Springer Proceedings in Business and Economics

Lăcrămioara Radomir ·
Raluca Ciornea · Huiwen Wang ·
Yide Liu · Christian M. Ringle ·
Marko Sarstedt *Editors*

State of the Art in Partial Least Squares Structural Equation Modeling (PLS-SEM)

Methodological Extensions and Applications in the Social Sciences and Beyond

Springer

Springer Proceedings in Business and Economics

Springer Proceedings in Business and Economics brings the most current research presented at conferences and workshops to a global readership. The series features volumes (in electronic and print formats) of selected contributions from conferences in all areas of economics, business, management, and finance. In addition to an overall evaluation by the publisher of the topical interest, scientific quality, and timeliness of each volume, each contribution is refereed to standards comparable to those of leading journals, resulting in authoritative contributions to the respective fields. Springer's production and distribution infrastructure ensures rapid publication and wide circulation of the latest developments in the most compelling and promising areas of research today.

The editorial development of volumes may be managed using Springer Nature's innovative EquinOCS, a proven online conference proceedings submission, management and review system. This system is designed to ensure an efficient timeline for your publication, making Springer Proceedings in Business and Economics the premier series to publish your workshop or conference volume.

This book series is indexed in SCOPUS.

Lăcrămioara Radomir • Raluca Ciornea •
Huiwen Wang • Yide Liu • Christian M. Ringle •
Marko Sarstedt
Editors

State of the Art in Partial Least Squares Structural Equation Modeling (PLS-SEM)

Methodological Extensions and Applications in the Social Sciences and Beyond

Springer

Editors
Lăcrămioara Radomir
Faculty of Economics and Business Administration
Babeş-Bolyai University
Cluj-Napoca, Romania

Raluca Ciornea
Faculty of Economics and Business Administration
Babeş-Bolyai University
Cluj-Napoca, Romania

Huiwen Wang
School of Economics and Management
Beihang University
Beijing, China

Yide Liu
School of Business
Macau University of Science and Technology
Taipa, Macao

Christian M. Ringle
Department of Management Sciences and Technology
Hamburg University of Technology
Hamburg, Germany

Marko Sarstedt
Ludwig-Maximilians-University, Munich, Germany
Babeş-Bolyai University
Cluj-Napoca, Romania

ISSN 2198-7246 ISSN 2198-7254 (electronic)
Springer Proceedings in Business and Economics
ISBN 978-3-031-34588-3 ISBN 978-3-031-34589-0 (eBook)
https://doi.org/10.1007/978-3-031-34589-0

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

The use of partial least squares structural equation modeling (PLS-SEM) has gained enormous momentum in the past decade in various business research fields such as accounting, information systems, marketing, strategic management, tourism but also in other non-business disciplines such as computer sciences, engineering, environmental sciences, medical sciences, political sciences, and psychology. The widespread adoption of the method is attributed to the confluence of various factors. PLS-SEM facilitates the estimation of models with many constructs and complex inter-relationships, while also accommodating advanced modeling such as higher-order constructs, nonlinear relationships, and conditional process models. Other strengths of PLS-SEM include the ability to support model comparisons and to test the predictive power of models. Especially the predictive assessment of the results obtained by PLS-SEM allows researchers to substantiate their findings and managerial recommendations, which are predictive in nature.

Continued efforts by researchers to refine and improve the method have yielded rich resources (e.g., textbooks and edited volumes on the method, methodological articles, and review papers on the method's use), which have further fueled the method's dissemination by creating an understanding how PLS-SEM can support researchers in accomplishing the study goals. In addition, access to software with user-friendly graphical user interfaces and to freely available PLS-SEM packages in the R Statistical Environment have encouraged the use of the method among non-technical researchers and among the researchers who are mindful of costs.

This proceedings book includes a collection of manuscripts presented during the 2022 International Conference on Partial Least Squares Structural Equation Modeling Conference (PLS2022) that was held September 6–9, 2022 at the Faculty of Economics and Business Administration of the Babeș-Bolyai University in Cluj-Napoca, Romania. The conference has been designed to cater the needs of researchers and practitioners who empirically apply and methodologically advance the PLS-SEM method. It is part of a tradition of great conferences in the context of the PLS-SEM method such as the PLS2005 in Barcelona, Spain, the PLS2015 in Seville, Spain, and the PLS2017 in Macau, China. As with the previous conferences, the PLS2022 served as a vehicle to share and discuss new ideas, help each other, and explore new ideas in a friendly environment among friends.

Photograph by Mumtaz Ali Memon, NUST Business School (NBS), Pakistan

The diverse scientific inquiries addressed in this collection of manuscripts testify both the widespread adoption of the method and researchers' interest to further address methodological issues in PLS-SEM. This book also supplements the resources available for the PLS-SEM community, which, we hope, will expand knowledge and foster novel research in various fields of inquiry.

We would like to seize this opportunity to thank all the authors and the numerous colleagues who have devoted their time to reviewing the submissions, thereby helping further improving the manuscripts. Without your consolidated effort, this edited volume wouldn't have been possible. Thank you!

Cluj-Napoca, Romania	Lăcrămioara Radomir
Cluj-Napoca, Romania	Raluca Ciornea
Beijing, China	Huiwen Wang
Taipa, Macau	Yide Liu
Hamburg, Germany	Christian M. Ringle
Munich, Germany	Marko Sarstedt

Contents

Part IV Innovation Management and Entrepreneurship

Part V Human Resource Management and Organizational Design and Behavior

Part I
Methodology

Empirical Validation of the 10-Times Rule for SEM

Ralf Wagner and Malek Simon Grimm

1 Introduction

Structural equation modeling has become indispensable in scholarly empirical research due to its ability to investigate latent constructs. However, researchers always face the central dilemma of determining a sufficient sample size to ensure a converting model, unbiased estimates, and sufficient statistical power. Various rules of thumb have been established to determine the optimal sample size, such as a sample of at least 200 participants (Boomsma 1983), 5–10 observations per estimated parameter (Bentler and Chou 1987), and 10 observations per variable (Nunnally 1975). Monte Carlo simulation and the inverse square root method are alternative and more complex methods that require a deeper understanding of statistical power analysis (Ranatunga et al. 2020).

For decades, researchers have cited the 10-times rule as representing a sufficient sample size in structural equation models (SEMs) (Thompson et al. 1995). The 10-times rule suggests that the minimum sample size should be 10 times the maximum number of arrowheads pointing at a latent variable anywhere in the partial least squares (PLS) path model (Hair et al. 2021). According to Google Scholar (October 2022), Thompson et al.'s (1995) initial paper by has been cited over 7000 times. However, an empirical validation of this rule of thumb is still lacking. This paper takes a simulation approach to validate the 10-times rule. Only a few studies have investigated the sample size requirements for SEMs (cf. Wolf et al. 2013; Stone and Sobel 1990), but no extant study has specifically attempted to validate the prominent 10-times rule.

R. Wagner (✉) · M. S. Grimm
University of Kassel, Chair for Sustainable Marketing, Kassel, Germany
e-mail: rwagner@wirtschaft.uni-kassel.de; m.grimm@wirtschaft.uni-kassel.de

L. Radomir et al. (eds.), *State of the Art in Partial Least Squares Structural Equation Modeling (PLS-SEM)*, Springer Proceedings in Business and Economics,
https://doi.org/10.1007/978-3-031-34589-0_1

To pursue the research question, we drew upon the European Customer Satisfaction Index (ESCI) data set provided by SmartPLS 3.0 and the respective model, which contains 24 independent variables. We used a random algorithm to induce randomly missing cases (MCs) into the data set to decrease the sample size randomly. A 60-times trait ($N = 240$) was tested against the original trait ($N = 250$) and traits with 50 times ($N = 200$), 40 times ($N = 160$), 30 times ($N = 120$), 20 times ($N = 80$), and 10 times ($N = 40$). Each trait was tested by an aggregated simulation of 30 data sets per trait, so that each data set had a unique pattern of MCs. We tested model stability by investigating the model fit indices SRMR, d_ULS, d_G, X^2, and NFI. It was found that the 60-times trait was nearly as precise as the original trait, whereas the remaining traits suffered significantly in measurement quality.

2 Methodology

The simulation data were evaluated using SmartPLS 4.0 and the associated ESCI data set and model containing 24 independent variables. All the available model fit indices (SRMR, d_ULS, d_G, X^2, and NFI) as well as the adjusted R^2 values were investigated. To address the research question, we induced MCs into the original data set, which contained 250 cases without any missing values (MVs). The characteristics of the MCs can be described by the missing completely at random (MCAR) condition.

The MCAR condition requires that the probability of the MVs (and, in this case, MCs) in a given variable is unrelated to all other measures (Grimm and Wagner 2020; Parwoll and Wagner 2012). This condition was ensured by using a code that generated randomly MCs within the spectrum of N with respect to the desired number of MCs. Seven traits were investigated. To prevent occasionally occurring patterns, 30 data sets with unique MCs patterns were generated for each trait (except the original trait). Accordingly, the results depict the average of 30 data sets for the test traits.

3 Results

Table 1 illustrates the results of the model fit indices as well as the respective differences from the original trait; Table 2 illustrates the results of the adjusted R^2 values.

The diverse values shown in Tables 1 and 2 imply that the 60-times trait is the most stable in comparison to the original trait. Biases arise if the number of MCs increases. Because the adjusted R^2 will vary according to the variable and from survey to survey, no general conclusion can be derived with this assessment. Notably, the values are increasingly biased as the number of MCs increases. The

Table 1 Estimation results and differences in percentage of the model fit indices

Trait	Estimates	Model fit indices				
		SRMR	D_ULS	d_G	X^2	NFI
Original trait	Estimate	0.13	5.21	0.79	993.22	0.67
	Δ	0%	0%	0%	0%	0%
60× trait	Estimate	0.13	5.21	0.81	968.59	0.66
	Δ	1%	0%	2%	− 2%	−1%
50× trait	Estimate	0.13	5.36	0.88	875.43	0.65
	Δ	3%	3%	12%	−12%	−4%
40× trait	Estimate	0.13	5.44	0.98	772.31	0.62
	Δ	3%	4%	25%	−22%	−7%
30× trait	Estimate	0.13	5.37	1.17	672.99	0.6
	Δ	3%	3%	48%	−32%	−10%
20× trait	Estimate	0.14	6.33	1.67	597.36	0.52
	Δ	10%	21%	108%	−38%	−11%
10× trait	Estimate	0.17	8.67	3.88	546.42	0.40
	Δ	27%	62%	339%	−38%	−38%

Table 2 Estimation results and differences in percentage of the adjusted R^2 values

Trait	Estimates	R^2 Values					
		Complaints	Expectation	Loyalty	Quality	Satisfaction	Value
Original trait	Estimate	0.28	0.25	0.45	0.31	0.68	0.34
	Δ	0%	0%	0%	0%	0%	0%
60× trait	Estimate	0.27	0.25	0.68	0.44	0.31	0.34
	Δ	−3%	0%	50%	43%	−55%	−1%
50× trait	Estimate	0.28	0.26	0.68	0.44	0.31	0.34
	Δ	−1%	4%	51%	43%	−54%	1%
40× trait	Estimate	0.28	0.27	0.67	0.45	0.32	0.34
	Δ	−2%	7%	48%	44%	54%	1%
30× trait	Estimate	0.29	0.30	0.69	0.47	0.35	0.35
	Δ	2%	19%	52%	53%	−48%	4%
20× trait	Estimate	0.28	0.28	0.68	0.47	0.33	0.37
	Δ	11%	11%	0%	4%	10%	−4%
10× trait	Estimate	0.28	0.25	0.45	0.31	0.68	0.34
	Δ	0%	0%	0%	0%	0%	0%

results in Table 1 also imply that a bias is induced and increases when the number of MCs increases.

One-sample *t*-tests were computed to test the statistical significance with the original trait (Fig. 1), with each single trait being tested against the respective original trait. The *t*-tests were run with the original values of the respective model fit indices. To provide a better comparison, Fig. 1 includes percentage values that indicate the percentage of difference from the original value.

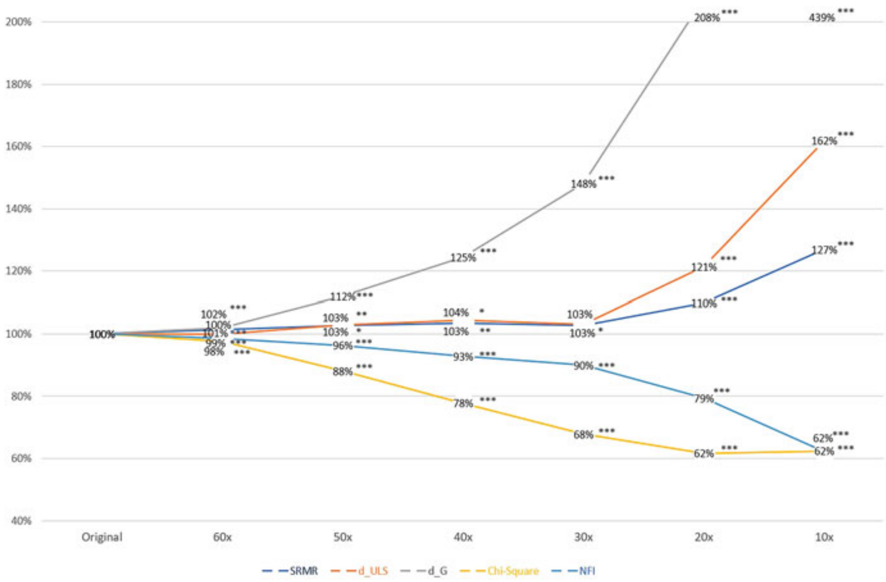

Fig. 1 Direct comparison of the model fit indices with statistical significance test. *Notes:* $^*p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$

The results suggest that d_G, NFI, and X^2 in particular are substantially and significantly biased and that d_ULS and SRMR are biased if the number of MCs increases. Overall, the 60-times trait is the most stable in comparison to the original trait, even though the models are slightly skewed with respect to the indices d_G, NFI, and X^2. Because the NFI measure is calculated on the basis of the X^2 value, the assessment of biases is related within that measure. The most relevant characteristic is the elbow that becomes apparent at the 30-times trait.

4 Conclusion

Challenging the common assumption that the PLS algorithm is robust against MVs and small sample sizes, this simulation study yields important implications for the theory and practice of recommending a minimum sample size. The results indicate that the 10-times rule is a misleading heuristic. With a decrease in sample size, the likelihood of biases increases. The 60-times trait performed best in this study and in comparison to the original trait. An elbow in the bias plot suggests that researchers are well advised to use at least a 30-times trait. Further research should adopt different models as test instances and should address the facets of statistical power (inverse square root of error variance) in relation to the fit statistics (heteroskedasticity), an overfitting assessment (outer sample validation routine), endogeneity tests, and non-MCAR MVs.

References

Bentler PM, Chou C-P (1987) Practical issues in structural modeling. Sociol Methods Res 16(1): 78–117

Boomsma A (1983) On the robustness of LISREL (maximum likelihood estimation) against small sample size and non-normality. Sociometric Research Foundation, Amsterdam

Google Scholar (2022) https://scholar.google.com/ Accessed Oct 2022

Grimm MS, Wagner R (2020) The impact of missing values on PLS, ML and FIML model fit. Advance online publication. Arch Data Sci Ser A 6(1):1–17

Hair JF, Hult GTM, Ringle CM, Sarstedt M, Danks NP, Ray S (2021) Partial least squares structural equation modeling (PLS-SEM) using R. Springer, Cham

Nunnally JC (1975) Psychometric theory, 11th edn. McGraw-Hill, New York

Parwoll M, Wagner R (2012) The impact of missing values on PLS model fitting. In: Gaul WA, Geyer-Schulz A, Schmidt-Thieme L, Kunze J (eds) Challenges at the interface of data analysis, computer science, and optimization. Springer, Berlin, pp 537–544

Ranatunga RVSPK, Priyanath HMS, Megama RGN (2020) Methods and rule-of-thumbs in the determination of minimum sample size when applying structural equation modelling: a review. J Social Sci Res 15:102–109

Stone CA, Sobel ME (1990) The robustness of estimates of total indirect effects in covariance structure models estimated by maximum. Psychometrika 55(2):337–352

Thompson R, Barclay D, Higgins CA (1995) The partial least squares approach to causal modeling: personal computer adoption and use as an illustration. Technology Studies: Special Issues on Research Methodology 2:284–324

Wolf EJ, Harrington KM, Clark SL, Miller MW (2013) Sample size requirements for structural equation models: An evaluation of power, bias, and solution propriety. Educ Psychol Meas 76(6):913–934

Missing Values in RGCCA: Algorithms and Comparisons

Caroline Peltier, Laurent Le Brusquet, François-Xavier Lejeune, Ivan Moszer, and Arthur Tenenhaus

C. Peltier (✉)
iCONICS core facility, Institut du Cerveau et de la Moelle Epinière, Inserm U1127, CNRS UMR 7225, Sorbonne Université, Paris, France

Centre des Sciences du Goût et de l'Alimentation, CNRS, INRAE, Institut Agro, University of Bourgogne Franche-Comté, Dijon, France

CNRS, INRAE, PROBE Research Infrastructure, ChemoSens Facility, Dijon, France

INRAE, Dijon, France
e-mail: caroline.peltier@inrae.fr

L. Le Brusquet
Laboratoire des Signaux et Systèmes (L2S), CentraleSupélec, Université Paris-Saclay, Gif-Sur-Yvette, France

CentraleSupélec, Gif-sur-Yvette, France
e-mail: laurent.lebrusquet@l2s.centralesupelec.fr

F.-X. Lejeune · I. Moszer
iCONICS core facility, Institut du Cerveau et de la Moelle Epinière, Inserm U1127, CNRS UMR 7225, Sorbonne Université, Paris, France

Institut du Cerveau, Hôpital Pitié, Paris, France
e-mail: f-x.lejeune@icm-institute.org

A. Tenenhaus
iCONICS core facility, Institut du Cerveau et de la Moelle Epinière, Inserm U1127, CNRS UMR 7225, Sorbonne Université, Paris, France

Laboratoire des Signaux et Systèmes (L2S), CentraleSupélec, Université Paris-Saclay, Gif-Sur-Yvette, France

CentraleSupélec, Gif-sur-Yvette, France
e-mail: arthur.tenenhaus@centralesupelec.fr

L. Radomir et al. (eds.), *State of the Art in Partial Least Squares Structural Equation Modeling (PLS-SEM)*, Springer Proceedings in Business and Economics,
https://doi.org/10.1007/978-3-031-34589-0_2

1 Introduction

Multidisciplinary approaches are now common in scientific research and provide multiple and heterogeneous sources of measures of a given phenomenon. These sources or blocks can be viewed as a collection of interconnected datasets, and dedicated algorithms are mandatory for providing relevant information from multiblock data.

Regularized generalized canonical correlation analysis (RGCCA) is a general statistical framework for multiblock data analysis which gathers many multiblock methods (PCA, PLS, CCA, consensus PCA, PLS-PM, etc.) through a single and very simple iterative algorithm. The global convergence of the algorithm was demonstrated in Tenenhaus et al. (2017).

However, multiblock data often have missing structure, i.e., data in one or more blocks may be completely unobserved for a sample (block-wise structure) or partially unobserved (random structure). The probability to observe missing data increases with the number of blocks. It is therefore mandatory to properly handle these missing structures within the framework of RGCCA.

In this work, several solutions were investigated and compared on simulations. An R package, "RGCCA," implementing all the methods is under development and is available on GitHub (https://github.com/rgcca-factory/RGCCA).

2 Background

2.1 Regularized Generalized Canonical Correlation Analysis

Let $\mathbf{X}_1, \ldots, \mathbf{X}_J$ be J blocks of variables, each block representing a set of p_j variables observed on n individuals. The number and the nature of the variables usually differ from one block to another, but the individuals must be the same across blocks. RGCCA is based on the following optimization problem (Tenenhaus et al. 2017).

$$\text{Maximize} \sum\nolimits_{j,k=1}^{J} c_{jk} g\left(\operatorname{cov}\left(\mathbf{X_j} \cdot \mathbf{w}_j, \mathbf{X}_k \cdot \mathbf{w}_k\right)\right) \text{s.t.} \mathbf{w}_\mathbf{j}^\mathbf{t} M_j \mathbf{w}_j = 1, \forall j \in \{1, \ldots, J\}$$

where $g(\mathrm{x})$ is any continuously differentiable convex function.

The connection matrix $\mathbf{C} = (c_{\mathrm{jk}})$ is a symmetric $J{*}J$ matrix of nonnegative elements describing the network of connections between blocks that the user wants to take into account. Usually, $c_{\mathrm{jk}} = 1$ for two connected blocks and 0 otherwise.

$\mathbf{M}_\mathrm{j}$ is a positive definite matrix, such as $\mathbf{M}_\mathrm{j} = \tau_\mathrm{j}\mathbf{I} + \left(1 - \tau_\mathrm{j}\right)\mathbf{X}_\mathrm{j}^\mathrm{t}\mathbf{X}_\mathrm{j}$ with $0 \leq \tau_\mathrm{j} \leq 1$. In the context of RGCCA, τ_j is called shrinkage constant. τ_j varies between 0 and 1 and interpolates smoothly between maximizing the covariance and maximizing the correlation.

In this work, we focused on the situation where $\tau_j = 1$ and all blocks are connected or connected to a superblock, defined as the concatenation of the individuals blocks.

2.2 *Missing Data Literature*

Pigott (2001) reviewed different types of strategies to deal with missing data: complete data technique, technique based on available data, imputation methods (including expectation-maximization algorithms), and multi-imputation methods.

2.3 *Methodology*

2.3.1 Complete Data Technique (Complete)

The complete approach consists in computing RGCCA only on the complete individuals. This is the simplest method and is the only one available until now.

2.3.2 Technique Based on Available Data (Passive)

This approach (the so-called passive) follows the approach presented in PLS-PM (Tenenhaus et al. 2005). RGCCA is closely related to the nonlinear estimation by iterative partial least squares algorithm (NIPALS; Wold 1966) as the components, and the axes are estimated alternatively. As in Tenenhaus et al. (2005), means and standard deviations of the variables are computed on all the available data; covariance matrices are computed using all the pairwise available data. This pairwise deletion procedure shows the drawback of possibly computing covariances of different sample sizes and/or different individuals.

2.3.3 Imputation Method (Iterative)

The imputation method (Fig. 1) is iterative and based on the alternated two steps (i) calculating the RGCCA weight vectors $\mathbf{w}_j$ and components $\mathbf{y}_j = \mathbf{X}_j\mathbf{w}_j$ and (ii) imputation of the missing values. This imputation was based on the hypothesis that each column of $\mathbf{X}_j$ can be estimated by a vector proportional to the RGCCA block component $\mathbf{y}_j$. This estimation is especially relevant when each block is unidimensional. Thus, the used reconstruction formulae for imputation was $\mathbf{X}_j = \mathbf{y}_j \gamma_j^T$ where $\boldsymbol{\gamma}_\mathbf{j}\, j$ is a column vector of size p_j containing the regression coefficients of $\mathbf{y}_j$ in the regression of $\mathbf{X}_j$ on $\mathbf{y}_j$.

Algorithm 1 Iterative algorithm for RGCCA with missing values

Consider $\mathbf{K}_j$ the $n \times p_j$ matrix such as $\mathbf{K}_j(l,k) = 0$ if is $\mathbf{X}_j(l,k)$ is missing and 1 otherwise

Initialization : $\mathbf{X}_1^0 = \mathbf{X}_1, \ldots, \mathbf{X}_J^0 = \mathbf{X}_J$ where the missing values are replaced by the colmeans

for s=0, 1 until convergence, **do**

Compute RGCCA on $\mathbf{X}_1^s, \ldots, \mathbf{X}_J^s$ and return $\mathbf{y}_1^s, \ldots, \mathbf{y}_J^s$.
This step includes the centering and scaling of each $\mathbf{X}_j^s$, giving $\tilde{\mathbf{X}}_j^s = (\mathbf{X}_j^s - \mathbf{M}_j^s)/\mathbf{S}_j^s$ where $\mathbf{M}_j^s$ and $\mathbf{S}_j^s$ are $n \times p_j$ matrices containing means and standard deviations of $\mathbf{X}_j^s$ and / is the division term by term

for $j \in [1:J]$ **do**

$$\mathbf{X}_j^{s+1} = \mathbf{K}_j \star \mathbf{X}_j^0 + (\mathbf{1}_n \mathbf{1}_{p_j}^\top - \mathbf{K}_j) \star (\mathbf{M}_j^s + \mathbf{S}_j^s \star [\mathbf{y}_j^s \gamma_j^{s\top}]))$$

where $\mathbf{y}_j^s$ is the block component associated with $\mathbf{X}_j^s$, γ_j^s contains regression coefficients of $\mathbf{y}_j^s$ in the regression of $\tilde{\mathbf{X}}_j^s$ on $\mathbf{y}_j^s$ and $\star$ is the Hadamard product

end for

end for

Fig. 1 Iterative algorithm for RGCCA with missing values

This method is directly inspired by the fixed effect model used in Josse and Husson (2012) for PCA with missing data. However, when the missing structure is blockwise, this approach gives the same results for all missing individuals, even if they have very different (non-missing) values in the other blocks.

To overcome this issue, we propose to use a "superblock" strategy. In the multiblock literature, a superblock $\mathbf{X}_{J+1} = [\mathbf{X}_1, \ldots, \mathbf{X}_J]$ is defined as the concatenation of all blocks. In this framework, each block is connected to the superblock (that is, $c_{j,\,J+1} = 1$ for $j = 1, \ldots, J$ and 0 otherwise), and optimization problem (1) reduces to

$$\text{Maximize} \sum\nolimits_{j=1}^{J} g\left(\text{cov}\left(\mathbf{X_j w}_\text{j}, \mathbf{X}_{\text{J}+1}\mathbf{w}_{\text{J}+1}\right)\right) \text{s.t.} \mathbf{w}_j^t \mathbf{M_J w}_\text{j} = 1, \forall j \in \{1, \ldots, J+1\}$$

This optimization problem is fully described in Tenenhaus et al. (2017). When a superblock $\mathbf{X}_{J+1}$ is considered, the imputation is obtained by using the following reconstruction formulae $\mathbf{X}_{J+1} = \mathbf{y}_{J+1} \boldsymbol{\gamma}_{J+1}^T$, where $\boldsymbol{\gamma}_{J+1}$ is a column vector of size $p = \Sigma p_j$ containing the regression coefficients of $\mathbf{y}_{J+1}$ in the regression of $\mathbf{X}_{J+1}$ on $\mathbf{y}_{J+1}$.with an algorithm similar to Algorithm 1.

2.3.4 Dataset Presentation

These methods were tested on the Russett dataset, available within the RGCCA package. The Russett dataset (Russett 1964) are studied in Gifi (1990) and aims to study the relationships between agricultural inequality, industrial development, and political instability. Three blocks of variables were defined accordingly (3, 2, and 2 variables for agriculture, industry, and political, respectively) for 47 countries.

2.3.5 Comparison Method

From the Russett dataset, 20 datasets with 5, 10, 15, 20, and 25% of random missing datasets were simulated. For each block, the norm of the difference between the first weight vector obtained from the full case $\mathbf{w}_j$ and the one from the missing case $\mathbf{w}_j$ is calculated. Consequently, the closer to 0 this norm is, the better the method is. We used this method to compare complete, available, iterative, and superblock approach.

3 Results and Discussion

From our simulations, it appears that the iterative algorithm (with and without superblock) converges monotonically. Furthermore, the simulations showed that the implemented methods outperformed the complete approach on Russett data (Fig. 2) especially for superblock or iterative method.

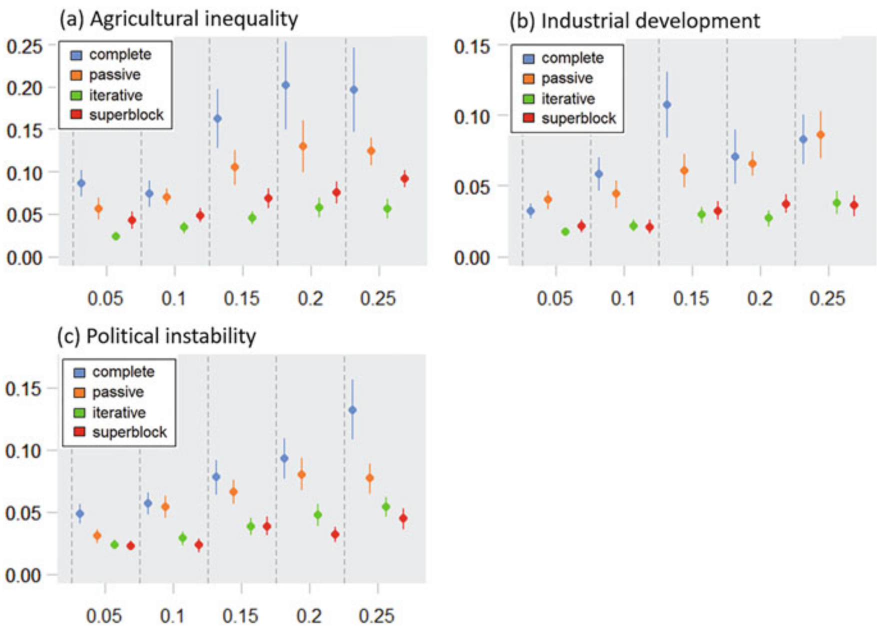

Fig. 2 The norm of the difference (mean and standard error) between the first axis of RGCCA based on the complete dataset and the first axis of the different “missing” methods according to the proportion of missing values in the dataset

4 Conclusions and Implications for Theory and Practice

This work presents different methods for taking missing values into account in RGCCA. The iterative method presented in this chapter gave good results on the Russett dataset. The convergence of this algorithm was observed but is still to be studied. Multiple imputations for visualizing the variability induced by the imputation could be implemented according to the model used. Furthermore, the strong hypothesis of unidimensionality of blocks used in this chapter could be relaxed by considering more than one component per block, including deflation steps in the iterative algorithm. Finally, this work was illustrated on the Russett dataset with a small number of variables, blocks, and samples, but aims to be also tested on more datasets.

References

Gifi A (1990) Nonlinear multivariate analysis. Wiley, Chichester

Josse J, Husson F (2012) Handling missing values in exploratory multivariate data analysis methods. Journal de la Société Française de Statistique 153(2):79–99

Pigott TD (2001) A review of methods for missing data. Educ Res Eval 7(4):353–383

Russett BM (1964) Inequality and instability: The relation of land tenure to politics. World Polit 16(3):442–454

Tenenhaus M, Esposito Vinzi V, Chatelin YM, Lauro C (2005) PLS path modeling. Comput Stat Data Anal 48(1):159–205

Tenenhaus M, Tenenhaus A, Groenen PJF (2017) Regularized generalized canonical correlation analysis: a framework for sequential multiblock component methods. Psychometrika 82(3): 737–777

Wold H (1966) Estimation of principal components and related models by iterative least squares. In: Krishnajah PR (ed) Multivariate analysis. Academic Press, New York, pp 391–420

https://github.com/rgcca-factory/RGCCA

Comparing Local vs Global Clustering with FIMIX-PLS: Application to Marketing

Sophie Dominique, Mohamed Hanafi, Fabien Llobell, Jean-Marc Ferrandi, and Véronique Cariou

1 Introduction

The use of Partial Least Squares Structural Equation Modeling (PLS-SEM) to estimate model parameters with interrelationships between observed concepts and their corresponding latent variables has gained popularity in marketing research since past decades (Sarstedt et al. 2022). When applied to a set of observations, the determination of a single set of parameters, encompassing the latent variables scores and their associated path coefficients estimations, implicitly assumes that heterogeneity is negligible. In practice, this assumption is not realistic in the social sciences, and there is generally unobserved heterogeneity that can taint the validity of PLS-SEM results. To uncover this potential heterogeneity in the data, one can apply a clustering strategy that aims to identify homogeneous segments of observations sharing the same pattern relationships (i.e., the same path coefficients within a segment).

S. Dominique (✉)
Oniris, INRAE, STATSC, Nantes, France

Lumivero, XLSTAT, Paris, France
e-mail: sdominique@lumivero.com

M. Hanafi · V. Cariou
Oniris, INRAE, STATSC, Nantes, France
e-mail: mohamed.hanafi@oniris-nantes.fr; veronique.cariou@oniris-nantes.fr

F. Llobell
Lumivero, XLSTAT, Paris, France
e-mail: fllobell@lumivero.com

J.-M. Ferrandi
LEMNA, Oniris, Nantes, France
e-mail: jean-marc.ferrandi@oniris-nantes.fr

L. Radomir et al. (eds.), *State of the Art in Partial Least Squares Structural Equation Modeling (PLS-SEM)*, Springer Proceedings in Business and Economics,
https://doi.org/10.1007/978-3-031-34589-0_3

For the last two decades, several clustering techniques have been developed within the PLS-SEM framework, such as FIMIX-PLS (Hahn et al. 2002; Ringle et al. 2010), REBUS (Esposito Vinzi et al. 2008), PLS-POS (Becker et al. 2013), GAS (Ringle et al. 2014), PLS-IRRS (Schlittgen et al. 2016), or more recently PLS-SEM KMEANS (Fordellone and Vichi 2020). Undoubtedly, the finite mixture PLS method (FIMIX-PLS) remains the most popular of them. In FIMIX-PLS, the measurement model is considered common to all observations, while observations can differ from each other only at the structural level. This approach is well suited to the context of social science because it leads to the rational assumption that the measurement scales are identical from one segment to another and observation to observation.

Going further, in some situations, heterogeneity can be concentrated only in a part of the structural model and not affect the entire set of path coefficients parameters. This leads to determining part of the path coefficients as common to all the observations and the other part as parameters reflecting heterogeneity, which depends on the observation segments. A marketing expert's knowledge is required to identify the constant and variable parts of the model for all observations. Our work aims to adapt FIMIX-PLS procedure so that the segmentation is only performed on a subpart of the structural model. Subsequently, local partitioning is introduced as a moderating variable in the PLS-SEM applied to all the observations. We advocate such a rationale to provide clusters, which aim to be more stable and interpretable. This strategy is illustrated using a case study pertaining to marketing. Local vs global partitioning with FIMIX-PLS are compared both in terms of interpretability and model quality.

2 Methodology

This analysis was carried out on a sample of 315 French people who are members of a CSA (Community Supported Agriculture) in Nantes in 2011 (Dufeu and Ferrandi 2013). CSA belongs to the family of short supply chains, with products sold directly from producers to consumers. Their members are generally opposed to other modes of retailing, particularly mass distribution. Mainly based on the principle of mutual commitment of both parts, CSA seeks to create strong links and proximity between producers and consumers.

The objective of this study on the Nantes CSAs was to model the link between the proximity perceived by members toward their CSA and their trust, satisfaction, and commitment in this CSA that are the key variables of relational marketing.

In marketing the concept of proximity is composed of several dimensions. Beyond the geographical proximity, the direct and repeated exchanges between all the actors (relational proximity), the sharing of values (identity proximity), and the knowledge on the production and distribution process (process proximity) are the dimensions evaluated in our work (Bergadaà and Del Bucchia 2009). They can all constitute antecedents of trust, satisfaction, and loyalty. Figure 1 represents the path

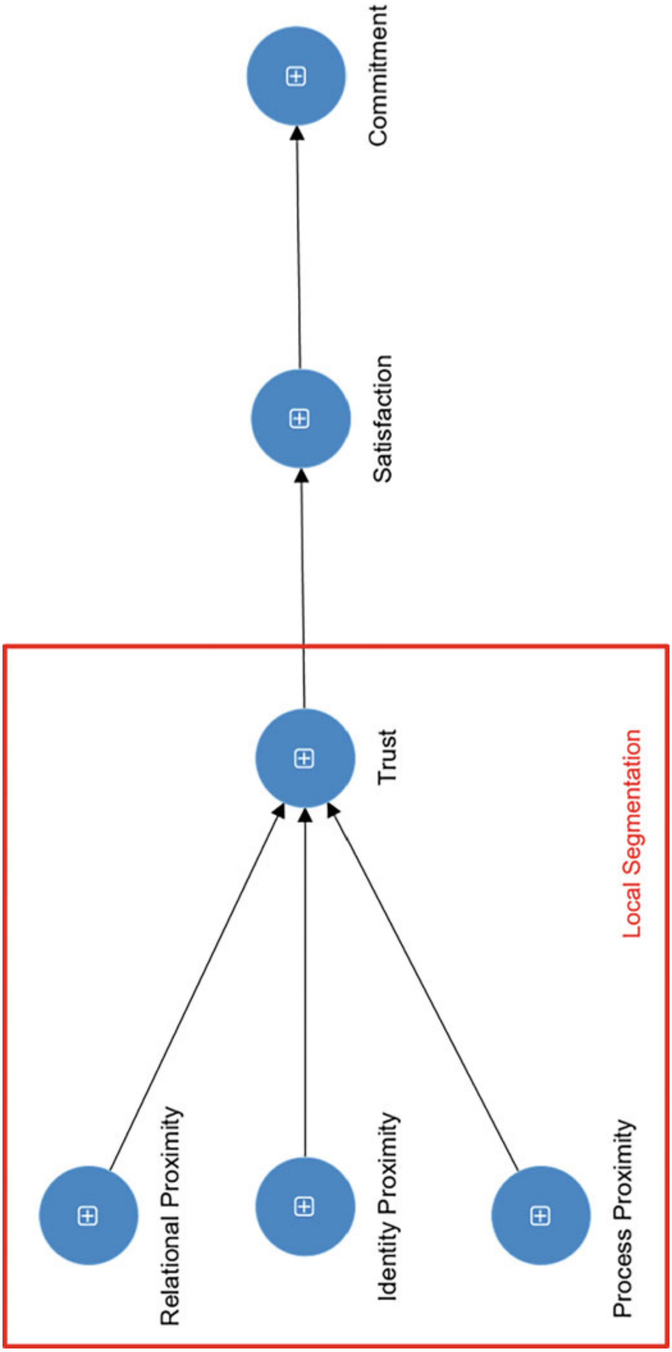

Fig. 1 Studied model

diagram associated with the six latent concepts detailed above. Trust, satisfaction, and commitment are the founding concepts of relational marketing and are therefore common to all segments that can be defined.

The first subpart of the model was chosen to segment the respondents. Indeed, proximity is considered here as an antecedent of the relationship marketing chain; thus, we assume that heterogeneity lies mainly on the relationships between proximity and trust. A first partitioning algorithm was carried out on this subpart of the model leading to a so-called local partition.

In parallel, a global partition was determined on the basis of the complete structural model. To compare the two strategies, parameters associated with the complete structural model were estimated independently for each segment of the local partition. Finally, the two partitions obtained (local vs global) were compared in terms of path coefficients and variance accounted for (R^2). Both partitions were obtained with FIMIX-PLS algorithm implemented in the SmartPLS software.

3 Results and Discussion

The partition obtained from the sub-model consisted of two segments. In contrast, the segmentation on the complete model led to retain identified three segments of observations. This number of segments was chosen with the BIC and CAIC indices (Sarstedt et al. 2011).

3.1 Global Partition

The global partition consists of one main segment ($N_1 = 201$) and two smaller ones ($N_2 = 80$ and $N_3 = 34$). Table 1 presents the results of the model associated with each cluster.

The first segment is characterized by a valuation of process and identity proximities (path coefficient, respectively, equals to 0.630 and 0.252). By feeling close to these dimensions, consumers are confident in their CSA which in turn leads to increased satisfaction and commitment. The R^2 values for each latent variable are greater than 0.6, indicating that they are well explained by the variables directly related to them and therefore that the structure model associated with segment 1 performs well.

Only the identity dimension of proximity explains the trust of segment 2 consumers (path coefficient = 0.347). Notwithstanding, proximity alone explains the trust dimension poorly ($R^2 = 0.147$).

Finally, the last segment is characterized by a strong valuation of identity proximity associated with a lower valuation of relational proximity. Moreover, the R^2 values are very close to 1; this indicates that the model is fully explained by the relationships defined between the latent variables, which is difficult to achieve in

Table 1 Results of the global partition models

	Segment 1 ($N = 201$)	Segment 2 ($N = 80$)	Segment 3 ($N = 34$)
	Path coefficient		
Relational prox. → Trust	0.074	−0.149	0.130*
Identity prox. → Trust	0.252***	0.347*	0.914***
Process prox. → Trust	0.630***	−0.055	−0.012
Trust → Satisfaction	0.911***	0.648***	0.957***
Satisfaction → Commitment	0.811***	0.798***	0.999***
	R^2		
Trust	0.774	0.147	0.919
Satisfaction	0.831	0.420	0.916
Commitment	0.658	0.636	0.998

Notes: *p-value < 0.05; **p-value < 0.01; ***p-value < 0.001

practice. Moreover, one emphasizes the small size of the segment ($N_3 = 34$). Focusing on the optimization of the structural model, the FIMIX-PLS strategy has led to a partition with a very specific cluster corresponding to an overfitted model with R^2 close to 1.

3.2 Local Partition

The local partitioning strategy, that is, to say on the first part of the structural model as shown in Fig. 1, led to consider a partition with two clusters whose respective size are $N_1 = 228$ and $N_2 = 87$. The results associated with the two models are figured out in Table 2.

In the largest segment, all three dimensions of proximity are antecedents to the concept of trust, with the proximity of the process being particularly important to the

Table 2 Results of the local partition models

	Segment 1 ($N = 228$)	Segment 2 ($N = 87$)
	Path coefficient	
Relational prox. → Trust	0.154***	−0.245*
Identity prox. → Trust	0.234***	0.425***
Process prox. → Trust	0.634***	−0.204*
Trust → Satisfaction	0.869***	0.746***
Satisfaction → Commitment	0.844***	0.804***
	R^2	
Trust	0.857	0.250
Satisfaction	0.755	0.556
Commitment	0.712	0.646

Notes: *p-value < 0.05; **p-value < 0.01; ***p-value < 0.001

trust that consumers place in their CSA (path coefficient = 0.634). The quality of the structural model related to this segment is satisfactory, with R^2 values greater than 0.7.

The smallest segment values only identity proximity (path coefficient = 0.425). The negative sign of the path coefficients linking process and relational proximity to trust may reveal that consumers have more trust if they share the company's identity values, but less if the process and relational proximity dimensions are present.

3.3 Local Partition and Global Partition

By comparing the sizes of both partitions, in Table 3, we notice that segments 1 and 2 of each partition group have the same observations. Moreover, the Adjusted Rand Indices (ARI) and Normalized Mutual Information (NMI) are satisfactory, and thus the two partitions are close (ARI = 0.32, NMI = 0.23).

From Table 3, it can also be seen that the third segment obtained in the global partition is included into the first segment of the local partition. The average R^2 of each segment of the local partition is greater than that obtained with the global partition (segment 1: $\overline{R^2_{\text{local1}}} = 0.775$ vs $\overline{R^2_{\text{global1}}} = 0.754$; segment 2: $\overline{R^2_{\text{local2}}} = 0.484$ vs $\overline{R^2_{\text{global2}}} = 0.401$). Furthermore, we achieve a larger value for the local partition by computing the average of the R^2 weighted by the number of observations on all clusters of each partition (weighted $\overline{R^2_{\text{local}}} = 0.694$ vs weighted $\overline{R^2_{\text{global}}} = 0.685$). Therefore, the local partition that groups segment 3 into segment 1 performs better.

Thus, this tends to confirm our hypothesis that this third cluster is probably an artifact. The strategy of performing a partition on a subpart of the structural model improves both the quality of the segmentation and its interpretability.

To conclude, local partitioning aims at bringing a prior knowledge on the source of the heterogeneity by imposing constraints on the structure for the determination of the path coefficients. These constraints consist in imposing some of the path coefficients to be common for all segments. The underlying constraints aim at (1) improving the quality of the partition and (2) facilitating the interpretation of the clusters obtained.

Future works are needed to directly take into account this a priori knowledge to be integrated as constraints within the FIMIX-PLS clustering criterion.

Table 3 Cross-tabulation of the size of the two partitions

Global partition / Local partition	Segment 1	Segment 2	Segment 3	Total
Segment 1	177	21	30	228
Segment 2	24	59	4	87
Total	201	80	34	315

References

Becker J-M, Rai A, Ringle CM, Völckner F (2013) Discovering unobserved heterogeneity in structural equation models to avert validity threats. MIS Q 37(3):665–694

Bergadaà M, Del Bucchia C (2009) La recherche de proximité par le client dans le secteur de la grande consommation alimentaire. Revue Management et Avenir 21:121–135

Dufeu I, Ferrandi J-M (2013) Les ressorts de l'engagement des consommateurs dans une forme particulière de consommation collaborative: les AMAP. Décis Market 72:157–178

Esposito Vinzi V, Trinchera L, Squillacciotti S, Tenenhaus M (2008) REBUS-PLS: a response-based procedure for detecting unit segments in PLS path modelling. Appl Stoch Model Bus Ind 24(5):439–458

Fordellone M, Vichi M (2020) Finding groups in structural equation modeling through the partial least squares algorithm. Comput Stat Data Anal 147:106957

Hahn C, Johnson MD, Herrmann A, Huber F (2002) Capturing customer heterogeneity using a finite mixture PLS approach. Schmalenbach Bus Rev 54(3):243–269

Ringle CM, Sarstedt M, Mooi EA (2010) Response-based segmentation using finite mixture partial least squares. Theoretical foundations and an application to American customer satisfaction index data. In: Stahlbock R, Crone S, Lessmann S (eds) Data mining. Annals of information systems, vol 8. Springer, Boston, pp 19–49

Ringle CM, Sarstedt M, Schlittgen R (2014) Genetic algorithm segmentation in partial least squares structural equation modeling. OR Spectr 36(1):251–276

Sarstedt M, Becker J-M, Ringle CM, Schwaiger M (2011) Uncovering and treating unobserved heterogeneity with FIMIX-PLS: which model selection criterion provides an appropriate number of segments? Schmalenbach Bus Rev 63(1):34–62

Sarstedt M, Hair JF, Pick M, Liengaard BD, Radomir L, Ringle CM (2022) Progress in partial least squares structural equation modeling use in marketing research in the last decade. Psychol Mark 39(5):1035–1064

Schlittgen R, Ringle CM, Sarstedt M, Becker J-M (2016) Segmentation of PLS path models by iterative reweighted regressions. J Bus Res 69(10):4583–4592

Partial Least Squares Structural Equation Modeling-Based Discrete Choice Modeling: An Illustration in Modeling Hospital Choice with Latent Class Segmentation

Andreas Fischer, Marcel Lichters, and Siegfried P. Gudergan

1 Introduction

Understanding and predicting individual choices—such as those between different means of transport, hospitals, or retailers—is extremely important in business research (Louviere et al. 2008). One commonly used method to explain and predict consumer choices is discrete choice modeling (DCM), first introduced by Luce and Tukey (1964), formalized by McFadden (1974), and introduced to marketing as choice-based conjoint (CBC) analysis by Louviere and Woodworth (1983). This method allows researchers to estimate the relative likelihood of choosing one option from a set of alternatives. This allows for estimating the utility values (i.e., preference of an attribute level over others, usually zero-centered within each attribute) and determining every attribute's relative importance weight (i.e., the importance of every attribute as a whole in the consumer's decision process) (Train 2009).

A. Fischer (✉)
Institute of Human Resource Management and Organizations, Hamburg University of Technology, Hamburg, Germany
e-mail: andreas.fischer@tuhh.de

M. Lichters
Faculty of Economics and Business Administration, Chemnitz University of Technology, Chemnitz, Germany
e-mail: marcel.lichters@wiwi.tu-chemnitz.de

S. P. Gudergan
College of Business, Law & Governance, James Cook University, Douglas, QLD, Australia

Department of Global Business and Trade, Vienna University of Economics and Business, Vienna, Austria

Aalto University School of Business, Aalto University, Helsinki, Finland
e-mail: siggi.gudergan@jcu.edu.au

L. Radomir et al. (eds.), *State of the Art in Partial Least Squares Structural Equation Modeling (PLS-SEM)*, Springer Proceedings in Business and Economics,
https://doi.org/10.1007/978-3-031-34589-0_4

However, since DCM provides results on an aggregated data level, researchers often use the latent class analysis (LCA) introduced by Greene and Hensher (2003) or hierarchical Bayes analysis (Lenk et al. 1996) to uncover distinct segments of individuals in the dataset. As an advantageous alternative, Hair et al. (2019a) suggest using partial least squares structural equation modeling (PLS-SEM) (Hair et al. 2016a, 2019b) for the estimation of individuals' preference functions drawing on discrete choice experiment (DCE) data. This method permits estimating the coefficients and the importance of both the attribute levels but also the entire attributes. In this research, we substantiate that LCA using PLS-SEM produces similar results compared to those generated using conventional DCM. An advantage of using PLS-SEM draws on its capability to reveal segments in consideration of the relationships between attributes as a whole (rather than attribute levels) and the choice variable. Thereby, the segmentation approach allows for uncovering segments with segment-specific differences related to the entire attribute rather than certain attribute levels when using conventional DCM. Specifying segments based on whole attributes is beneficial for describing and characterizing segments, interpreting results, and drawing conclusions.

2 Methodology

Our illustrative application of PLS-SEM in LCA utilizes DCE data from the healthcare sector. The model focuses on the choice of patience between different hospitals in Germany. Schuldt et al. (2017) estimated patient choices concerning hospitals, using a sample of 590 randomly selected participants in three different German cities of the federal state Saxony-Anhalt, which responded to the DCE study in a "paper-and-pen" questionnaire. Within this study, each participant provided answers to eight choice tasks, comprising two options each. Four attributes described each option (distance to hospital, information, number of treatments, and complication rate). Each of these attributes had two levels (distance 1 km or 20 km to hospital, high or low level of information for treatment, high or low number of treatments per year, high or low rate of complications). The authors used Sawtooth Software Lighthouse Studio Sawtooth Software (2019) to create individualized choice designs according to a balanced overlap strategy (Johnson et al. 2013; Orme and Chrzan 2017).

3 Results

For the DCM analysis, we revert to the results presented by Schuldt et al. (2017). These authors used a mixed effect logit model with a random effects intercept to estimate attribute level utilities and the relative importance of each attribute. The PLS-SEM results reported here draw on the use of SmartPLS 3 (Ringle et al. 2015).

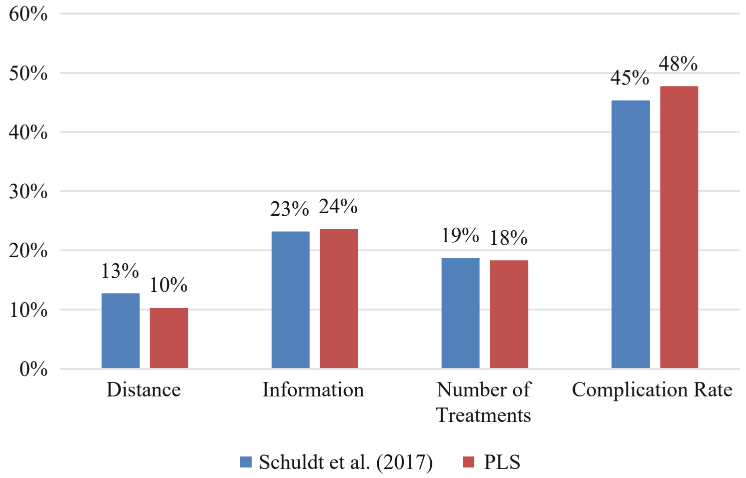

Fig. 1 Comparison of importance weights

Table 1 Part worth utilities and segment sizes PLS-SEM

	Attribute	Attribute level	1	2	3	4
Part worth utilities	Distance	1 km	0.054	0.182	0.019	0.00
		20 km	−0.054	−0.182	−0.019	0.00
	Information on treatment	Brief	0.151	0.161	0.642	0.00
		Detailed	−0.151	−0.161	−0.642	0.00
	Number of treatments	More than 100 times	0.127	0.254	0.106	0.00
		Less than 50 times	−0.127	−0.254	−0.106	0.00
	Complication rate	Low	0.659	0.227	0.200	1.00
		High	−0.659	−0.227	−0.200	−1.00
	Segment size		46.8%	36.4%	14.1%	2.7%

Figure 1 shows both the importance weights of our estimates and those of the estimates of Schuldt et al. (2017). The differences are very small and play no role in the interpretation of results. The complication rate is the most important attribute, followed by information and the number of treatments, and then distance.

A LCA allows for uncovering distinct segments of individuals with different preferences. Based on the latent variable scores we run a LCA by using the statistical software R (R Core Team 2019) and the flexmix package (Grün and Leisch 2007). Table 1 shows the results of the four-segment solution, Table 2 presents the results from Schuldt et al. (2017), which rely on LCA model proposed by DeSarbo et al. (1995) and implemented in Sawtooth Software Lighthouse Studio (Sawtooth Software 2019). In contrast to the traditional LCA, our method reveals segments at the level of path coefficients in the PLS-SEM and the level of attributes in the DCM.

Table 2 Part worth utilities and segment sizes (Schuldt et al. 2017)

	Attribute	Attribute level	1	2	3	4
Part worth utilities	Distance	1 km	0.511	−0.084	1.103	1.170
		20 km	−0.511	0.084	−1.103	−1.170
	Information on treatment	Brief	1.173	0.913	0.882	0.218
		Detailed	−1.173	−0.913	−0.882	−0.218
	Number of treatments	More than 100 times	0.906	0.466	2.505	−0.173
		Less than 50 times	−0.906	−0.466	−2.505	0.173
	Complication rate	Low	3.103	0.616	1.885	0.263
		High	−3.103	−0.616	−1.885	−0.263
	Segment size		60.1%	20.3%	11.0%	8.7%

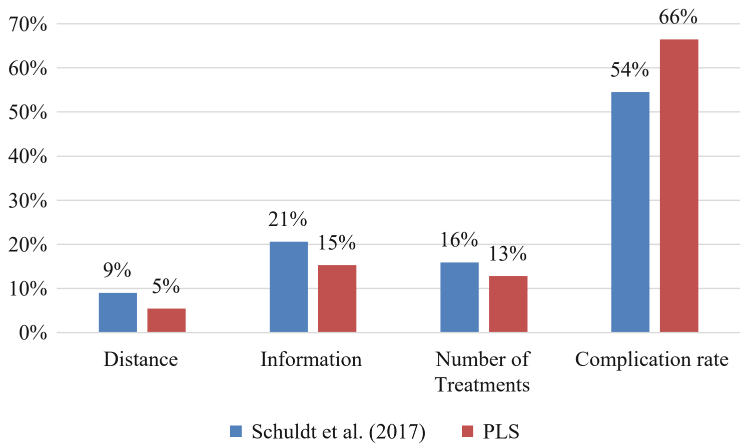

Fig. 2 Importance weights Segment 1

Figure 2 shows that the estimated relative importance scores for the largest segment are almost identical when comparing both methods. The complication rate is the most important attribute, after that information on treatment, then the number of treatments, and subsequently distance. The differences between the two methods play again no role in the interpretation. The results for the other segments are, however, different. While the coefficients have consistent signs, they differ in magnitude. The segment sizes are also different, but with both methods the largest two segments cover more than 80% of the individuals.

A closer look at the smallest segment—based on PLS-SEM estimation—offers, however, interesting insights. For each respondent, in this segment, each individual choice set throughout the DCM study presented one option with a high and a low complication rate alternative and the respondents choose systematically the alternative with the lower complication rate. Thus, our PLS segmentation successfully

Table 3 Fit indices for a one- to four-segment solution

	Number of segments			
Criteria	1	2	3	4
AIC	10559.642	9932.747	−7810.799	−7982.517
BIC	10595.405	10011.427	−7689.203	−7818.005
EIC	Na	0.815	0.793	0.781

identified decision-makers with non-compensatory decision-making behavior. Schuldt et al. (2017) did not reveal this segment.

To determine the optimal number of segments, Table 3 shows the Akaike information criterion (AIC; Akaike 1974), Bayesian information criterion (BIC; Schwarz 1978), and entropic information criterion (EIC; Ramaswamy et al. 1993) for a one- to a four-segment solution obtained by PLS-SEM and flexmix. In line with the original study, also our segmentation method would suggest four distinct segments of respondents.

For the selection of the number of segments, a deeper understanding of the information criteria is, however, important (Oliveira-Brochado and Vitorino Martins 2014). Information criteria are not always useful to determine the optimal numbers of segments, because they don't take into account, how well separated the segments are (Hair et al. 2016b). Further research is needed to understand the performance of the information criteria in the use of PLS-SEM in DCM.

4 Discussion

With this chapter, we substantiate the suitability of PLS-SEM to estimate DCMs drawing on DCE data. The DCM and PLS-SEM results are comparable to other estimation methods; in this case, a multilevel mixed-effects logit model and a LCA, drawing on Sawtooth Software Lighthouse Studio. Using PLS-SEM is promising, because it reduces computational runtimes as compared to hierarchical Bayes estimation, while at the same time, it reduces complexity by focusing on the attribute and not on the attribute level. Furthermore, the use of PLS-SEM enables the researcher to model latent benefit dimensions, by combining multiple attributes into a higher-order construct, when appropriate. More importantly, we show that PLS-SEM in combination with the LCA permits to uncover meaningful and distinct segments in the DCE data and allows uncovering of unobserved heterogeneity. Unobserved heterogeneity is a validity threat in statistical analysis (Becker et al. 2013). The estimation of DCM with PLS-SEM enables the researcher to conduct respondent-specific analysis (e.g., calculating individual R^2 values to uncover individual choice behavior) to develop a deeper understanding of the choices.

These findings are important on a practical level, as they enable companies, for example, to target individuals belonging to different segments. On a theoretical level, they enable researchers to understand more about the choice process (non-compensatory decision-making) and the advantages and disadvantages of different estimation methods. Nonetheless, some limitations pertain to the results that we are presenting here. First, we need to further analyze the differences between traditional LCA and our method, and we need to more fully understand, which data structures are better uncovered with our or the traditional method if the data structure conditions the suitability of a particular method. For instance, does the decision-making approach (compensatory versus non-compensatory) affect the suitability of employing PLS-SEM for the estimation of choices?

Acknowledgments We also would like to acknowledge the contributions provided by Professor Christian M. Ringle.

References

Akaike H (1974) A new look at the statistical model identification. In: Parzen E, Tanabe EK, Kitagawa G (eds) Selected papers of Hirotugu Akaike. Springer series in statistics. Springer, New York, pp 215–222

Becker J-M, Rai A, Ringle CM, Völckner F (2013) Discovering unobserved heterogeneity in structural equation models to avert validity threats. MIS Q 37(3):665–694

DeSarbo WS, Ramaswamy V, Cohen SH (1995) Market segmentation with choice-based conjoint analysis. Mark Lett 6(2):137–147

Greene WH, Hensher DA (2003) A latent class model for discrete choice analysis: contrasts with mixed logit. Transp Res B Methodol 37(8):681–698

Grün B, Leisch F (2007) Fitting finite mixtures of generalized linear regressions in R. Comput Stat Data Anal 51(11):5247–5252

Hair JF, Hult GTM, Ringle CM, Sarstedt M (2016a) A primer on partial least squares structural equation modeling (PLS-SEM). Sage, Los Angeles

Hair JF, Sarstedt M, Matthews LM, Ringle CM (2016b) Identifying and treating unobserved heterogeneity with FIMIX-PLS: part I–method. Eur Bus Rev 28(1):63–76

Hair JF, Ringle CM, Gudergan SP, Fischer A, Nitzl C, Menictas C (2019a) Partial least squares structural equation modeling-based discrete choice modeling: an illustration in modeling retailer choice. Bus Res 12(1):115–142

Hair JF, Risher JJ, Sarstedt M, Ringle CM (2019b) When to use and how to report the results of PLS-SEM. Eur Bus Rev 31(1):2–24

Johnson FR, Lancsar E, Marshall D, Kilambi V, Mühlbacher A, Regier DA et al (2013) Constructing experimental designs for discrete-choice experiments: report of the ISPOR conjoint analysis experimental design good research practices task force. Value Health 16(1):3–13

Lenk PJ, DeSarbo WS, Green PE, Young MR (1996) Hierarchical Bayes conjoint analysis: recovery of partworth heterogeneity from reduced experimental designs. Mark Sci 15(2): 173–191

Louviere JJ, Woodworth G (1983) Design and analysis of simulated consumer choice or allocation experiments: an approach based on aggregate data. J Mark Res 20(4):350–367

Louviere JJ, Street D, Burgess L, Wasi N, Islam T, Marley AA (2008) Modeling the choices of individual decision-makers by combining efficient choice experiment designs with extra preference information. J Choice Model 1(1):128–164

Luce RD, Tukey JW (1964) Simultaneous conjoint measurement: a new type of fundamental measurement. J Math Psychol 1(1):1–27

McFadden D (1974) Conditional logit analysis of qualitative choice behavior. In: Zarembka P (ed) Frontiers in econometrics. Academic Press, New York, pp 105–139

Oliveira-Brochado A, Vitorino Martins F (2014) Identifying small market segments with mixture regression models. Int J Latest Trends Fin Econ Sci 4(4):812–820
Orme BK, Chrzan K (2017) Becoming an expert in conjoint analysis: choice modelling for Pros: Sawtooth Software
R Core Team (2019) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna
Ramaswamy V, DeSarbo WS, Reibstein DJ, Robinson WT (1993) An empirical pooling approach for estimating marketing mix elasticities with PIMS data. Mark Sci 12(1):103–124
Ringle CM, Wende S, Becker J-M (2015) SmartPLS 3. SmartPLS GmbH, Boenningstedt
Sawtooth Software (2019) Lighthouse Studio 9 (Version 9.6.1). Sequim, WA: Sawtooth Software, Inc. https://www.sawtoothsoftware.com/products/online-surveys Accessed 17 Jan 2022
Schuldt J, Doktor A, Lichters M, Vogt B, Robra B-P (2017) Insurees' preferences in hospital choice—a population-based study. Health Policy 121(10):1040–1046
Schwarz G (1978) Estimating the dimension of a model. Ann Stat 6(2):461–464
Train KE (2009) Discrete choice methods with simulation. Cambridge University Press, London

The Use of PLS-SEM in Engineering: A Tool to Apply the Design Science

Ari Melo-Mariano and Ana Bárbara Plá

1 Introduction

The growth in using PLS-SEM in the social sciences is a fact. Hair et al. (2019a) explain that since 2015 there has been a significant advance in the number of published papers in the social sciences area, which evidences the growth of the method's importance. However, PLS-SEM use has not been limited to the social sciences; it has demonstrated its use's versatility in distinct areas such as Knowledge Management (Cepeda-Carrion et al. 2019), Hospitality and Tourism (Usakli and Kucukergin 2018), Information Systems Research (e.g., Hair et al. 2017), Psychology (e.g., Willaby et al. 2015), Medicine (e.g., Menni et al. 2018), and Engineering (Aibinu and Al-Lawati 2010; Durdyev et al. 2018).

The adaptability of the PLS-SEM method to different contexts demonstrates its flexibility in different knowledge areas, positioning it as a "border" tool once it can consolidate different science fields, becoming applicable in contexts where the limits between one science field and another are not well known, as Industry 4.0, or even in the large data volume, as Big Data.

According to Schwab (2019), to think about Industry 4.0 is to comprehend the increasing harmonization and integration in the different areas' discoveries making the fourth industrial revolution unique, the fruit of innovation that resulted from these technologies collaborations from distinct areas that are already real and are transforming the society.

A. Melo-Mariano (✉) · A. B. Plá
University of Brasília, Production Engineering Department, DataLab, University Campus Darcy Ribeiro, Brasília-DF, Brazil
e-mail: arimariano@unb.br

L. Radomir et al. (eds.), *State of the Art in Partial Least Squares Structural Equation Modeling (PLS-SEM)*, Springer Proceedings in Business and Economics,
https://doi.org/10.1007/978-3-031-34589-0_5

Industry 4.0 has been a significant engineering concern, mainly as regards the interconnection with other knowledge areas and the need for a greater comprehension of the production context.

The same interaction inquiries between the areas occur in the data context currently receiving the title of Big Data. For example, Shmueli (2017) explains that one of the Big Data challenges is in the integration of different data sources (with different volumes, variety, and velocity), many times unknown for the Engineering, that is used to work with inanimate data, which can generate erroneous analyses when not considering behavioral aspects.

These Engineering challenges favor the search for models and methodologies that dialogue well with different knowledge areas, maximizing results for the research in the area. Engineering has been using a few decades of distinguished research approaches compared to the natural sciences, the Design Science Research. Dresch et al. (2015) explain that traditional scientific research presents descriptions and explanations about existing phenomena, but Engineering is not limited to existing phenomena and systems. It also includes the project and study of phenomena and objects that do not exist yet but will be created by the engineer for the common good, which shows the need for a distinguished research approach. Thus, Engineering research assumes a position based on an artifact that should be developed, validated, and applied, generating an impact on society.

The new challenges faced by Engineering and its research assumptions based on design approximate the knowledge area to the current understanding of PLS-SEM from the composites' perspective (Henseler 2017; Schuberth et al. 2018), by adopting the view of artificial sciences, advocated by Simon (1978), separating what is natural from what is artificial.

This research seeks to answer the question: How can the PLS-SEM be inserted in the scientific research context in Engineering?

The existing Engineering context is marked by Industry 4.0 and a significant data background (known as the Big Data era), which sometimes favors exploratory research, sometimes confirmatory research, with new constructs, experimental models, or almost experimental, and mainly the scarcity of consolidated literature. These premises can be well developed in the PLS-SEM context. This way is possible to comprehend how the PLS-SEM can be a useful Engineering tool that can collaborate with a better understanding and management of the borders between the different knowledge areas and kinds, being a connection point and favoring in a certain way the areas in which Engineering will be connected with to overcome the new challenges.

Thus, this chapter aims to realize the methodological connections between the PLS-SEM and Engineering through Design Science.

This chapter is divided into one chapter of background, followed by methodology, results, and discussions, and finally, the conclusions and implications for theory and practice.

2 Background

2.1 Design Science

According to Dresch et al. (2015), the concept of Design Science (or project science) was born from observations of the Nobel Prize winner Herbert Simon, from what he interprets as *Artificial Science*, a book launched in 1969. In this work, he detaches what is natural from what is artificial regarding the research object.

In this new context, the artificial is everything that is idealized/designed by man, as machines, organizations, and economy, and in this way, the research engages in the study of the creation and projection of new artifacts, or yet, to support the actual problem solving that is not sustained by the paradigm of the natural sciences (Dresch et al. 2015).

Aken (2004) describes the design science concept, which includes engineering, medical sciences, and modern psychotherapy, and has as its research mission: develop valid and reliable knowledge to design solutions to problems. It is emphasized that the goal is not the action itself but the knowledge that will be generated and could be used to design new solutions, and then, it will be made an action from that point.

Design Science is a scientific paradigm that emerges to help researchers with a goal, a prescription, and, consequently, the creation of knowledge about how to project (Cauchick et al. 2019). It is an approach alternative that promotes investigations through artifacts that will contribute to the creation of new systems or in the improvement of the existing ones, aiming for better results in the actual problem solving, differentiating from the natural sciences that aim the study of known phenomena and problems, and abandoning the concept of new.

This way, artificial science sees the actions involving man as artifacts, a connection between the internal environment (artifact organization) and external environment (operating conditions), projected to answer a particular purpose that should be validated and given a satisfactory solution.

This construction character favors Engineering research, mainly artifacts in the solution creation context. This way to think about models created as ingredients of a construct is to think in artifacts and, at the same time, in composite models of free constructions, typical of Engineering use.

Design Science as a research approach operationalizes its stages in a context called Design Science Research, which is essential to comprehend the difference in the use of both terminologies (one as a method and the other as steps to the method).

2.2 PLS-SEM (Partial Least Square-Structured Equation Model)

The technology increment and the need to improve the understanding of data to decision-making has caused a search for methods to project and deliver relevant results. In this context emerges the partial least square structural equation modeling—PLS-SEM, a union of two critical areas, the econometrics, utilizing predictive models, and the psychometrics, using behavioral models (Hair et al. 2019a).

Structural equation modeling (SEM) is a technique that allows the analysis of multiple relations between factors that are not observable and are challenging to measure through variables that function as indicators. Equations are used to evaluate the relationship between these variables using a statistical toolset.

The variables represent constructs, which are the unobservable factors that will be analyzed. The SEM then enables modeling and estimating parameters of the relation between the constructs.

Henseler (2017) defines construct as "constructions that are theoretically justified." In other words, it is a term that describes an event that generates theoretical interest. Usually, the constructs are formed by measures made by observation or indicators that can represent the concepts.

The SEM realized through the partial least squares (PLS) method becomes a technique that can estimate the cause-and-effect relationship between the constructs (Hair et al. 2012). The PLS cannot represent visual concepts in complex models. However, according to Shmueli et al. (2016), the PLS can "produce parameter estimates of complex models without many of the distributional and other constraints of traditional parametric methods."

The PLS-SEM is more attractive primarily when the research goal focuses on predicting and explaining the variance of the primary constructs through the different explanation constructs (Hair et al. 2012).

Usually, the constructs that study behavior are treated via common factors, but in recent years, composite models have been discovered. Behavioral constructs are, generally, latent variables that can be understood and ontological entities, that is, of the science of being, as people's attributes. However, the design constructs are created as the fruit of thinking (composite constructions) (Henseler 2017), the same as the artifact in Design Science.

Similar to the Design Science concepts, the PLS-SEM has its models anchored in artifacts that should be identified in the literature, developed, validated, and applied, delivering valuable results.

3 Methodology

This was exploratory research with a qualitative-quantitative approach. It was conducted in three steps: first, describing the principles of Design Science and PLS-SEM; second, presenting the principles of Design Science and PLS-SEM to 15 professionals and professors from different engineering fields to opine about the principles' adhesion; and third, conducting a systematic review through an adaptation of the theory of consolidated meta-analytical approach from Mariano and Rocha (2017).

In step 1, many principles were searched in article databases, getting to books (that had to be acquired by the researchers) that explained the concepts of Design Science and PLS-SEM in a more didactic way, which was very important to step 2, since the experts and professors evaluations depended of the understanding of both approaches. In step 2, professionals and professors from different Engineering areas were searched in person and sometimes online. The information contained in the Background section of this article and the Design Science stages was presented to them, named Design Science Research. Some of the interviewees requested additional material, such as Henseler's articles. These results showed opinions about the possible connections between Design Science and PLS-SEM in a text presented in the results section.

Finally, step 3 is through a systematic review of Web of Science (WoS) and Scopus databases. The review was made using the strings presented by Khan et al. (2019) ("partial least squares structural equation modeling" or "partial least squares structural equation modeling" or "partial least squares path modeling" or "partial least squares path modeling" or "PLS path modeling" or "PLS path modeling" or "path model with latent variables" or "PLS-SEM" or "PLS path model" or "PLS-PM" or "SmartPLS" or "PLSgraph" or "PLS-Graph" or "XLSTAT" or "semPLS" or "matrixpls" or "ADANCO" or "PLSgui" or "PLS-GUI" or "LVPLS").

All the papers in Engineering areas in Scopus and WoS were filled with the search endings referring to PLS-SEM to understand the evolution and use of the method in Engineering after that. They analyzed the authors through the number of quotes to have a cluster of the most influential authors (Fig. 1).

The four clusters were analyzed to understand the ecosystem in which all the use of PLS-SEM in Engineering was organized.

Finally, the union between Design Science and the terms that refer to PLS-SEM was sought to understand what had already been done from this perspective (although it was presented last, it was one of the first steps in this chapter, which motivated the work due to the insufficient number of indexed articles).

To analyze the quote network between authors, it was used the VOS viewer. The VOS viewer is free software available to create and visualize bibliographic maps.

According to Van Eck and Waltman (2009), the VOS viewer allows the bibliographic map to be shown in different ways, reinforcing an aspect that would prefer. This allows analysis with more options. In general, it indicates a large amount of data uses, but its use is also beneficial in cases of little data.

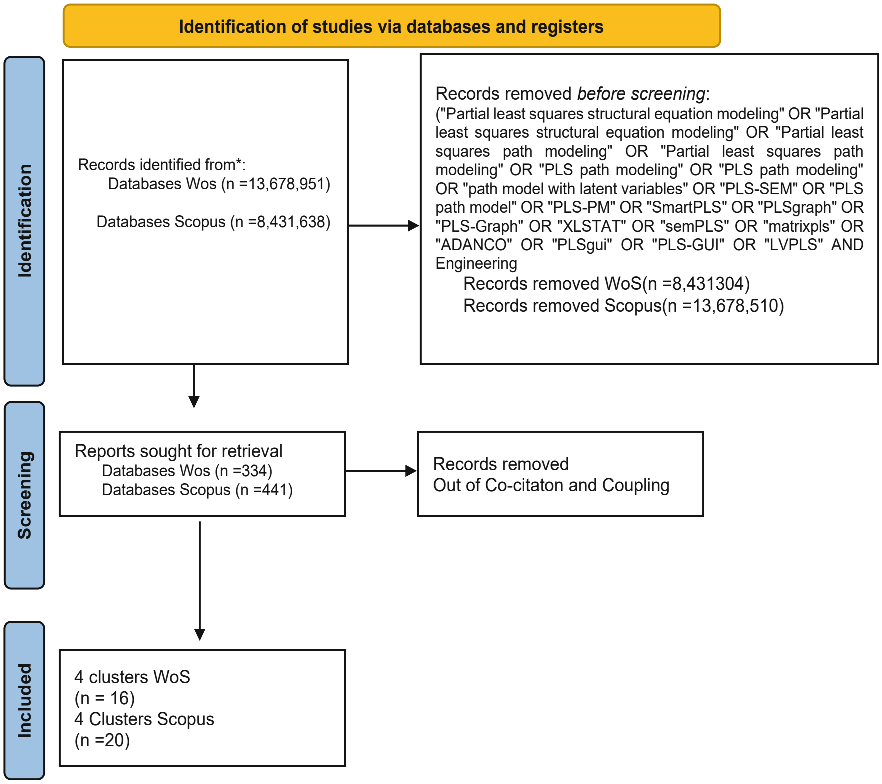

Fig. 1 Research PRISMA process adapted

4 Results and Discussion

4.1 Similarities Between Design Science and PLS-SEM (Partial Least Square-Structured Equation Model)

A first impression is that Design Science is defined through an organization from the bottom up. This organization begins with artifacts following constructs, models, methods, instances, and design propositions (Cauchick et al. 2019):

- Construct: Concepts used to describe problems or specify solutions.
- Models: Set of elements and relations that represent the general structure of reality.
- Method: Set of logical steps needed to effectuate a specific activity.
- Instances: Execution of the artifacts in their natural environment, highlighting the viability and efficiency of the artifacts.
- Design proposition: Technological rules or project rules are considered theoretical contributions of Design Science.

The research with the Design Science approach is oriented to problem-solving, being of practical nature. With the creation of artifacts support, these solutions aim to benefit society increasingly. This organization resembles the modeling phases of the PLS-SEM.

A second impression about Design Science is that the solution it develops doesn't have to be a great solution but a satisfactory one. In other words, it means that the main goal is to solve the problem, not necessarily that it is made in the better way it can, which could have impediments to its application in the real world. In the same way, the PLS-SEM is flexible and can present different solutions in different contexts of the research, such as confirmatory, exploratory, predictive, explanatory, and descriptive (Henseler 2018).

According to Dresch et al. (2015), the result is considered satisfactory to Design Science when there is a consensus that there is a progress in the solution achieved with the new artifact, in comparison with previous solutions, when exists solutions exist.

Another factor to be seen is related to Design Science Research, the stages of Design Science. From the comprehension of the Design Science concept, which is a paradigm, it is possible to understand Design Science Research, a research methodology that applies Design Science. Therefore, it is a suggested structure when one wishes to prescribe solutions or to develop and/or analyze artifacts (Aken 2004).

In Fig. 2, it is possible to verify the steps needed to execute the Design Science Research. The methodology starts with problem identification, which consists in identifying, structuring, and comprehending the situation which it is intended to present a solution. Problem understanding is essential to creating a valuable and proper solution. Then it must be done a systematic review of the literature, to justifying all the research that will be done, in other words, to ground the ideas that will be presented in the paper.

The next step is to identify options for artifacts to be developed to solve the problem previously detected. This step is usually creative and subjective. After analyzing the options, a proper artifact must be proposed. Finally, with the artifact defined, the project to develop the artifact must be created. In Design Science Research, the researcher also has the role of an artifact designer.

According to the interview, both steps occur accordingly to what occurs in the PLS-SEM because the relations between the model's variables must be guided by the scientific literature or through specialists that can ensure the relation (Hair et al. 2019a).

The next step consists of the execution of the project that will result in the development or construction of the artifact chosen in its functional state. The researcher constructs the internal environment of the artifact through algorithms, graphic models, mockups, and other available tools.

In order to validate the artifact, the next step is to evaluate the artifact. After defining the requirements that should be validated and which is the desirable performance of the artifact, the artifact should be evaluated to see if it meets what is expected of it. Finally, the explicit learnings must be executed after the evaluation

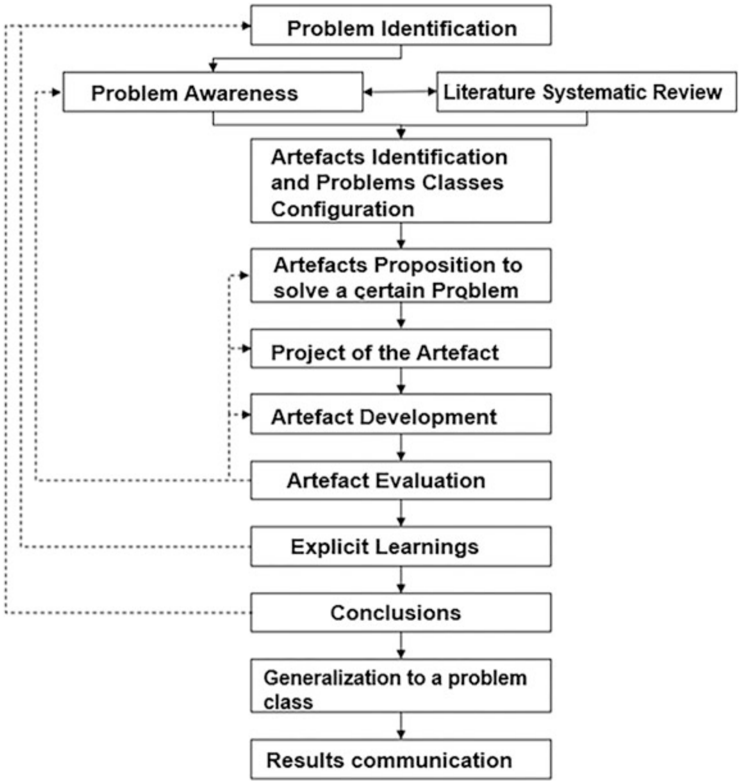

Fig. 2 DSR Methodology. *Source:* Cauchick et al. (2019)

since it is through it that the improvements needed can be perceived, avoiding problems in the future use of the artifact.

It was noticed that this same process of validation occurs inside the PLS-SEM, through the reliability (individual, internal), validity (convergent, discriminating), and multicollinearity, to all the A mode and B mode, with the validity in a construct level and indicator level (Hair et al. 2019a, b).

To start the conclusion, all the things that occurred during the research should be reported. After that, the generalization to a problem class should be made. It must be said that the technological rule quoted by Aken (2004) is defined as "a general knowledge that connects an artifact with a result desired in an application field." He says that a technological rule is a primary product of Design Science, which is the same as the generalization of a problem class. In other words, this stage is about the fact that the solution found is not necessarily just a specific solution to a problem but the knowledge that could be applied to a problem class similar to the problem faced.

Lastly, the results obtained must be reported. Worth it to say that the paper must be accessible and must have a description of what has been done and of all the results found, including problems and limitations faced, since this information will help new research and new pieces of knowledge.

The comparison analysis noticed that besides the creation of the artifact as a result of the Design Science Research to solve problems, there is also the generation of prescriptive knowledge that meets the same goal of the artifact lined up with the PLS-SEM context.

This way, it can be perceived that both contexts are connected through its application stages, making the PLS-SEM a method that incorporates the Design Science principles.

Once the comparisons were finished, a vision was made of the PLS-SEM use over the years in Engineering and its evolution of publications and quotes between authors in Web of Science and Scopus, finishing with the use of the PLS-SEM in Engineering through Design Science.

The results demonstrate that the oldest record is the Web of Science database in the Engineering field, using PLS-SEM in the article “Partial Least-Squares Path Modeling with Latent-Variables” by Gerlach et al. (1979). In this work, the authors explain that the partial least squares model is a solution to evaluate complex models, ensuring an evaluation through many sources. This way, it can be perceived that the PLS-SEM can offer solutions to complex models, being very useful to the current challenges, like Industry 4.0 or Big Data, because in both situations, complex models are treated, making the PLS-SEM become a border tool, consolidating different kinds of knowledge.

The oldest paper about PLS-SEM + Engineering in the Scopus database is “Transnational terrorism: Prospectus for a causal modeling approach” by Hopple (1982), where the author proposes a causal model to comprehend the causes of transnational terrorism. The author explains that the difficulty of this type of model needs a softer modeling, ensuring flexibility to its calculus. Furthermore, the creation of models is a human artifact, considered an artifact or artificial object. These concepts agglutinate the proximity between Engineering through Design Science and the PLS-SEM.

The most quoted record in the Web of Science database in the Engineering field was the paper “Using PLS path modeling in new technology research: updated guidelines” (490 quotes) by Henseler et al. (2016), where authors ratify the increasing use of the tool in the technology field. In the paper, the authors discuss the recent developments in the structural equations via variance area and the possibility of working with composites and factors, being characterized as a formidable statistical tool. It can be perceived that working with composites shows up again as a preponderant factor in the use of techniques in Engineering.

In the Scopus database, the most quoted record is the book *Modelling Transport* (1524 quotes) by de Dios Ortúzar and Willumsen (2011), where the authors present different kinds of modeling to private and public sectors, providing depth in each topic, pedagogically, with the discussion of the role of theory, data, model specification, estimative, validation, and application. This way, it can be noticed that the models always followed different Engineering areas, translating the recreation of the

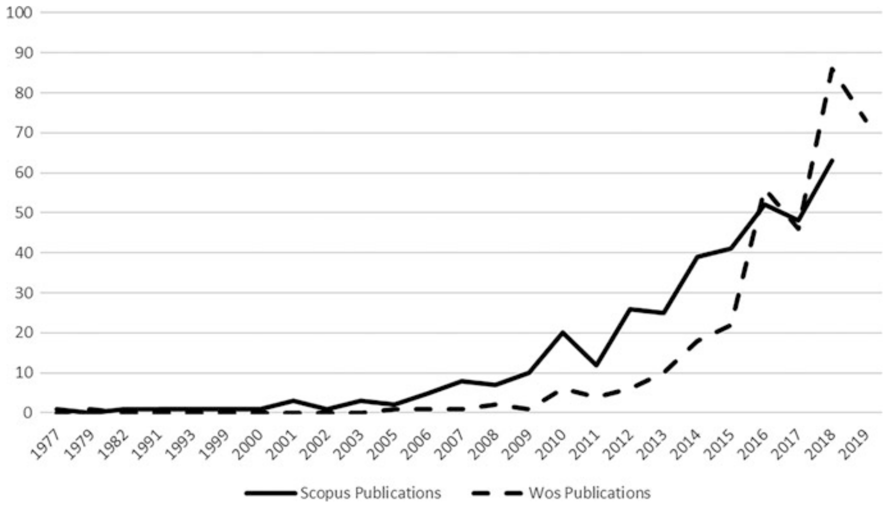

Fig. 3 Publication of articles per year in Scopus and WOS

real in models (artificial design) comprehensible and feasible to be manipulated on a smaller scale, ensuring simulations and observations.

The publications with the use of PLS-SEM in both databases present growth in Engineering, getting to 73 in Web of Science in 2019 and 68 in Scopus (Fig. 3).

The Web of Science database offers the distinction between the different areas of Engineering. This way was possible to count that from the 334 records, 114 papers belonged to Industrial Engineering, 80 to Electrical Engineering, 51 to Civil Engineering, and 45 to Environmental Engineering, and the rest was divided between the other areas.

This way, it can be noticed that the use of PLS, at least in the WoS, is diffuse between the different areas of Engineering. Therefore, it generated two network maps to find the relationship between authors and discover the central quote core (Fig. 4).

The first cluster is revealed by the papers of Ooi and followers, with the approach of the effects of new technologies and the way of accepting and using via structural equations. Gunasekaran forms a second cluster, where evaluated questions related to manufacturing, such as Lean and green product development. Fifty percent of the author's papers are made in Brazil, with the collaboration of Brazilian authors. The Chong nucleus approach papers focus on innovation in the supply chain. Finally, the Henseler kernel, in which composite-type structural equation models adhering to Design Science Theory is shown. The Henseler cluster seen in zoom also reveals the participation of Ringle presents in his paper "Gain more insight from your PLS-SEM results The importance-performance map analysis," the IPMA tool, inside the SmartPLS software, transforming the results in a business application by revealing the variables or indicators that must be prioritized, ratifying once again the use of

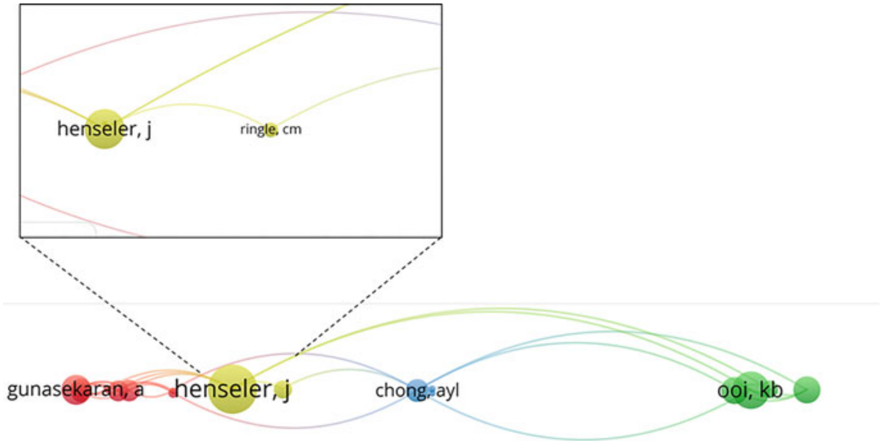

Fig. 4 Network of WoS quote relation. *Source:* Extracted from Vosviewer

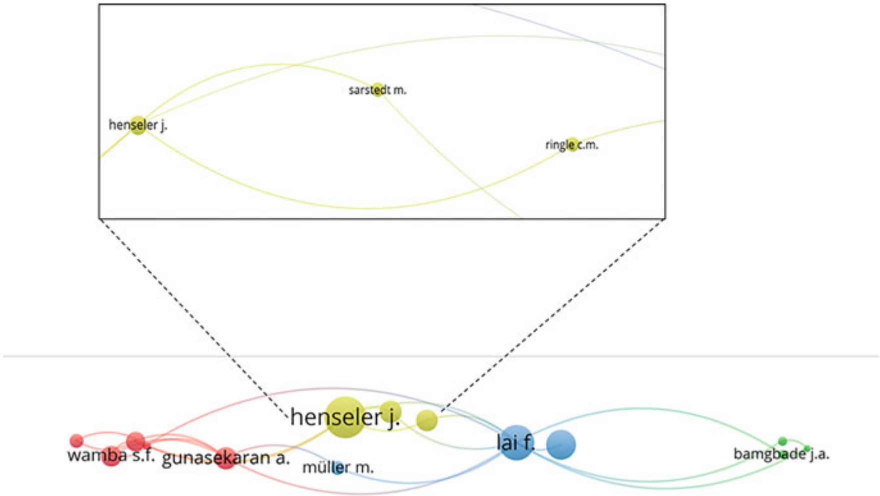

Fig. 5 Network of Scopus quote relation. *Source:* Extracted from Vosviewer

PLS-SEM in Engineering, once the Design Science always defends final results of their studies, applicable.

In the same way, an analysis of the network of authors about the quote dynamic in the Scopus database was made. It found four different nuclei (Fig. 5).

It can be noticed that Wamba and Gunasekaran formed the nucleus one. Wamba approaches analysis via PLS-SEM to the business performance in different strategic contexts, such as Big Data analysis, senior management commitment to environmental issues, and quality dynamics. Gunasekaran repeats the work identified in WoS and additionally is a coauthor of the paper about PLS-SEM and Big Data. Lai brings studies related to China and mainly in the operations area. Muller works with the study of the supply chains, mainly studying providers.

Bamgbade is shown in a cluster studying external and internal factors linked to sustainability in construction. Moreover, finally, Henseler, shows up with most of the papers found in WoS, but with the addition of one editorial about the use of PLS-SEM in the industrial ambit, published in Industrial Management and Data Systems (see Henseler 2016). Ringle and Sarstedt are shown together in articles related to the use of IPMA, just like in WoS.

Both figures were made zoom on the name of Henseler because he is a seminal author when considering the PLS-SEM context since the composite perspective, the approach that is directly associated with the context of the artificial design present in the Design Science is used in Engineering.

The same associations raised in this study were made before by Venable and Baskerville (2012), where the authors used Design Science Research to analyze the use of PLS-SEM, associating their stages, but in this occasion, the suggestion is the use of the PLS-SEM faces to operationalize the Design Science Research.

In 2018 is already shown a study about a model to predict the use of services in mobile health to bring attendance to regions with few resources (Mburu and Oboko 2018). This model is oriented by Design Science and uses PLS-SEM to operationalize the steps of Design Science Research.

This way, it can be perceived that the PLS-SEM is increasingly used in Engineering. The development of PLS-SEM and the new perspectives established by Henseler (Henseler 2017; Schuberth et al. 2018) was significant to establish a more extensive contact between PLS-SEM and its use in Engineering. Some studies have already begun to make associations between PLS-SEM and Design Science Research (Venable and Baskerville 2012) and the use of PLS-SEM to operationalize Design Science (Mburu and Oboko 2018). This way to call the attention to the use of PLS-SEM to the challenges faced by Engineering is to establish new possibilities of research in an interdisciplinary context such as Big Data and Industry 4.0, making the PLS-SEM a border tool.

5 Conclusion and Implications for Theory and Practice

It can be noticed that the use of PLS-SEM in Engineering has been growing each year and that the propositions made by Henseler (Henseler 2017; Schuberth et al. 2018) about composite models approximate the PLS-SEM to the Design Science Research approach. This way, the PLS-SEM is a tool that can contribute in a relevant way to the interaction challenges between distinct areas, as in the Big Data context, computer sciences, information systems, software engineering, statistic, and mathematics in a more generalist way, or specific contexts, as philosophy, marketing, psychology and anthropology (in artificial intelligence development), and biology and medicine (bioinformatic), among others. In such a rich context of information for research modeling, a soft tool, such as PLS-SEM, will help in results that will converge in the different points of view of the different areas.

In the same way, it can be noticed that to the Industry 4.0 context, the unification of considered complex approaches, in an isolated way, in research and application (Internet of things, cyber-physics systems, additive manufacture, among others), potentialized in the integration to create the fourth industrial revolution. This way, the PLS-SEM in Engineering, following the Design Science approach, can help create applied models that will help in this knowledge frontier, helping a better interaction between parts.

This way, the research problem, which was how the PLS-SEM could be inserted in the scientific research context in Engineering, was solved. As a suggestion for future research, we advise a research PLS-SEM following the Design Science Research steps.

References

Aibinu AA, Al-Lawati AM (2010) Using PLS-SEM technique to model construction organizations' willingness to participate in e-bidding. Autom Constr 19(6):714–724

Aken JEV (2004) Management research based on the paradigm of the design sciences: the quest for field-tested and grounded technological rules. J Manag Stud 41(2):219–246

Cauchick P, Lacerda D, Abackerli A, Carvalho M, Costa S, Lima E et al (2019) Metodologia científica para engenharia. Elsevier, Brasil

Cepeda-Carrion G, Cegarra-Navarro JG, Cillo V (2019) Tips to use partial least squares structural equation modelling (PLS-SEM) in knowledge management. J Knowl Manag 23(1):67–89

de Dios Ortúzar J, Willumsen LG (2011) Modelling transport, 4th edn. Wiley, New York

Dresch A, Lacerda DP, Autunes Júnior JAV (2015) Design science research: método de pesquisa para avanço da ciência e tecnologia. Bookman, Sao Paolo

Durdyev S, Zavadskas EK, Thurnell D, Banaitis A, Ihtiyar A (2018) Sustainable construction industry in Cambodia: awareness, drivers and barriers. Sustainability 10(2):392

Gerlach RW, Kowalski BR, Wold HO (1979) Partial least-squares path modelling with latent variables. Anal Chim Acta 112(4):417–421

Hair JF, Sarstedt M, Pieper TM, Ringle CM (2012) 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 Plan 45(5-6):320–340

Hair J, Hollingsworth CL, Randolph AB, Chong AYL (2017) An updated and expanded assessment of PLS-SEM in information systems research. Ind Manag Data Syst 117(3):442–458

Hair JF, Hult GTM, Ringle CM, Sarstedt M, Castillo-Apraiz J, Cepeda Carrión G, Roldán JL (2019a) Manual de partial least squares structural equation modeling (PLS-SEM), Spanish edn. Sage, Barcelona

Hair JF, Risher JJ, Sarstedt M, Ringle CM (2019b) When to use and how to report the results of PLS-SEM. Eur Bus Rev 31(1):2–24

Henseler J (2016) Guest editorial. Ind Manag Data Syst 116(9):1842–1848

Henseler J (2017) Bridging design and behavioral research with variance-based structural equation modeling. J Advert 46(1):178–192

Henseler J (2018) Partial least squares path modeling: Quo vadis? Qual Quan Int J Methodol 52(1): 1–8

Henseler J, Hubona G, Ray PA (2016) Using PLS path modeling in new technology research: updated guidelines. Ind Manag Data Syst 116(1):2–20

Hopple GW (1982) Transnational terrorism: prospectus for a causal modeling approach. Stud Conf Terrorism 6(1-2):73–100

Khan GF, Sarstedt M, Shiau WL, Hair JF, Ringle CM, Fritze MP (2019) Methodological research on partial least squares structural equation modeling (PLS-SEM) an analysis based on social network approaches. Internet Res 29(3):407–429

Mariano AM, Rocha MS (2017) Revisão da Literatura: Apresentação de uma Abordagem Integradora. In: XXVI Congreso Internacional de la Academia Europea de Dirección y Economía de la Empresa (AEDEM), Reggio Calabria, Italy, pp 427–443

Mburu S, Oboko R (2018) A model for predicting utilization of mHealth interventions in low-resource settings: case of maternal and newborn care in Kenya. BMC Med Inform Decis Mak 18(1):67

Menni C, Lin C, Cecelja M, Mangino M, Matey-Hernandez ML, Keehn L et al (2018) Gut microbial diversity is associated with lower arterial stiffness in women. Eur Heart J 39(25): 2390–2397

Schuberth F, Henseler J, Dijkstra TK (2018) Confirmatory composite analysis. Front Psychol 9: 2541

Schwab K (2019) A quarta revolução industrial. Edipro, Sao Paolo

Shmueli G (2017) Research dilemmas with behavioral big data. Big Data 5(2):98–119

Shmueli G, Ray S, Estrada JMV, Chatla S, B. (2016) The elephant in the room: predictive performance of PLS models. J Bus Res 69(10):4552–4564

Simon HA (1978) Rationality as process and as product of thought. Am Econ Rev 68:1–16

Usakli A, Kucukergin KG (2018) Using partial least squares structural equation modeling in hospitality and tourism: do researchers follow practical guidelines? Int J Contemp Hosp Manag 30(11):3462–3512

Van Eck N, Waltman L (2009) Software survey: VOSviewer, a computer program for bibliometric mapping. Scientometrics 84(2):523–538

Venable J, Baskerville R (2012) Eating our own cooking: toward a more rigorous design science of research methods. Electron J Bus Res Methods 10(2):141–153

Willaby HW, Costa DS, Burns BD, MacCann C, Roberts RD (2015) Testing complex models with small sample sizes: a historical overview and empirical demonstration of what partial least squares (PLS) can offer differential psychology. Personal Individ Differ 84:73–78

Discovering Issues in Cross-Cultural Adaptation of Questionnaire Through PLS-SEM Analysis

Fariha Reza and Huma Amir

1 Background

Customers who belong to the bottom of the pyramid (BoP) are marginalized from the mainstream consumer market because of their financial constraints. Literature suggests they desire to bridge the gap between them and the relatively more affluent people by purchasing products that reflect a better social status (Srivastava et al. 2020; Yurdakul et al. 2017). One contemporary example of such aspirational products is smartphones penetrating the BoP segment (Prahalad 2019).

While literature suggests that BoP customers purchase aspirational goods to avoid poverty-related shame, it is not fully understood whether such purchases increase healthy engagement with one's environment and subjective well-being (Dahana et al. 2018). Furthermore, earlier research suggested that BoP customers sometimes sacrificed their immediate needs to save enough to purchase their aspirational products (Atkin et al. 2021). Patience in their everyday lives facilitated such sacrifices and did not let them despair in challenging circumstances. Not much Western marketing literature is available on patience; therefore, studying patience in conjunction with motivation for purchasing aspirational goods, healthy engagement with one's circumstances, and the subjective well-being of BoP customers gave novel insights (Haybron 2016). Social identity theory provided the theoretical justification for studying motivation for buying aspirational products because of social comparisons that BoP customers made with those who were better off in status

F. Reza (✉)
Institute of Business Management, Karachi, Sindh, Pakistan
e-mail: fariha.raza@iobm.edu.pk

H. Amir
Institute of Business Administration, Karachi, Sindh, Pakistan
e-mail: hmair@iba.edu.pk

L. Radomir et al. (eds.), *State of the Art in Partial Least Squares Structural Equation Modeling (PLS-SEM)*, Springer Proceedings in Business and Economics,
https://doi.org/10.1007/978-3-031-34589-0_6

(Jolliffe and Prydz 2021). On the other hand, self-determination theory (Ryan and Deci 2000) provided the theoretical background for studying the relationship between aspirational purchases and healthy engagement with one's life and the subjective well-being of BoP customers (Jaikumar et al. 2018). Therefore, this research empirically tested the relationships between motivation to purchase a smartphone, intention to purchase a smartphone, self-determination, patience, subjective well-being, and attitude toward life of BoP customers.

2 Methodology

2.1 Population and Sampling

People who had, at times, made an occasional aspirational purchase, despite earning less than USD 8 per day, were the population of this research (United Nations Development Program [UNDP] 2008). Furthermore, such people were living in rented accommodations and were often unable to engage in socially relevant consumption practices, and consequently, they were facing negative social consequences of poverty (Yurdakul et al. 2017). A sample of 641 respondents was chosen through chain referral (Reza et al. 2021). Data was collected through a personally administered questionnaire.

2.2 Data Collection Tool and Method

Well-established scales were adopted to measure motivation to purchase a smartphone, intention to purchase a smartphone, self-determination, subjective well-being, patience, and attitude toward life of BoP customers (Reza et al. 2021). The scale items were translated from English to Urdu (Pakistan's national language) to make them understandable to the targeted population, keeping in view its shallow linguistic proficiency (Alyami et al. 2021). During pilot testing, it was realized that the 7-point Likert scale response format contained indistinguishable choices for the respondents; therefore, the response format was reduced to 5-point scale.

2.3 Choice of SmartPLS for Data Analysis

Various software and applications are available to analyze quantitative data; among them, SPSS and SmartPLS are more commonly used in market research (Memon et al. 2021). Compared with SPSS PROCESS macro, PLS-SEM software becomes a better choice because it can test multiple dependency relationships altogether, whereas PROCESS macro tests each dependency relationship individually (Hair et al. 2019; Hayes et al. 2017; Sarstedt et al. 2022). Hence, multiple dependency relationships were tested in this research through PLS-SEM analysis through SmartPLS.

In PLS-SEM analysis, measurement model fit is assessed through outer loadings, composite reliability, bootstrap confidence interval, convergent reliability, and discriminant validity (Sarstedt et al. 2019). After establishing the goodness of fit for the measurement model, the analysis proceeds to determine structural model fit through the coefficient of determination, effect size, Q-Square, and robustness. However, it should be noted that without determining the goodness of fit for a measurement model, a researcher should not test the dependency relationships between constructs that are shown in the structural model (Henseler et al. 2016).

3 Results and Discussion

A reflective measurement model was constructed in SmartPLS that complied with theories of motivation, self-determination, and subjective well-being (Reza et al. 2021). After dropping weak indicators from the model, the composite reliability of the scales that measured motivation, intention to make aspirational purchases, patience, and subjective well-being improved (Henseler et al. 2012) as shown in Table 1.

However, the low values of average variance extracted (AVE) depicted the low convergent validity of the scales (Hair et al. 2019), as shown in Table 2. Earlier research suggested that if the composite reliability of all the constructs is greater than 0.6, a researcher may proceed with structural model fit despite AVE being less than 0.5 (Ringle et al. 2015); therefore, researchers proceeded with the structural model fit analysis despite this limitation.

The discriminant validity of the constructs was determined with the heterotrait-monotrait ratio (Hair et al. 2021). This ratio was less than 0.9 for the constructs, as shown in Table 3, and hence acceptable (Hair et al. 2021; Sarstedt et al. 2019).

Table 1 Composite reliability before and after dropping weak indicators

	Composite reliability	
Scale	Using all indicators	After dropping weak indicators
ATL	0.867	0.881
ITP	0.837	0.838
Motivation	0.786	0.821
Patience	0.680	0.784
SD	0.692	0.678
SWB	0.463	0.753

Notes: ATL attitude toward life; *ITP* intention to purchase; *SD* self-determination; *SWB* subjective well-being

Table 2 Reliability and validity

Scale	Cronbach's alpha	rho_A	Composite reliability	AVE
ATL	0.884	0.888	0.881	0.275
ITP	0.838	0.841	0.838	0.464
Motivation	0.833	0.832	0.821	0.343
Patience	0.792	0.813	0.784	0.389
SD	0.697	0.765	0.678	0.291
SWB	0.743	0.812	0.753	0.518

Notes: ATL attitude toward life; *ITP* intention to purchase; *SD* self-determination; *SWB* subjective well-being; *AVE* average variance extracted

Table 3 HTMT ratio for discriminant validity

Scale	ATL	ITP	Motivation	Patience	SD
ITP	0.192				
Motivation	0.256	0.380			
Patience	0.613	0.121	0.228		
SD	0.363	0.441	0.410	0.243	
SWB	0.406	0.163	0.125	0.286	0.413

Notes: ATL attitude toward life; *ITP* intention to purchase; *SD* self-determination; *SWB* subjective well-being

The researchers revisited the tool translation and adaptation process to find out the reasons behind low divergent validity. Although the methodology conformed to the accepted practices established through literature, the unexpected results could be attributed to the multiple understandings of patience by the targeted population and the shallowness of communication resulting from educational and social deprivations.

4 Conclusion

It is reiterated here that multiple definitions of "attitude toward life" and "patience" exist in the literature (Bülbül and Arslan 2017; Kunieda 2019). However, clarity of thought requires the medium of language for expression (Kronrod 2022). Due to social exclusion and living reduced lives, the linguistic proficiency of the BoP segment is shallow (Piller 2016). As a result, the respondents had difficulty distinguishing the more nuanced differences between life attitudes and patience.

Another plausible reason for low divergent validity between attitude toward life and patience could be the respondents' religiosity. In the psychology of religion, patience is life-shaping virtue and, therefore, could have been understood as a life attitude (Bülbül and Arslan 2017).

The limited linguistic repertoire of the BoP segment and the religiosity were two plausible reasons that lowered the validity of well-established tools when used in a different context. These reasons emphasize the need to develop specific tools in the native language by native researchers. In this case, Pakistani researchers who are familiar with the sociocultural characteristics of the BoP population may come forward to design data collection tools that exhibit greater reliability and validity. Furthermore, using emojis instead of word choices may be explored when developing a tool for the BoP population (Sheth 2021). The pervasive use of cell phones (including smartphones) has raised the level of nonverbal communication through emojis in the BoP population. Therefore, using emojis might be a more sensitive way to measure the response of a BoP respondent.

While this paper highlighted the limitations in the translated tool's compromised convergent and divergent validity, new research may target tool refinement or new tool development to measure the BoP attitudes more accurately and with greater sensitivity. Academic and marketing professionals will appreciate such a development as it will enable them to study a marginalized yet multitrillion-dollar segment more effectively.

References

Alyami M, Henning M, Krägeloh CU, Alyami H (2021) Psychometric evaluation of the Arabic version of the fear of COVID-19 scale. Int J Ment Heal Addict 19(6):2219–2232

Atkin D, Colson-Sihra E, Shayo M (2021) How do we choose our identity? A revealed preference approach using food consumption. J Polit Econ 129(4):1193–1251

Bülbül AE, Arslan C (2017) Investigation of patience tendency levels in terms of self-determination, self-compassion and personality features. Univ J Educ Res 5(9):1632–1645

Dahana WD, Kobayashi T, Ebisuya A (2018) Empirical study of heterogeneous behavior at the base of the pyramid: the influence of demographic and psychographic factors. J Int Consum Mark 30(3):173–191

Hair JF, Risher JJ, Sarstedt M, Ringle CM (2019) When to use and how to report the results of PLS-SEM. Eur Bus Rev 31(1):2–24

Hair JF, Hult GTM, Ringle CM, Sarstedt M, Danks NP, Ray S (2021) Partial least squares structural equation modeling (PLS-SEM) using R. Springer, Cham

Haybron DM (2016) The philosophical basis of eudaimonic psychology. In: Joar V (ed) Handbook of eudaimonic well-being. Springer, Cham, pp 27–53

Hayes AF, Montoya AK, Rockwood NJ (2017) The analysis of mechanisms and their contingencies: PROCESS versus structural equation modeling. Australas Mark J 25(1):76–81

Henseler J, Fassott G, Dijkstra TK, Wilson B (2012) Analysing quadratic effects of formative constructs by means of variance-based structural equation modelling. Eur J Inf Syst 21(1): 99–112

Henseler J, Hubona G, Ray PA (2016) Using PLS path modeling in new technology research: updated guidelines. Ind Manag Data Syst 116(1):2–20

Jaikumar S, Singh R, Sarin A (2018) 'I show off, so I am well off': subjective economic well-being and conspicuous consumption in an emerging economy. J Bus Res 86:386–393

Jolliffe D, Prydz EB (2021) Societal poverty: a relative and relevant measure. World Bank Econ Rev 35(1):180–206

Kronrod A (2022) Language research in marketing. Found Trends® Market 16(3):308–421

Kunieda M (2019) Designing for meaningfulness in life: creating new meaning. https://aaltodoc.aalto.fi/bitstream/handle/123456789/39044/master_Kunieda_Masahiro_2019.pdf;sequence=1 Accessed 8 Nov 2021.

Memon MA, Ramayah T, Cheah J-H, Ting H, Chuah F, Cham TH (2021) PLS-SEM statistical programs: a review. J Appl Struct Eq Model 5(1):i–xiv

Piller I (2016) Monolingual ways of seeing multilingualism. J Multicult Discour 11(1):25–33

Prahalad D (2019, Jan 2) The new fortune at the bottom of the pyramid. Strategy+Business https://www.strategy-business.com/article/The-New-Fortune-at-the-Bottom-of-the-Pyramid. Accessed 30 May 2022.

Reza F, Amir H, Kazmi HAS (2021) Impact of smartphones, self-determination and patience on subjective well-being of bottom of pyramid customers. Revista Brasileira De Marketing 20(2): 279–308

Ringle CM, Silva DD, Bido D (2015) Structural equation modeling with the SmartPLS. Brazil J Market 13(2):56–73

Ryan RM, Deci EL (2000) Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. Am Psychol 55(1):68–78

Sarstedt M, Hair JF, Cheah J-H, Becker J-M, Ringle CM (2019) How to specify, estimate, and validate higher-order constructs in PLS-SEM. Australas Mark J 27(3):197–211

Sarstedt M, Hair JF, Pick M, Liengaard BD, Radomir L, Ringle CM (2022) Progress in partial least squares structural equation modeling use in marketing research in the last decade. Psychol Mark 39(5):1035–1064

Sheth J (2021) New areas of research in marketing strategy, consumer behavior, and marketing analytics: the future is bright. J Mark Theory Pract 29(1):3–12

Srivastava A, Mukherjee S, Jebarajakirthy C (2020) Aspirational consumption at the bottom of Pyramid: a review of Literature and Future Research Directions. J Bus Res 110:246–259

United Nations Development Programme UNDP (2008) Creating value for all: strategies for doing business with the poor. https://www.undp.org/rwanda/publications/creating-value-all-strategies-doing-business-poor Accessed 21 Sept 2022.

Yurdakul D, Atik D, Dholakia N (2017) Redefining the bottom of the pyramid from a marketing perspective. Mark Theory 17(3):289–303

Use of PLS-SEM Approach in the Construction Management Research

Sachin Batra

1 Introduction

Partial least squares structural equation modeling (PLS-SEM) is a contemporary multivariate data analysis method to estimate hypothetically established cause-effect relationship models (Zeng et al. 2021). Many scholars have demonstrated the use of PLS-SEM in different business disciplines, namely, marketing, strategic management, management information systems, international business, human resource management, operations management, supply chain management, accounting, and tourism and hospitality (Lee et al. 2011; Hair et al. 2012; Ringle et al. 2012; Peng and Lai 2012; Kaufmann and Gaeckler 2015; Richter et al. 2016; Sarstedt et al. 2020). But in the construction management domain, scholars are less acquainted with the PLS-SEM approach (Zeng et al. 2021). However, since 2010 when the first article was published in the construction management domain using the PLS-SEM approach, Scopus search results indicate an exponential increase in the number of articles, with 2304 published in 2021.

Therefore, in this chapter, the author attempted to provide a snapshot of the scientific activity of the use of the PLS-SEM approach in the construction management domain by answering the following research questions:

- What are the most influential studies using PLS-SEM in the construction sector?
- What is the co-citation network?
- What are the domains and levels of analysis in which PLS-SEM is used in construction management research?

S. Batra (✉)
NICMAR University, Pune, Maharashtra, India

L. Radomir et al. (eds.), *State of the Art in Partial Least Squares Structural Equation Modeling (PLS-SEM)*, Springer Proceedings in Business and Economics,
https://doi.org/10.1007/978-3-031-34589-0_7