R} \parallel_F^2 \tag{14}$$

where in (14):

(i) $\mathbf{R}$ is the empirical correlation matrix (see (15)),

(ii) $\widehat{\mathbf{R}} = \widehat{\mathbf{R}}(\boldsymbol{\rho})$ is the correlation matrix involved by the path model (see (16)),

(iii) tr denotes the trace of a square matrix and $\|\|_F$ denotes the Frobenius norm.

Definition 4.

i. The matrix denoted by $\mathbf{R}$ *and defined by*

$$\mathbf{R} = \begin{pmatrix} \mathbb{E}(\boldsymbol{\xi}^{(data)}\boldsymbol{\xi}^{(data)'}) & \mathbb{E}(\boldsymbol{\xi}^{(data)}\boldsymbol{\eta}^{(data)'}) \\ \mathbb{E}(\boldsymbol{\eta}^{(data)}\boldsymbol{\xi}^{(data)'}) & \mathbb{E}(\boldsymbol{\eta}^{(data)}\boldsymbol{\eta}^{(data)'}) \end{pmatrix} \tag{15}$$

is called the empirical correlation matrix, where $\boldsymbol{\xi}^{(data)}$ *and* $\boldsymbol{\eta}^{(data)}$ *are respectively the vectors of exogenous and the endogenous variables obtained from data.*

ii. The matrix denoted by $\widehat{\mathbf{R}}$ *and defined by*

$$\widehat{\mathbf{R}} = \widehat{\mathbf{R}}(\boldsymbol{\rho}) = \begin{pmatrix} \mathbb{E}(\boldsymbol{\xi}^{(model)}\boldsymbol{\xi}^{(model)'}) & \mathbb{E}(\boldsymbol{\xi}^{(model)}\boldsymbol{\eta}^{(model)'}) \\ \mathbb{E}(\boldsymbol{\eta}^{(model)}\boldsymbol{\xi}^{(model)'}) & \mathbb{E}(\boldsymbol{\eta}^{(model)}\boldsymbol{\eta}^{(model)'}) \end{pmatrix} \tag{16}$$

is called the correlation matrix implied by the model, where $\boldsymbol{\xi}^{(model)}$ *and* $\boldsymbol{\eta}^{(model)}$ *are respectively the vectors of exogenous and the endogenous variables obtained from the vector of the model's parameters defined in (13).*

It is clear that the minimization of the criterion F defined in (14) requires to answer the following two questions. First, how the matrix $\widehat{\mathbf{R}}$ is computed when $\boldsymbol{\rho}$ is known? And second, how to find the vector $\boldsymbol{\rho}$ which minimizes F?

In order to answer the first question, El Hadri and Hanafi (2015) introduced the Finite Iterative Method (FIM) in 2015 to compute $\widehat{\mathbf{R}}$ iteratively. FIM is defined by the following iterations:

Algorithm 1 Finite Iterative Method to compute the matrix $\widehat{\mathbf{R}}$ for q exogenous variables and p endogenous variables.

Initialization : Set $\widehat{\mathbf{R}}_{1:q,1:q} = \boldsymbol{\Phi}$, $\mathbf{A} = [\boldsymbol{\Gamma}, \mathbf{B}]$ and $j = 1$.

Iterate over $j = 1, \cdots, p$ the following steps :

Step 1. Compute $\widehat{\mathbf{R}}_{q+j,1:q+j-1} = \mathbf{A}_{j,1:q+j-1}\widehat{\mathbf{R}}_{1:q+j-1,1:q+j-1}$.

Step 2. Compute $\widehat{\mathbf{R}}_{1:q+j-1,q+j} = (\widehat{\mathbf{R}}_{q+j,1:q+j-1})'$.

Step 3. Set $\widehat{\mathbf{R}}_{q+j,q+j} = 1$.

FIM starts by setting $\widehat{\mathbf{R}}_{1:q,1:q} = \boldsymbol{\Phi}$ (correlation matrix between the q exogenous variables) where $\widehat{\mathbf{R}}_{1:q,1:q}$ is the sub-matrix of $\widehat{\mathbf{R}}$ obtained by extracting the first q rows and the first q columns. Thereafter, the sub-row $\widehat{\mathbf{R}}_{q+j,1:q+j-1}$ of the $(q+j)^{th}$ row of $\widehat{\mathbf{R}}$ containing the first $(q+j-1)$ elements is computed (step 1) as the product between (i) the sub-row $\mathbf{A}_{k,1:q+j-1}$ of the j^{th} row of $\mathbf{A}$ containing the first $(q+j-1)$ elements, (ii) and the block $\widehat{\mathbf{R}}_{1:q+j-1,1:q+j-1}$ where $\widehat{\mathbf{R}}_{1:q+j-1,1:q+j-1}$ is the sub-matrix of $\widehat{\mathbf{R}}$ obtained by extracting the first $(q+j-1)$ rows and the first $(q+j-1)$ columns. In step 2, the sub-column $\widehat{\mathbf{R}}_{1:q+j-1,q+j}$ of the $(q+j)^{th}$ column of $\widehat{\mathbf{R}}$ is computed as being the transpose of the sub-row $\widehat{\mathbf{R}}_{q+j,1:q+j-1}$. The $(q+j)^{th}$ diagonal element of $\widehat{\mathbf{R}}$ is set to 1 (step 3). Steps 1 to 3 are iterated p times over j.

And in order to answer the second question, El Hadri et al. (2023) introduce in 2022 a new procedure to estimate $\boldsymbol{\rho} = (\rho_1, \cdots, \rho_I)'$ where I is the size of $\boldsymbol{\rho}$. Formally, the elements of this vector are estimated alternately and successively (one after an other) by minimizing a univariate function each time defined as follows:

$$F_i(v) = F(\rho_1, \cdots, \rho_{i-1}, v, \rho_{i+1}, \cdots, \rho_I) \tag{17}$$

Formally, we fix $\rho_2, \cdots$ and ρ_I and solve for ρ_1. Once ρ_1 is found, we fix it at its value and keep $\rho_3, \cdots$ and ρ_I fixed and solve for ρ_2. We continue in the same way by fixing $\rho_1, \cdots$ and ρ_{I-1} at their recent values and we solve for ρ_I. Then we go back to update ρ_1 and the process is repeated until convergence. i.e. $F(\boldsymbol{\rho}^{(s)}) - F(\boldsymbol{\rho}^{(s+1)})$ is less than a small fixed value. Note that the initial values are arbitrarily chosen. The procedure is defined as the following :

Algorithm 2 Procedure to estimate the vector of parameters associated with a given recursive PAM.

Initialization : Choose arbitrary values $\rho_1^{(0)}, \ldots, \rho_I^{(0)}$ and set $s = 0$.

While $\| \boldsymbol{\rho}^{(s+1)} - \boldsymbol{\rho}^{(s)} \| \geq$ threshold.

Iterate over $t = 1, \cdots, I$.

Step 1 : Compute $\mathbf{M}_i^{(s)} = \widehat{\mathbf{R}}\left(\rho_1^{(s+1)}, \ldots, \rho_{i-1}^{(s+1)}, 0, \rho_{i+1}^{(s)}, \ldots, \rho_I^{(s)}\right)$ using FIM.

Step 2 : Compute $\tilde{\mathbf{N}}_i^{(s)} = \widehat{\mathbf{R}}\left(\rho_1^{(s+1)}, \ldots, \rho_{i-1}^{(s+1)}, 1, \rho_{i+1}^{(s)}, \ldots, \rho_I^{(s)}\right)$ using FIM.

Step 3 : Compute $\mathbf{N}_i^{(s)} = \tilde{\mathbf{N}}_i^{(s)} - \mathbf{M}_i^{(s)}$.

Step 4 : Compute $\rho_i^{(s+1)} = \dfrac{\left\langle \mathbf{N}_i^{(s)}, \mathbf{R} - \mathbf{M}_i^{(s)} \right\rangle}{\left\| \mathbf{N}_i^{(s)} \right\|_F^2}$.

End.

$\boldsymbol{\rho}^{(s)} \leftarrow \boldsymbol{\rho}^{(s+1)}$.

End.

The procedure begins with an arbitrary choice of initialization. Suppose that $\rho_1^{(s+1)}, \ldots, \rho_{i-1}^{(s+1)}$ are computed, the matrices $\mathbf{M}_i^{(s)}$ and $\tilde{\mathbf{N}}_i^{(s)}$ are computed by FIM by setting respectively $\rho_i = 0$ and $\rho_i = 1$ (step 1 and step 2). These two matrices allow to compute the matrix $\mathbf{N}_i^{(s)}$ (step 3). Then ρ_i is updated in step 4 by the minimizer of the function given in (17). Steps 1 to 4 are iterated I times over i. The procedure is iterated until the quantity $\| \boldsymbol{\rho}^{(s+1)} - \boldsymbol{\rho}^{(s)} \|$ reaches a given threshold.

2.3 Reflective Path Analysis Model

Definition 5. *A recursive PAM is called Reflective Path Analysis Model (RPAM) if it contains one exogenous variable, all other variables are endogenous and explained by the exogenous variable and no endogenous variable is explained by another endogenous variable (figure (2 (b)).*

A direct consequence of definition 5 is that for a RPAM, $\mathbf{\Phi} = 1$, $\mathbf{\Gamma} = (\gamma_{11}, \gamma_{21}, \cdots, \gamma_{p1})'$, $\mathbf{B} = \mathbf{0}$, $\boldsymbol{\rho} = \mathbf{\Gamma}'$ and $I = p$.

Proposition 1. *The correlation matrix implied by a RPAM involving p endogenous variables and defined in (16) is :*

$$\widehat{\mathbf{R}} = \begin{pmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_p \\ \rho_1 & 1 & \rho_1\rho_2 & \cdots & \rho_1\rho_p \\ \rho_2 & \rho_1\rho_2 & 1 & & \\ \vdots & \vdots & \vdots & \ddots & \rho_{p-1}\rho_p \\ \rho_p & \rho_1\rho_p & \cdots & \rho_{p-1}\rho_p & 1 \end{pmatrix} \tag{18}$$

Proof. See Appendix. □

In addition, the empirical correlation matrix of dimension $(p \times p)$ among endogenous variables is denoted by $\mathbf{\Omega}$:

$$\mathbf{\Omega} = [\omega_{j,k}]_{1 \leq j,k \leq p} = \mathbb{E}(\boldsymbol{\eta}^{(data)} \boldsymbol{\eta}^{(data)'}) \tag{19}$$

and the empirical correlation vector of dimension p between the p endogenous variables and the exogenous variable is denoted by $\mathbf{H}$:

$$\mathbf{H} = [h_j]_{1 \leq j \leq p} = \mathbb{E}(\boldsymbol{\eta}^{(data)} \boldsymbol{\xi}^{(data)'}) \tag{20}$$

It follows that the empirical correlation matrix defined in (15) is expressed as:

$$\mathbf{R} = \begin{pmatrix} 1 & h_1 & h_2 & \cdots & h_p \\ h_1 & 1 & \omega_{2,1} & \cdots & \omega_{p,1} \\ h_2 & \omega_{2,1} & 1 & & \\ \vdots & \vdots & \vdots & \ddots & \omega_{p,p-1} \\ h_p & \omega_{p,1} & \cdots & \omega_{p,p-1} & 1 \end{pmatrix} \tag{21}$$

Corollary 1. *The update expression given in step 4 of algorithm 2 (the argmin of the function given in (17) is :*

$$\rho_j^{(s+1)} = \frac{h_{j1} + \sum_{l=1}^{j-1} \omega_{jl} \rho_l^{(s+1)} + \sum_{l=j+1}^{p} \omega_{jl} \rho_l^{(s)}}{1 + \sum_{l=1}^{j-1} \left(\rho_l^{(s+1)}\right)^2 + \sum_{l=j+1}^{p} \left(\rho_l^{(s)}\right)^2}$$

Proof. See Appendix. □

3 PLSFIM algorithm

This section is devoted to introducing a new PLSPM algorithm. The novelty affects mainly the estimation of weights performed in (8). To do so, at each iteration s, ($s = 0, 1, \cdots$) and for each k, ($1 \leq k \leq K$), the weights vector $\tilde{\mathbf{w}}_k^{(s+1)}$ will be estimated using corollary 1 in section 2. For this purpose, we start by introducing the following definition.

Definition 6. *For k, ($1 \leq k \leq K$), and s, ($s = 0, 1, \cdots$), the model containing the p_k manifest variables $\mathbf{X}_{k1}, \cdots, \mathbf{X}_{kp_k}$ and the internal estimation $\mathbf{z}_k^{(s)}$ is called the k^{th} augmented block at iteration s.*

It is clear that when Mode A is used for the k^{th} block, the associated augmented block is RPAM. In particular, $\mathbf{z}_k^{(s)}$ is the exogenous variable and $\mathbf{X}_{k1}, \cdots, \mathbf{X}_{kp_k}$ are the endogenous variables. Using the procedure proposed by El Hadri et al. (2023), the associated vector of weights $\tilde{\mathbf{w}}_k^{(s+1)}$ can estimated by algorithm 3 below.

Algorithm 3 Algorithm to estimate vector of weights associated with the k^{th} block.

Initialization : Choose arbitrary values $\left[\tilde{w}_{k1}^{(s+1)}\right]^{(0)}, \ldots, \left[\tilde{w}_{kp_k}^{(s+1)}\right]^{(0)}$ and set $t = 0$.

While $\left\| \left[\tilde{\mathbf{w}}_k^{(s+1)}\right]^{(t+1)} - \left[\tilde{\mathbf{w}}_k^{(s+1)}\right]^{(t)} \right\| \geq threshold$

Iterate for $j = 1, \cdots, p_k$

Compute $\left[\tilde{w}_{kj}^{(s+1)}\right]^{(t+1)} = \frac{1}{n} \times \frac{\mathbf{X}'_{kj}\mathbf{z}_k^{(s)} + \sum_{l=1}^{j-1} \mathbf{X}'_{kj}\mathbf{X}_{kl} \left[\tilde{w}_{kl}^{(s+1)}\right]^{(t+1)} + \sum_{l=j+1}^{p_k} \mathbf{X}'_{kj}\mathbf{X}_{kl} \left[\tilde{w}_{kl}^{(s+1)}\right]^{(t)}}{1 + \sum_{l=1}^{j-1} \left(\left[\tilde{w}_{kl}^{(s+1)}\right]^{(t+1)}\right)^2 + \sum_{l=j+1}^{p_k} \left(\left[\tilde{w}_{kl}^{(s+1)}\right]^{(t)}\right)^2}.$

End.

$\left[\tilde{\mathbf{w}}_k^{(s+1)}\right]^{(t)} \leftarrow \left[\tilde{\mathbf{w}}_k^{(s+1)}\right]^{(t+1)}$.

End.

Algorithm 3 allows to build a new algorithm to compute scores of LVs in PLSPM framework. We call this algorithm *Finite Iterative Method based algorithm (PLSFIM)* and it is defined as follows (figure 4).

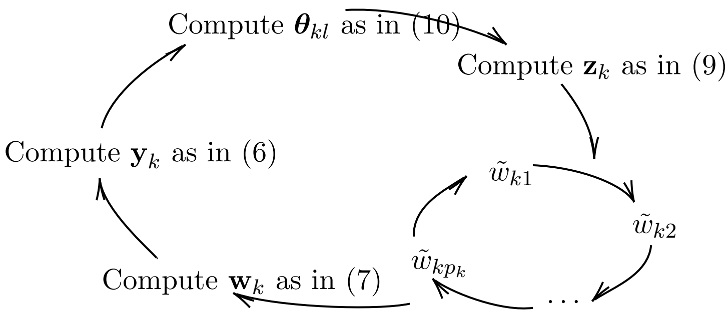

Figure 4: PLSFIM algorithm to compute the scores of LVs.

It is important to note that PLSFIM and PLSPM algorithms are clearly different. In deed, the estimates of vectors of weights are provided by these algorithms using different criteria. Basically, PLSPM algorithm computes the associated weights for the k^{th} block (see the first equation in (8)) by minimizing the following criterion:

$$\frac{1}{2}\sum_{j=1}^{p_k}\left[\tilde{w}_{kj} - \mathbf{X}'_{kj}\mathbf{z}_k^{(s)}\right]^2 \quad (22)$$

whereas, from (23), the criterion used in PLSFIM algorithm is :

$$\frac{1}{2}\sum_{j=1}^{p_k}\left[\tilde{w}_{kj} - \mathbf{X}'_{kj}\mathbf{z}_k^{(s)}\right]^2 + \frac{1}{2}\sum_{j,l=1,j\neq l}^{p_k}\left(\tilde{w}_{kj}\tilde{w}_{kl} - \mathbf{X}'_{kj}\mathbf{X}_{kl}\right)^2 \quad (23)$$

The quantity $\frac{1}{2}\sum_{j,l=1,j\neq l}^{p_k}\left(\tilde{w}_{kj}\tilde{w}_{kl} - \mathbf{X}'_{kj}\mathbf{X}_{kl}\right)^2$ which is the difference between the two criteria represents the ignored part when the model is considered as p_k simple regression models. This constitutes the main justification for the failure of Mode A stated in section 1.

One can remark that the PLSFIM seems to be more computationally demanding. In this situation, much better estimation scores of LVs are expected to compensate these computational efforts. In this context, is PLSFIM more efficient than PLSPM? In order to respond to this question, the communality index defined in (5) is considered as a criterion of efficiency. Ideally, one can provide a formal proof that compare Com^{PLSFIM} and Com^{PLSPM}. Unfortunately, this proof seems not to be an easy task and needs deep reflection. Consequently it will not be discussed in the present paper and we limit this comparison in a numerical level. The following section is dedicated to this feature.

4 Numerical Comparisons

Several studies using real and simulated data will be given in this section to highlight the advantages of PLSFIM compared to the classical algorithm PLSPM respectively described in figure 4 and figure 3 for reflective blocks. Note that all results concern Lohmöller's procedure. We remark that Hanafi-Wold's procedure provides exactly the same results.

4.1 Comparison on real data

To clarify the comparison between the two algorithms, it is very useful to refer to a practical example where PLSPM has been applied very extensively. In this context, The European Consumer Satisfaction Index (ECSI) Tenenhaus et al. (2005) is considered. It is an analytical tool designed to provide a solid basis for selecting the right marketing strategy. Using ECSI, the company can discover what the most important factors are for the satisfaction and loyalty of customers. The corresponding dataset consists of 250 consumer responses to service units and 27 variables (27 MVs) grouped into 6 blocks (6 LVs). This suggests that Mode A is used to compute the weights in PLSPM. Finally, the R plspm Sanchez (2013) package is used to implement PLSPM. Figure 5 below represents the conceptual configuration of the ECSI model.

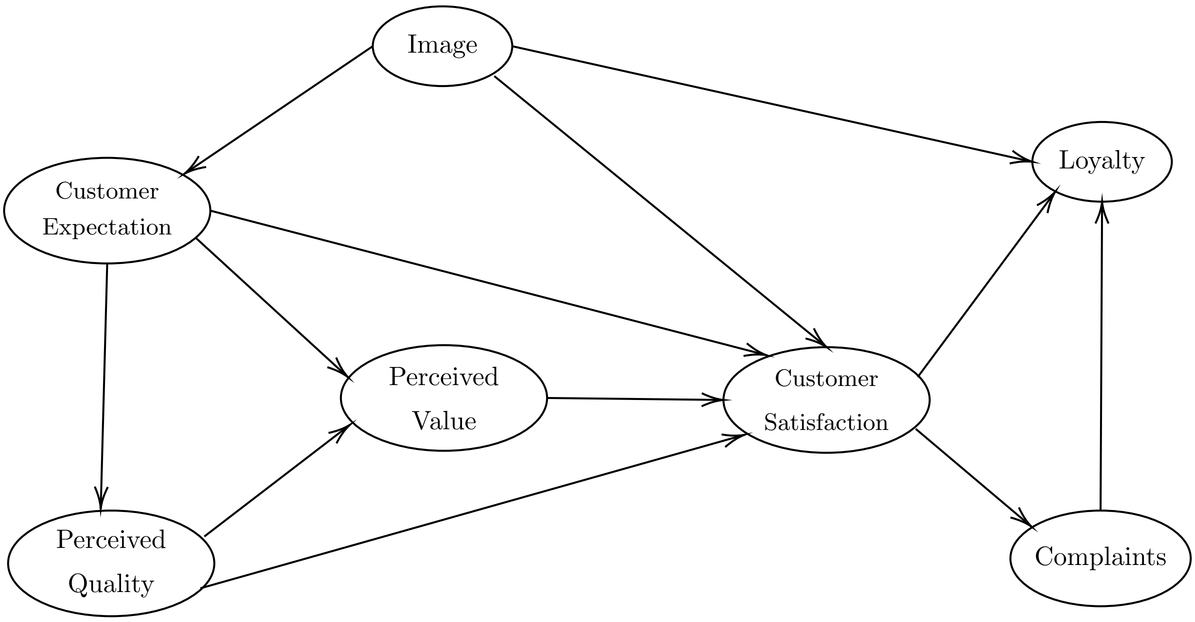

Figure 5: Causality model describing causes and consequences of customer satisfaction.

PLSFIM and PLSPM are applied to this model to compute the scores of LVs. They are initialized by the same vectors of weights which values are all chosen to be equal to 0.5 and the threshold is fixed to 10^{-7}. In addition, the threshold used in PLSFIM is fixed at 10^{-9}. The estimates of weights and scores of LVs obtained for centroid and factorial schemes by the two algorithms are given in Appendix 5.

The respective values of communality index defined in (5) obtained by PLSPM and PLSFIM are $Com^{PLSPM} = 0.659692$ and $Com^{PLSFIM} = 0.660761$ for centroid scheme and $Com^{PLSPM} = 0.659697$ and $Com^{PLSFIM} = 0.660757$ for factorial scheme. This allows us to conclude that the part of variance of MVs explained by LVs estimated by PLSFIM is more important than those estimated by PLSPM.

4.2 Comparison on simulated data 1

In the perspective of investigating if PLSFIM algorithm is always more efficient than PLSPM algorithm, we consider the same model ECSI described in Figure 5 but with simulated data. In this context and in order to avoid the problem of non convergence of PLSPM Henseler (2010), we control the correlation between the MVs, and the Cronbach's alpha to ensure the unidimensionality for all blocks Tenenhaus et al. (2005). 100 data sets are generated using `dplyr`, `tidyr`, and `faux` packages in the R software such that the correlation between the MVs is ranged in the interval $[0.7; 0.9]$ and the Cronbach's alpha fixed between 0.91 and 0.98 for all blocks. Thereafter, PLSPM and PLSFIM are initialized with the same vectors of weights and applied with the threshold 10^{-7} to estimate the scores of LVs. The communality index defined in (5) is then computed for both algorithms. Figure 6 below depicts the evolution of this index for centroid and factorial schemes.

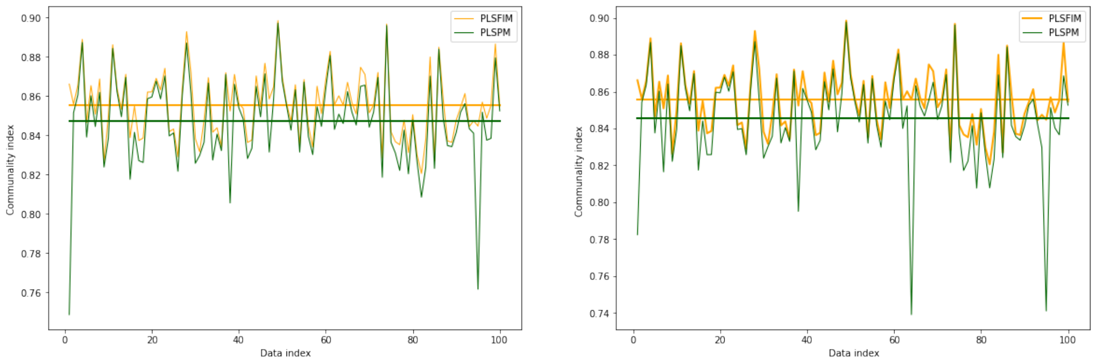

Figure 6: Evolution of the communality index for centroid (Left) and factorial (Right) schemes for simulation 1.

Figure 6 shows that the communality provided by the PLSFIM is always higher than that obtained by PLSPM. In deed, the respective mean values are 0.855 and 0.845 for centroid scheme and the respective mean values are 0.855 and 0.847 for factorial scheme. Based on this simulation, we can confirm that the LVs estimated by PLSFIM are more related to their own MVs than those estimated by PLSPM. However, this advantage of PLSFIM is not substantially significant.

4.3 Comparison on simulated data 2

In this section, we propose to show that the advantage of PLSFIM algorithm over PLSPM algorithm can be more significant. To do so, we consider the model given in figure 7 below.

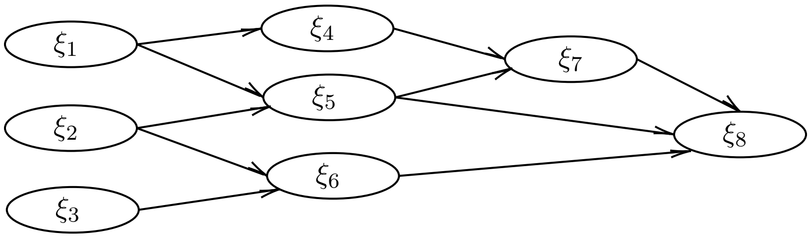

Figure 7: Conceptual model with 8 blocks.

This model contains 8 blocks with different numbers of MVs, see Table 1 below.

Block	ξ_1	ξ_2	ξ_3	ξ_4	ξ_5	ξ_6	ξ_7	ξ_8
Number of MVs	5	6	7	4	5	4	4	8

Table 1: Number of MVs per block for model associated with figure 7

Thereafter, 100 datasets are generated on 4 configurations such that the correlation between the MVs is respectively ranged in the intervals $[0.5; 0.6]$, $[0.6; 0.7]$, $[0.7; 0.8]$ and $[0.8; 0.9]$ using the same method as in section 4.2. Thereafter, the communality index is computed for each configuration and for each algorithm. Figure 8 below represents the evolution of this index for centroid and factorial schemes. In addition, table 2 below summarizes the means of communalities for both algorithms, both schemes and four configurations.

Figure 8 shows that communality index obtained by using PLSFIM is clearly more important than the one obtained by PLSPM. More precisely, table 2 shows that the communality obtained using PLSFIM is informatively greater than the one obtained by PLSPM whatever the chosen scheme and the considered configuration.

Correlation	Centroid		Factorial	
	PLSPM	PLSFIM	PLSPM	PLSPIM
[0,5;0,6]	0.45	0.56	0.46	0.56
[0,6;0,7]	0.52	0.62	0.53	0.62
[0,7;0,8]	0.57	0.68	0.60	0.69
[0,8;0,9]	0.64	0.76	0.65	0.75

Table 2: Means of communalities for PLSPM and PLSFIM, both schemes and four configurations.

Figure 8: Evolution of the communality index for simulation 2 for centroid (Left) and factorial (Right) schemes.

5 Conclusion and perspectives

We have introduced in this paper a new algorithm to estimate the so-called weights in PLS-SEM framework when all blocks are considered to be reflective. The proposed algorithm is based on the Finite Iterative Method and allows to build a new algorithm to compute the scores of LVs.

On one hand, this contribution concerns only the outer model. Can we propose analog results for the inner model? The response to this question is the subject of future work and it is in advance. Moreover, new quality indices can be constructed to test the fit of models.

On the other hand, the advantages of this new algorithm are highlighted. Indeed we have presented a real study and two numerical simulations that show that the proposed algorithm is more efficient than the classical algorithm in terms of commonality index. Meanwhile, it will be interesting to provide formal proof of this efficiency.

Acknowledgements

The authors gratefully acknowledge the support of editors during the publication process as well as the useful feedback provided by two anonymous referees.

References

N. K. Avkiran and C. M. Ringle. Partial least squares structural equation modeling: Recent advances in banking and finance. 2018.

J. Benitez, J. Henseler, A. Castillo, and F. Schuberth. How to perform and report an impactful analysis using partial least squares: Guidelines for confirmatory and explanatory is research. *Information & management*, 57(2):103168, 2020.

K. A. Bollen. *Structural equations with latent variables*, volume 210. John Wiley & Sons, 1989.

W. Chin. Pls-graph user's guide. *CT Bauer College of Business/University of Houston*, 2001.

G. Cho and J. Y. Choi. An empirical comparison of generalized structured component analysis and partial least squares path modeling under variance-based structural equation models. *Behaviormetrika*, 47(1):243–272, 2020.

C. Crocetta, L. Antonucci, R. Cataldo, R. Galasso, M. G. Grassia, C. N. Lauro, and M. Marino. Higher-order pls-pm approach for different types of constructs. *Social Indicators Research*, 154:725–754, 2021.

L. Eboli, C. Forciniti, and G. Mazzulla. Formative and reflective measurement models for analysing transit service quality. *Public Transport*, 10:107–127, 2018.

Z. El Hadri and M. Hanafi. The finite iterative method for calculating the correlation matrix implied by a recursive path model. *Electronic Journal of Applied Statistical Analysis*, 8(1), 2015.

Z. El Hadri, A. Sahli, and M. Hanafi. Simple and fast convergent procedure to estimate recursive path analysis model. *Behaviormetrika*, 50(1):317–333, 2023.

R. Fabbricatore, M. Iannario, R. Romano, and D. Vistocco. Component-based structural equation modeling for the assessment of psycho-social aspects and performance of athletes: Measurement and evaluation of swimmers. *AStA advances in statistical analysis*, 107(1):343–367, 2023.

J. F. Hair, M. Sarstedt, T. M. Pieper, and C. M. Ringle. The use of partial least squares structural equation modeling in strategic management research: a review of past practices and recommendations for future applications. *Long range planning*, 45 (5-6):320–340, 2012a.

J. F. Hair, M. Sarstedt, C. M. Ringle, and J. A. Mena. An assessment of the use of partial least squares structural equation modeling in marketing research. *Journal of the academy of marketing science*, 40:414–433, 2012b.

J. F. Hair Jr, G. T. M. Hult, C. M. Ringle, M. Sarstedt, N. P. Danks, and S. Ray. *Partial least squares structural equation modeling (PLS-SEM) using R: A workbook*. Springer Nature, 2021.

M. Hanafi. Pls path modelling: computation of latent variables with the estimation mode b. *Computational Statistics*, 22(2):275–292, 2007.

M. Hanafi, P. Dolce, and Z. El Hadri. Generalized properties for hanafi–wold's procedure in partial least squares path modeling. *Computational Statistics*, 36:603–614, 2021.

M. Hanafi, Z. El Hadri, A. Sahli, and P. Dolce. Overcoming convergence problems in pls path modelling. *Computational Statistics*, 37(5):2437–2470, 2022.

J. Henseler. On the convergence of the partial least squares path modeling algorithm. *Computational statistics*, 25:107–120, 2010.

H. Hwang, Y. Takane, and A. Tenenhaus. An alternative estimation procedure for partial least squares path modeling. *Behaviormetrika*, 42:63–78, 2015.

H. Hwang, M. Sarstedt, J. H. Cheah, and C. M. Ringle. A concept analysis of methodological research on composite-based structural equation modeling: bridging plspm and gsca. *Behaviormetrika*, 47:219–241, 2020.

R. B. Kline. *Principles and practice of structural equation modeling*. Guilford publications, 2023.

J.-B. Lohmöller. *Latent variable path modeling with partial least squares*. Springer Science & Business Media, 2013.

V. D. Mayrink, R. V. Panaro, and M. A. Costa. Structural equation modeling with time dependence: an application comparing brazilian energy distributors. *AStA Advances in Statistical Analysis*, 105:353–383, 2021.

M. M. Pelagatti, M. Fattore, and G. Vittadini. Inconsistencies of the pls-pm approach to structural equation models with formative-reflective schemes. *Electronic Journal of Applied Statistical Analysis*, 5(3):333–338, 2012.

D. X. Peng and F. Lai. Using partial least squares in operations management research: A practical guideline and summary of past research. *Journal of operations management*, 30(6):467–480, 2012.

C. Ringle, D. Da Silva, and D. Bido. Structural equation modeling with the smartpls. *Bido, D., da Silva, D., & Ringle, C.(2014). Structural Equation Modeling with the Smartpls. Brazilian Journal Of Marketing*, 13(2), 2015.

G. Russolillo. *Partial Least Squares Methods for Non-Metric Data.* PhD thesis, University of Naples Federico II, Italy, 2009.

A. Sahli, S. A. Ebnou Abdem, M. Hanafi, and Z. El Hadri. Extending el-hadri-sahli-hanafi procedure for path analysis with non standardized variables. *Quality & Quantity*, pages 1–16, 2024.

G. Sanchez. Pls path modeling with r. *Trowchez Editions*, 2013.

S. Streukens, S. Leroi-Werelds, and K. Willems. Partial least squares path modeling: Basic concepts, methodological issues and applications, 2017.

Y. Takane and H. Hwang. Comparisons among several consistent estimators of structural equation models. *Behaviormetrika*, 45:157–188, 2018.

M. Tenenhaus, V. E. Vinzi, Y.-M. Chatelin, and C. Lauro. Pls path modeling. *Computational statistics & data analysis*, 48(1):159–205, 2005.

V. Vinzi. Handbook of partial least squares, 2010.

H. Wold. Nonlinear estimation by iterative least squares procedures in: David, fn (hrsg.), festschrift for j. *Neyman: Research Papers in Statistics, London*, 1966.

A. XLSTAT. Statistical and data analysis solution. *Addinsoft XLSTAT: New York, NY, USA*, 2019.

Appendix 1

Proof. (of proposition 1). Using the FIM algorithm 1, the implied correlation matrix $\widehat{\mathbf{R}}$ is computed iteratively as follows :

i. Initialisation: $\widehat{\mathbf{R}}_{1:1,1:1} = \mathbf{\Phi} = 1$.

ii. Let j be an integer such that $1 \leq j \leq p$ and suppose that

$$\widehat{\mathbf{R}}_{1:j,1:j} = \begin{pmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{j-1} \\ \rho_1 & 1 & \rho_1\rho_2 & \cdots & \rho_1\rho_{j-1} \\ \rho_2 & \rho_1\rho_2 & 1 & & \\ \vdots & \vdots & \vdots & \ddots & \rho_{j-2}\rho_{j-1} \\ \rho_{j-1} & \rho_1\rho_{j-1} & \cdots & \rho_{j-2}\rho_{j-1} & 1 \end{pmatrix}.$$

iii. $\widehat{\mathbf{R}}_{j+1,1:j} = \mathbf{A}_{j,1:j}\widehat{\mathbf{R}}_{1:j,1:j} = \begin{pmatrix}\rho_j & 0\cdots & 0\end{pmatrix}\begin{pmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{j-1} \\ \rho_1 & 1 & \rho_1\rho_2 & \cdots & \rho_1\rho_{j-1} \\ \rho_2 & \rho_1\rho_2 & 1 & & \\ \vdots & \vdots & \vdots & \ddots & \rho_{j-2}\rho_{j-1} \\ \rho_{j-1} & \rho_1\rho_{j-1} & \cdots & \rho_{j-2}\rho_{j-1} & 1 \end{pmatrix}$

$= \begin{pmatrix}\rho_j & \rho_1\rho_j\cdots & \rho_{j-1}\rho_j\end{pmatrix}$

iv. $\widehat{\mathbf{R}}_{1:j,j+1} = \widehat{\mathbf{R}}'_{j+1,1:j} = \begin{pmatrix}\rho_j \\ \rho_1\rho_j \\ \vdots \\ \rho_{j-1}\rho_j\end{pmatrix}$

v. $\widehat{\mathbf{R}}_{j+1,j+1} = 1$

vi. $\widehat{\mathbf{R}}_{1:j,j+1} = \begin{pmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_j \\ \rho_1 & 1 & \rho_1\rho_2 & \cdots & \rho_1\rho_j \\ \rho_2 & \rho_1\rho_2 & 1 & & \\ \vdots & \vdots & \vdots & \ddots & \rho_{j-1}\rho_j \\ \rho_j & \rho_1\rho_j & \cdots & \rho_{j-1}\rho_j & 1 \end{pmatrix}$

□

Appendix 2

Proof. (of corollary 1).
On one side, proposition 1 gives

$$\widehat{\mathbf{R}}(\rho_j = 0) = \begin{pmatrix} 1 &&&&&&&& \\ \rho_1 & 1 &&&&&&& \\ \rho_2 & \rho_1\rho_2 & 1 &&&&&& \\ \vdots & \vdots & \vdots & \ddots &&&&& \\ \rho_{j-1} & \rho_1\rho_{j-1} & \cdots & \rho_{j-2}\rho_{j-1} & 1 &&&& \\ 0 & 0 & \cdots & \cdots & 0 & 1 &&& \\ \rho_{j+1} & \rho_1\rho_{j+1} & \cdots & \cdots & \cdots & 0 & 1 && \\ \vdots & \vdots & \vdots & \cdots & \cdots & \cdots & \cdots & 1 & \\ \rho_p & \rho_1\rho_p & \cdots & \cdots & 0 & \cdots & \cdots & \rho_{p-1}\rho_p & 1 \end{pmatrix}$$

And

$$\widehat{\mathbf{R}}(\rho_j=1)=\begin{pmatrix}1 & & & & & & & & \\ \rho_1 & 1 & & & & & & & \\ \rho_2 & \rho_1\rho_2 & 1 & & & & & & \\ \vdots & \vdots & \vdots & \ddots & & & & & \\ \rho_{j-1} & \rho_1\rho_{j-1} & \cdots & \rho_{j-2}\rho_{j-1} & 1 & & & & \\ 1 & \rho_1 & \cdots & \cdots & \rho_{j-1} & 1 & & & \\ \rho_{j+1} & \rho_1\rho_{j+1} & \cdots & \cdots & \cdots & \rho_{j+1} & 1 & & \\ \vdots & \vdots & \vdots & \cdots & \cdots & \cdots & \cdots & 1 & \\ \rho_p & \rho_1\rho_p & \cdots & \cdots & \rho_p & \cdots & \cdots & \rho_{p-1}\rho_p & 1\end{pmatrix}$$

Thus

$$\mathbf{N}_j=\widehat{\mathbf{R}}(\rho_j=1)-\widehat{\mathbf{R}}(\rho_j=0)=\begin{pmatrix}0 & & & & & & & & \\ 0 & 0 & & & & & & & \\ 0 & 0 & 0 & & & & & & \\ \vdots & \vdots & \vdots & \ddots & & & & & \\ 0 & 0 & \cdots & 0 & 0 & & & & \\ 1 & \rho_1 & \cdots & \cdots & \rho_{j-1} & 0 & & & \\ 0 & 0 & \cdots & \cdots & \cdots & 0 & 0 & & \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & 0 & \\ 0 & 0 & \cdots & \cdots & \rho_p & \cdots & \cdots & \cdots & 0\end{pmatrix}$$

Thus

$$\|\mathbf{N}_j\|_F^2=2\left(1+\sum_{l=1,l\neq j}^{p}\rho_l^2\right)$$

And on the other side, (21) gives

$$\mathbf{R}-\mathbf{M}_j=\begin{pmatrix}0 & & & & & & & & \\ h_1-\rho_1 & 0 & & & & & & & \\ h_2-\rho_2 & \omega_{12}-\rho_1\rho_2 & 0 & & & & & & \\ \vdots & \vdots & \vdots & \ddots & & & & & \\ h_{j-1}-\rho_{j-1} & \omega_{1,j-1}-\rho_1\rho_{j-1} & \cdots & \omega_{j-2,j-1}-\rho_{j-2}\rho_{j-1} & 0 & & & & \\ h_j & \omega_{1,j} & \cdots & \cdots & \omega_{j-1,j} & 0 & & & \\ h_{j+1}-\rho_{j+1} & \omega_{1,j+1}-\rho_1\rho_{j+1} & \cdots & \cdots & \cdots & 0 & 0 & & \\ \vdots & \vdots & \vdots & \cdots & \cdots & \cdots & \cdots & 0 & \\ h_p-\rho_p & \omega_{1,p}-\rho_1\rho_p & \cdots & \cdots & \omega_{p,j} & \cdots & \cdots & \omega_{p-1,p}-\rho_{p-1}\rho_p & 0\end{pmatrix}$$

Thus

$$\langle \mathbf{N}_j, \mathbf{R} - \mathbf{M}_j \rangle = 2\left(h_j + \sum_{l=1, l\neq j}^{p} \omega_{jl}\rho_l\right)$$

Finally

$$\widehat{\rho}_j = \frac{\langle \mathbf{N}_j, \mathbf{R} - \mathbf{M}_j \rangle}{\|\mathbf{N}_j\|_F^2} = \frac{h_j + \sum\limits_{l=1, l\neq j}^{p} \omega_{jl}\rho_l}{1 + \sum\limits_{l=1, l\neq j}^{p} \rho_l^2}$$

□

Appendix 3

Centroid scheme					
Image		Expectation		Quality	
PLSFIM	PLSPM	PLSFIM	PLSPM	PLSFIM	PLSPM
0.240	0.206	0.250	0.235	0.237	0.241
0.311	0.297	0.273	0.281	0.272	0.268
0.300	0.306	0.227	0.223	0.224	0.225
0.197	0.181	0.248	0.261	0.248	0.245
0.220	0.286	0.267	0.265	0.245	0.246
Value		Satisfaction		Loyalty	
PLSFIM	PLSPM	PLSFIM	PLSPM	PLSFIM	PLSPM
0.336	0.354	0.312	0.317	0.360	0.374
0.305	0.285	0.312	0.317	0.260	0.251
0.267	0.249	0.261	0.248	0.357	0.371
0.308	0.327	0.257	0.260	0.251	0.223
Factorial scheme					
Image		Expectation		Quality	
PLSFIM	PLSPM	PLSFIM	PLSPM	PLSFIM	PLSPM
0.240	0.205	0.250	0.236	0.237	0.242
0.311	0.297	0.273	0.281	0.272	0.268
0.300	0.307	0.227	0.224	0.224	0.225
0.197	0.181	0.248	0.259	0.248	-0.245
0.219	0.284	0.267	0.265	0.235	0.246
Value		Satisfaction		Loyalty	
PLSFIM	PLSPM	PLSFIM	PLSPM	PLSFIM	PLSPM
0.336	0.354	0.312	0.317	0.360	0.374
0.306	0.285	0.312	0.317	0.251	0.250
0.267	0.249	0.261	0.248	0.354	0.371
0.308	0.327	0.257	0.259	0.251	0.222

Table 3: Weights obtained using PLSPM and PLSFIM for centroid and factorial scheme associated with ECSI model given in part 4.1.

Centroid scheme												
Item	Image		Expectation		Quality		Value		Satisfaction		Loyalty	
	PLSFIM	PLSPM	PLSFIM	PLSPM	PLSFIM	PLSPM	PLSFIM	PLSPM	PLSFIM	PLSPM	PLSFIM	PLSPM
1	-0.164	-0.145	0.362	0.357	-0.471	-0.475	0.050	0.038	-0.205	-0.202	0.176	0.194
2	0.931	0.911	0.738	0.720	0.374	0.372	0.398	0.416	0.382	0.376	0.526	0.526
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
249	0.068	0.148	-0.206	-0.186	0.337	0.347	0.681	0.66 3	0.537	0.533	-0.227	-0.204
250	0.503	0.532	-0.631	-0.643	-0.706	-0.701	0.367	0.369	-0.043	-0.046	-0.239	-0.236
Factorial scheme												
Item	Image		Expectation		Quality		Value		Satisfaction		Loyalty	
	PLSFIM	PLSPM	PLSFIM	PLSPM	PLSFIM	PLSPM	PLSFIM	PLSPM	PLSFIM	PLSPM	PLSFIM	PLSPM
1	-0.164	-0.144	0.362	0.357	-0.472	-0.476	0.051	0.040	-0.205	-0.203	0.176	0.194
2	0.931	0.912	0.738	0.722	0.374	0.371	0.398	0.416	0.382	0.376	0.526	0.526
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
249	0.068	0.149	-0.206	-0.186	0.338	0.348	0.681	0.663	0.537	0.533	-0.227	-0.203
250	0.502	0.530	-0.630	-0.643	-0.707	-0.712	0.367	0.369	-0.043	-0.046	-0.239	-0.236

Table 4: Scores of LVs obtained by PLSPM and PLSFIM for centroid and factorial schemes associated with ECSI model given in part 4.1.