

PDF hosted at the Radboud Repository of the Radboud University Nijmegen

The following full text is an author's version which may differ from the publisher's version.

For additional information about this publication click this link.

<http://hdl.handle.net/2066/91378>

Please be advised that this information was generated on 2021-05-09 and may be subject to change.

A METHODOLOGICAL COMPARISON BETWEEN PLS PATH MODELING AND GENERALIZED STRUCTURED COMPONENT ANALYSIS

Laura Trinchera, Wynne W. Chin, Vincenzo Esposito Vinzi & Jörg Henseler

*SUPELEC, Département Signaux & Systèmes Électroniques
Plateau de Moulon - 3, rue Joliot-Curie, 91192 Gif-sur-Yvette, France,
E-mail: laura.trinchera@supelec.fr*

*University of Houston, Department of Decision and Information Sciences
C.T. Bauer College of Business 334 Melcher Hall Houston, Texas 77204-6282
E-mail: wchin@uh.edu*

*ESSEC Business School of Paris,
Avenue Bernard Hirsch, B.P. 50105, 95021 Cergy-Pontoise, France,
E-mail: vinzi@essec.edu*

*Radboud University Nijmegen, Nijmegen School of Management
Thomas van Aquinostraat 3 6525 GD Nijmegen The Netherlands
E-mail: j.henseler@fm.ru.nl*

Mots-clés: Component-based approach to SEM, GSCA, PLS-PM

Résumé: L'approche PLS aux modèles à équations structurelles (PLS Path Modeling) et l'analyse en composantes structurelles généralisées (Generalized Structured Component Analysis, GSCA) sont deux techniques d'estimation des modèles à équations structurelles basées sur les composantes. Les deux méthodes visent à estimer les relations de cause à effet entre deux ou plusieurs variables latentes au moyen d'une série d'indicateurs observés. Dans chacune des deux méthodes la variable latente est définie comme une combinaison linéaire de ses propres indicateurs observés, soit comme une composante. La principale différence entre les deux approches est dans la méthode d'estimation des paramètres du modèle. L'approche PLS utilise un algorithme itératif basé sur des régressions interdépendantes, tandis que l'analyse en composantes structurelles généralisées définit une formule unique pour le modèle algébrique sous-jacent et utilise l'algorithme des moindres carrés alternés (Alternating Least Squares). Malgré plusieurs études ont été effectuées pour comparer l'approche PLS et les approches basées sur l'estimation de la matrice de covariance, seule une récente étude par simulation implique également l'analyse en composantes structurelles généralisées (Hwang, 2010). Dans ce papier, une nouvelle étude par simulation est présentée afin d'évaluer les performances des deux approches. Nous montrons les liens entre l'analyse en composantes structurelles généralisées et la méthode *Maximum Sum of Explained Variance* (Glang, 1988), ainsi que entre la méthode de Glang

et l'approche PLS. Dans le cadre de l'analyse en composantes structurelles généralisées le modèle de mesure semble jouer un rôle majeur dans l'estimation des paramètres du modèle.

Abstract: PLS Path Modeling and Generalized Structured Component Analysis are two *component-based* approaches to Structural Equation Models. Both methods aim at estimating the causal relationships linking two or more latent variables by means of a set of observed indicators. Moreover, both methods define the latent variable as a linear combination of its own observed variables, i.e. as a component. The main difference between the two approaches is in the estimation of the model parameters. In fact, PLS Path Modeling uses an iterative algorithm based on a series of interdependent regressions, while the Generalized Structured Component Analysis defines a unique algebraic formulation for the model and uses an Alternative Least Square algorithm. Despite several studies have been performed to compare the PLS Path Modeling and the *covariance-based* approaches to Structural Equation Modeling, only a recent simulation study involves also the Generalized Structured Component Analysis (Hwang *et al.*, 2010). Here, a new simulation study is presented in order to assess the performances of both PLS Path Modeling and Generalized Structured Component Analysis. We discuss the links between Generalized Structured Component Analysis and the Maximum Sum of Explained Variance method (Glang, 1988), as well as the links between the Glang's method and the PLS Path Modeling. Moreover, in Generalized Structured Component Analysis the measurement model seems to play a major role when estimating the model parameters.

Introduction

Two different approaches have been developed since the 1970s to estimate SEM parameters. A first approach aims at reproducing the sample covariance matrix of the manifest variables by means of the model parameters. The fundamental hypothesis underlying this approach is that the implied covariance matrix of the manifest variables is a function of the model parameters. Due to the key role played by the covariance matrix in the estimation process, this approach to SEM has been named *covariance-based*. A different approach is the so-called *component-based* one, where the estimation of latent variable scores plays a central role.

There exist several methods to estimate SEM parameters in a *component-based* framework. The most widely known is Partial Least Squares Path Modeling (PLS-PM) or PLS approach to SEM (Wold, 1975; Tenenhaus *et al.*, 2005). More recently, Hwang & Takane (2004) presented the Generalized Structured Component Analysis (GSCA).

In the next section we briefly introduce GSCA. Then, a study on the similarities and differences among these two *component-based* approaches is presented.

GSCA as a component-based approach to SEM

Generalized Structured Component Analysis (GSCA) has been recently proposed by Hwang and Takane (2004) to estimate Structural Equation Models. The basic idea of GSCA is to integrate the two sub-models defining the measurement and the structural models in a unique formulation, i.e.:

$$\begin{bmatrix} \mathbf{x}_i \\ \xi_i \end{bmatrix} = \begin{bmatrix} \mathbf{\Lambda} \\ \mathbf{B} \end{bmatrix} \xi_i + \begin{bmatrix} \epsilon_i \\ \zeta_i \end{bmatrix} \quad (1)$$

where \mathbf{x}_i is the vector containing all the manifest variables for unit i , ξ_i is the vector of all the latent variables for the i -th unit, $\mathbf{\Lambda}$ is the matrix of the loadings, \mathbf{B} is the square matrix containing the path coefficients of the structural model (an element of \mathbf{B} is equal to zero if the relationship is not included in the model), and ϵ_i and ζ_i are the two vectors of the residuals in the measurement and the structural models respectively.

Moreover, in GSCA, latent variables are defined as weighted components of the observed variables, i.e.:

$$\xi_i = \mathbf{W}\mathbf{x}_i \quad (2)$$

where \mathbf{W} is the matrix containing the component weights.

Then, the equation (1) can be rewritten as:

$$\begin{bmatrix} \mathbf{I} \\ \mathbf{W} \end{bmatrix} \mathbf{x}_i = \begin{bmatrix} \mathbf{0} & \mathbf{\Lambda} \\ \mathbf{0} & \mathbf{B} \end{bmatrix} \begin{bmatrix} \mathbf{I} \\ \mathbf{W} \end{bmatrix} \mathbf{x}_i + \begin{bmatrix} \epsilon_i \\ \zeta_i \end{bmatrix} \quad (3)$$

where \mathbf{I} is an identity matrix.

Defining $\mathbf{A} = \begin{bmatrix} \mathbf{0} & \mathbf{\Lambda} \\ \mathbf{0} & \mathbf{B} \end{bmatrix}$, $\mathbf{r}_i = \begin{bmatrix} \epsilon_i \\ \zeta_i \end{bmatrix}$ and $\mathbf{u}_i = \begin{bmatrix} \mathbf{I} \\ \mathbf{W} \end{bmatrix} \mathbf{x}_i$, the last equation can be rewritten as:

$$\mathbf{u}_i = \mathbf{A}\mathbf{u}_i + \mathbf{r}_i. \quad (4)$$

As it is easy to notice in GSCA all the manifest variables, as well as all the latent variables, are included in the supervector \mathbf{u}_i of dimension $(J + Q)$, where Q is the total number of LVs in the model and J is the number of endogenous LVs in the structural model. Moreover all the model parameters (i.e. loadings and path coefficients) are included in the squared matrix \mathbf{A} of dimension $(J + Q)$. As the authors underlined “*differently from the PLS-PM, in GSCA the structural and the measurement models are not addressed separately, on the contrary they are combined in a unique algebraic formulation*” (Hwang & Takane, 2004). This allows the authors to identify a unique function to maximize.

Therefore, the parameters of GSCA (\mathbf{W} and \mathbf{A}) are estimated so that the sum of the squares of all residuals \mathbf{r}_i for the i -th unit is as small as possible. In other words, the following least-squares criterion is minimized:

$$\vartheta = \sum_{i=1}^n (\mathbf{u}_i - \mathbf{A}\mathbf{u}_i)' (\mathbf{u}_i - \mathbf{A}\mathbf{u}_i) \quad (5)$$

with respect to \mathbf{W} and \mathbf{A} and under the constraint that the latent variable scores are normalized, i.e.: $\sum_{i=1}^n \xi_{iq}^2 = 1$.

This is equivalent to minimizing:

$$\vartheta = \text{trace} [(\mathbf{U} - \mathbf{UA})'(\mathbf{U} - \mathbf{UA})] \quad (6)$$

An Alternating Least Squares (ALS) algorithm (De Leeuw *et al.*, 1976) is used so as to minimize equation (5). The convergence is assured since ALS monotonically decreases the value of the chosen criterion. Nevertheless, it is not assured that the convergence is reached in a global minimum. To overcome this problem different procedures are available: using “good” initial values or running the algorithm with different starting values. In particular, Hwang and Takane (2004) suggest using a Constrained Component Analysis to obtain “good” starting values for \mathbf{W} , and then simply obtain \mathbf{A} as least square estimate given \mathbf{W} .

PLS-PM vs. GSCA: comparisons and discussion

Recently, Hwang *et al.* (2010) presented a comparative study on the performance of GSCA, PLS-PM and *covariance-based* approach to SEM. In this study they have shown that the GSCA model is similar to the Reticular Action Model for covariance structure analysis by McArdle and McDonald (1984) with the main difference that in GSCA the LVs are obtained as a linear combination of their MVs. They tested the performance of PLS-PM, GSCA and the *covariance-based* approach to SEM under different simulation schemes, with conditions related to sample size, model specification and data distribution. Moreover, since the three approaches need different parameters to be estimated, they only paid attention to the estimated values of loadings and path coefficients. According to their results, Hwang *et al.* claim that GSCA should be preferred to PLS-PM when the considered model is correctly specified. Instead, both PLS-PM and GSCA show similar performances in the case of misspecified models. It is important to notice that the simulation scheme related to the correctly specified model hypothesis considers cross-loadings, i.e. assumes that a same MV is associated to more than one LV. In our opinion, Hwang *et al.* (2010) conclusions should be reviewed in compliance with the particular model chosen for the simulations. As a matter of fact, the presence of cross-loadings is not a standard case in PLS-PM applications.

Despite several studies have been performed to compare PLS-PM to *covariance-based* approaches to SEM, the Hwang *et al.* (2010) study is the first one that includes GSCA. Here, a new simulation study is presented in order to assess the performances of both PLS-PM and GSCA under conditions that are more suitable to the specific features of both methods. Namely, we refer to different conditions for model specification, number of MVs related to a LV, number of dimensions underlying a set of MVs, prediction relevance of these dimensions. Both methods will be tested on simulated data using a specifically developed R code. In accordance with the empirical findings of a study presented by

Tenenhaus (2008), GSCA seems to be more related to Principal Component Analysis of the sets of MVs rather than to structural equation modeling between the LVs. In other words, the measurement model seems to play a major role when estimating model parameters in GSCA. To conclude we show the links between GSCA and the Maximum Sum of Explained Variance method by Glang (1988) and how the Glang's method is equivalent to PLS as long as all regressions in the structural model are simple.

References

- [1] 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.
- [2] Glang, M. (1988) Maximierung der Summe erklärter Varianzen in linearrekursiven Strukturgleichungsmodellen mit multiple Indikatoren: Eine Alternative zum Schätzmodus B des Partial-Least-Squares-Verfahren, *Phd thesis*, Universität Hamburg, Hamburg, Germany.
- [3] Hwang H., Malhotra N. K., Kim Y., Tomiuk M. A., Hong S. (2010) A Comparative Study on Parameter Recovery of Three Approaches to Structural Equation Modeling, *Journal of Marketing Research*, 47 (4), 699-712.
- [4] Hwang H., Takane Y. (2004) Generalized structured component analysis, *Psychometrika*, 69, 81-99.
- [5] McArdle J. J., McDonald R. P. (1984) Some Algebraic Properties of the Reticular Action Model for Moment Structures, *British Journal of Mathematical and Statistical Psychology*, 37, 234-251.
- [6] Tenenhaus M, (2008) Component-based Structural Equation Modelling, *Total Quality Management & Business Excellence*, 19 (7), 871-86.
- [7] Tenenhaus M., Esposito Vinzi V., Chatelin Y.M., Lauro C. (2005) PLS path modeling, *Computational Statistics and Data Analysis*, 48, 159-205.
- [8] Wold H. (1975) Modelling in complex situations with soft information, in *Third World Congress of Econometric Society*, Toronto, Canada.