European Business Review

Identifying and treating unobserved heterogeneity with FIMIX-PLS: Part II - A case study

Lucy M. Matthews Marko Sarstedt Joseph F. Hair Christian M. Ringle

Received 14 September 2015
Accepted 20 September 2015

Identifying and treating unobserved heterogeneity with FIMIX-PLS

Part II – A case study

Lucy M. Matthews
Department of Marketing, Middle Tennessee State University, Murfreesboro, Tennessee, USA

Marko Sarstedt
Faculty of Economics and Management, Otto-von-Guericke-University, Magdeburg, Germany, and Faculty of Business and Law, University of Newcastle, Newcastle, Australia

Joseph F. Hair
Coles College of Business, Kennesaw State University, Kennesaw, Georgia, USA, and

Christian M. Ringle
Institute of Human Resource Management and Organizations, Hamburg University of Technology (TUHH), Hamburg, Germany, and Faculty of Business and Law, University of Newcastle, Newcastle, Australia

Abstract

Purpose – Part I of this article (*European Business Review*, Volume 28, Issue 1) offered an overview of unobserved heterogeneity in the context of partial least squares structural equation modeling (PLS-SEM), its prevalence and challenges for social sciences researchers. This paper aims to provide an example that explains how to identify and treat unobserved heterogeneity in PLS-SEM by using the finite mixture PLS (FIMIX-PLS) module in the SmartPLS 3 software (Part II).

Design/methodology/approach – This case study illustrates the application of FIMIX-PLS using a popular corporate reputation model.

Findings – The case study demonstrates the capability of FIMIX-PLS to identify whether unobserved heterogeneity significantly affects structural model relationships. Furthermore, it shows that FIMIX-PLS is particularly useful for determining the number of segments to extract from the data.

Research limitations/implications – Since the introduction of FIMIX-PLS, a range of alternative latent class techniques has appeared. These techniques address some of the limitations of the approach relating to, for example, its failure to handle heterogeneity in measurement models, or its distributional assumptions. This research discusses alternative latent class techniques and calls for the joint use of FIMIX-PLS and PLS prediction-oriented segmentation.

European Business Review
Vol. 28 No. 2, 2016
pp. 208-224

0955-534X
DOI 10.1108/EBR-09-2015-0095

This article refers to the FIMIX-PLS module of the SmartPLS 3 software (www.smartpls.com). Christian M. Ringle acknowledges a financial interest in SmartPLS.

Originality/value – This article is the first to offer researchers, who have not been exposed to the method, an introduction to FIMIX-PLS. Based on a state-of-the-art review of the technique, the paper offers a step-by-step tutorial on how to use FIMIX-PLS by using the SmartPLS 3 software.

Keywords PLS-SEM, Structural equation modeling, Partial least squares, FIMIX-PLS, Finite mixture models, Unobserved heterogeneity

Paper type Research paper

Introduction

Supplementing the article entitled "Identifying and Treating Unobserved Heterogeneity with FIMIX-PLS: Part I – Method", by Hair *et al.* (2016b) in the *European Business Review* (Vol. 28 No. 1), this article provides an example of how to identify and interpret unobserved heterogeneity in partial least squares structural equation modeling (PLS-SEM) by using the FIMIX-PLS module in the SmartPLS 3.2.3 software (Ringle *et al.*, 2015, 2005). As discussed in Part I, checking for unobserved heterogeneity is important to ensure that the results from an aggregate data level analysis are not biased, which would be the case if there were two or more unidentified, dissimilar groups within a data set. When encountered, heterogeneous groups need to be identified, assessed and treated. By following the systematic FIMIX-PLS analysis procedure suggested by Hair *et al.* (2016b) – specifically, see their Figure 3 – this case study offers guidelines for applying the technique to specific research problems.

Case study

Step 1: run the FIMIX-PLS procedure

To illustrate the use of FIMIX-PLS, we draw on the corporate reputation model by Eberl (2010), which Hair *et al.* (2014a, 2016a) use in their PLS-SEM book. The model's purpose is to explain the effects of corporate reputation on customer satisfaction and, ultimately, customer loyalty (*CUSL*). Corporate reputation represents a company's overall evaluation by its stakeholders (Helm *et al.*, 2010). Following Schwaiger (2004), corporate reputation is measured using two dimensions. One dimension represents the cognitive evaluations of the company and measures the construct describing the company's competence (*COMP*). The second dimension captures affective judgments and assesses perceptions of the company's likeability (*LIKE*). Schwaiger (2004) further identifies four antecedent dimensions of reputation – quality (*QUAL*), performance (*PERF*), attractiveness (*ATTR*) and corporate social responsibility (*CSOR*) – measured by a total of 21 formative indicators. The measurement approach has been validated in different countries and applied in various research studies (Eberl and Schwaiger, 2005; Raithel and Schwaiger, 2015; Raithel *et al.*, 2010; Schloderer *et al.*, 2014). Research has shown that compared to alternative reputation measures, the approach performs favorably in terms of convergent validity and predictive validity (Sarstedt *et al.*, 2013).

Figure 1 is a screenshot from the SmartPLS 3 software and shows the corporate reputation model. The original data set stems from Hair *et al.* (2014a) and includes 344 observations. However, some indicators have missing values. We apply the casewise deletion option to the indicators used in the PLS path model, which deletes all the responses with missing values. After casewise deletion, 336 observations remain [1]. In the following, we document each step of the FIMIX-PLS analysis, using SmartPLS 3 software.

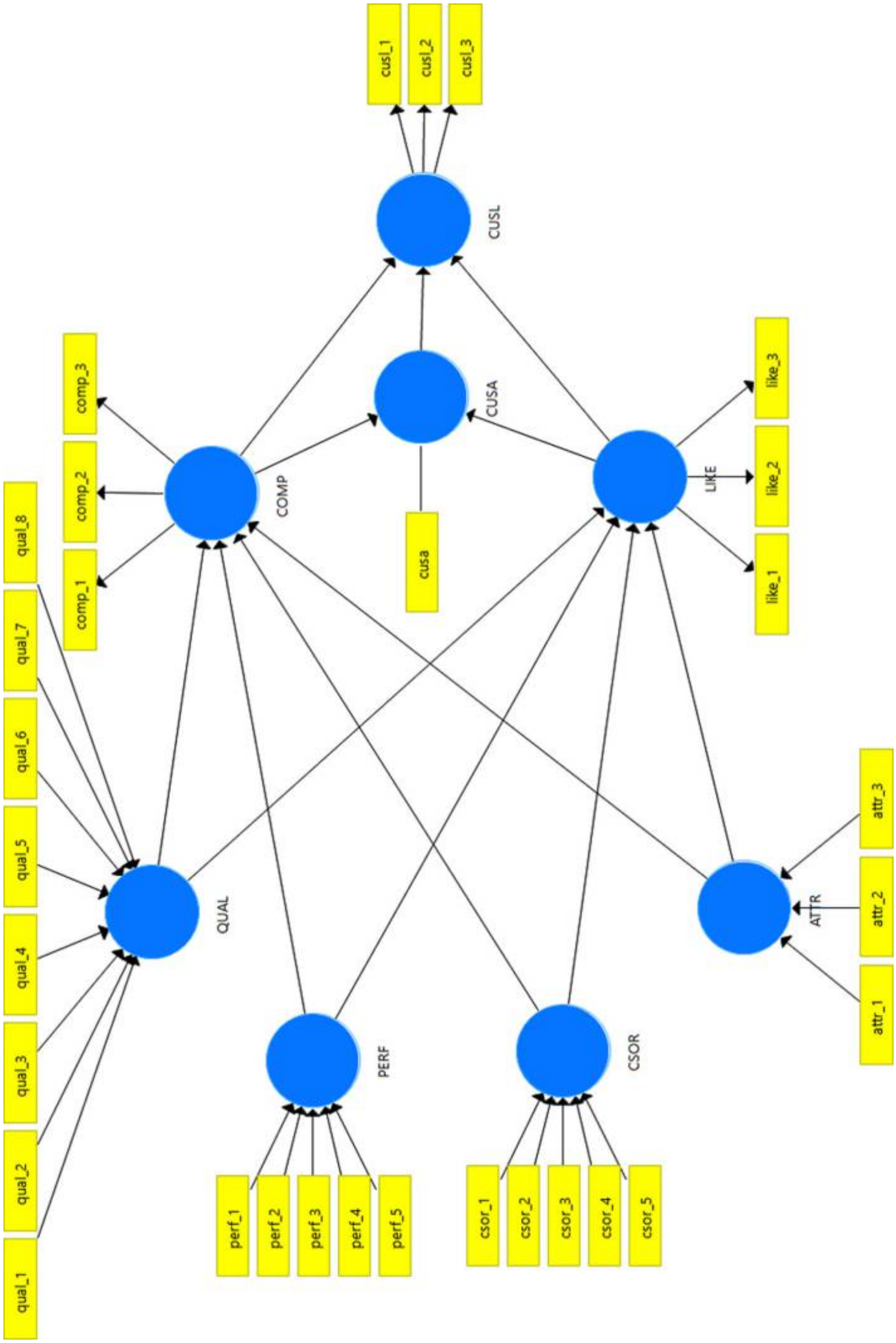


Figure 1.
Corporate reputation model in SmartPLS 3

To initiate the FIMIX-PLS analysis, first draw the model as shown in Figure 1 and select the $N = 336$ as the active data set. Next, click on Calculate → Finite Mixture (FIMIX) Segmentation in the menu bar. Alternatively, you can left-click on the wheel symbol in the tool bar and select the corresponding option in the combo box that opens. After selecting the FIMIX-PLS option, the dialog box in Figure 2 appears. For the initial analysis, start with a one-segment solution and use the default settings for the stop criterion ($1 \cdot 10^{-10} = 1.0\text{E-}10$), maximum number of iterations (5,000) and the number of repetitions (10). The dialog box has two further tabs to specify the standard PLS-SEM algorithm settings and the treatment of missing values. Use the default setting for the PLS-SEM algorithm (Hair *et al.*, 2014a). Finally, click on Start Calculation.

After convergence, SmartPLS 3 opens a new report tab, which shows the results of the FIMIX-PLS analysis. Before analyzing the results in detail, we need to re-run FIMIX-PLS for higher-segment solutions. To determine the upper bound of the range of segment solutions to consider, check the minimum sample size requirements as specified in Hair *et al.* (2016a). With a maximum number of eight arrowheads pointing at any construct in the model (formative indicators of *QUAL*) and assuming a five per cent significance level, as well as a minimum R^2 of 0.25, we would need 54 observations to reliably estimate the model. The greatest integer from dividing the sample size (i.e., 336) by the minimum sample size (i.e., 54) yields a theoretical upper bound of $6.22 = 6$. However, given the complexity of the model, we only consider a one- to five-segment solution, especially as an equal distribution of observations – which would be necessary to meet the minimum sample size requirements – is highly unlikely. Thus, re-run FIMIX-PLS for two to five segments, using the above algorithm settings and save the results report for each run.

Step 2: determine the number of segments

To determine the number of segments to retain from the data, we need to examine the fit indices, which you can find under Quality Criteria → Fit Indices for each of the

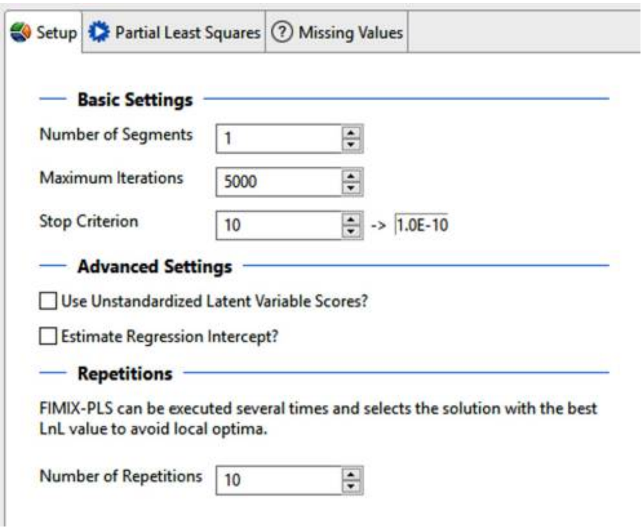


Figure 2.
FIMIX-PLS start dialog box

FIMIX-PLS results reports. To facilitate their comparison across the different segment number solutions, you can export the values to a spreadsheet software, such as Microsoft Excel. To do so, click on Export to clipboard: CSV in the Fit Indices tab and paste the values into an Excel file. Table I provides an overview of the log likelihood values, information criteria and the normed entropy statistic (EN) – as described in Part I (Hair *et al.*, 2016b) – for a one- to five-segment solution.

Two aspects are worth remembering: First, your values will look different from those reported in Table I because FIMIX-PLS initializes randomly and therefore produces different results each time it is run. Nevertheless, the result implications should not differ fundamentally regarding the number of segments to retain. Second, remember that for each fit measure, the optimal solution is the number of segments with the lowest value (see bold numbers in Table I), except in terms of EN, where higher values indicate a better separation of the segments.

Unfortunately, the modified Akaike information criteron with factor 3 (AIC_3) and the consistent Akaike information criterion (CAIC) do not indicate the same number of segments, and neither do AIC_3 and the Bayesian information criterion (BIC). As noted in Part I of this article (Hair et al., 2016b), the Akaike information criterion (AIC) overestimates and the minimum description length with factor 5 (MDL_5) underestimates the correct number of segments. AIC indicates a five-segment solution, suggesting that the correct number is clearly lower than this. On the other extreme, CAIC and, particularly, MDL_5, both indicate a one-segment solution, suggesting that two or more segments should be considered. Further analysis shows that the two best-performing criteria (Sarstedt *et al.*, 2011), the modified Akaike information criterion with factor 4 (AIC_4) and BIC, both indicate two segments, thus providing initial support for this solution. However, the two-segment solution exhibits an EN value below 0.50, suggesting that the two segments are not well separated.

Examining the relative segment sizes in Table II shows that selecting more than two segments is not reasonable. For example, for a three-segment solution, the breakdown of segment sizes is Segment 1 with 52.0 per cent (of 336 = 175 observations), Segment 2 with 41.2 per cent (of 336 = 138 observations) and Segment 3 with only 6.8 per cent (of 336 = 23 observations). As can be seen, with 20 observations, Segment 3 is too small for a segment-specific PLS-SEM analysis. At this point, we could also consider dropping the third segment, as it is too small to warrant valid analysis and, instead, focus on the analysis and interpretation of the other, larger segments.

	No. of segments				
Criteria	1	2	3	4	5
LnL	−1,406.860	−1,353.829	−1,319.790	−1,291.483	−1,269.733
AIC	2,847.720	2,777.658	2,745.580	2,724.966	**2,717.465**
AIC_3	2,864.720	2,812.658	2,798.580	**2,795.966**	2,806.465
AIC_4	2,881.720	**2,847.658**	2,851.580	2,866.966	2,895.465
BIC	2,912.611	**2,911.257**	2,947.886	2,995.981	3,057.188
CAIC	**2,929.611**	2,946.257	3,000.886	3,066.981	3,146.188
MDL_5	**3,308.175**	3,725.653	4,181.114	4,648.040	5,128.080
EN	n/a	0.430	0.638	**0.650**	0.610

Table I.
Fit indices for a one- to five-segment solution

Overall, the results suggest that there is no substantial level of heterogeneity in the data. Researchers can therefore limit their analysis to the aggregate data set, as done in Hair *et al.* (2014a, 2016a). However, to illustrate the application of Steps 3 and 4 of the systematic procedure for applying FIMIX-PLS (Figure 5 in Hair *et al.*, 2016b), we continue with the analysis for a two-segment solution.

Step 3: explanation of the latent segment structure

The explanation of the segment structure in Step 3 focuses on the observations' segment membership probabilities, which can be found under Final Results → Final Partition in the FIMIX-PLS results report. In a first step, we need to assign each observation to one of the two segments based on the maximum segment membership probabilities. This step is not done automatically in SmartPLS 3.2.3 but needs to be completed using, for example, Microsoft Excel. Future versions of SmartPLS will offer this automatic option in a re-organized results report. For now, we refer to the following procedure, which also is relevant when running the previous software version SmartPLS 2 (Ringle *et al.*, 2015). Therefore, go to the Final Partition tab and click on Copy to clipboard at the top right. Next, paste the data from the clipboard into Excel and save the file on your computer. Excel's formula tool allows you to easily compute each observation's segment affiliation based on the probabilities of segment membership:

- Create a new column called *Max* and use the MAX function in Excel for each observation (i.e., row). For example, go to cell E2 and write MAX(B2:C2) and press the return key.
- Create two new columns called *Group 1* and *Group 2*. Use the IF/THEN formula for each of these two columns to indicate whether the maximum probability value is given in Segment 1 or Segment 2. For example, go to cell G2 and type IF(E2 = B2;1;0) and press return. Excel will return the value 1 if the maximum probability in cell E2 equals the probability given in Segment 1 in cell B2, otherwise zero. Similarly, go to cell H2 and write IF(H2 = C2;2;0). Excel will return the value 2 if this observation belongs to Segment 2, otherwise zero.
- Create a new column called *FIMIX-PLS Groups* and use Excel's SUM function to sum the values of columns *Group 1* and *Group 2* per observation (i.e., row). The resulting value equals each observation's group affiliation. For example, go to cell J2 and type in SUM(G2;H2) and press return.

	Relative segment sizes				
No. of segments	Segment 1	Segment 2	Segment 3	Segment 4	Segment 5
2	0.515	0.485			
3	0.520	0.412	0.068		
4	0.480	0.408	0.079	0.033	
5	0.389	0.314	0.168	0.075	0.054

Notes: The table shows the relative segment sizes in declining order per solution (i.e., row); due to label switching, a specific segment can have different labels across the solutions (McLachlan and Peel, 2000); the SmartPLS 3 software uses the relative segment sizes in declining order when assigning the segment numbers to the final FIMIX-PLS segments

Table II. Relative segment sizes (N = 336)

By continuing these analyses for the remaining observations, we obtain the partitioning of all the observations into one of the two segments produced by FIMIX-PLS. In our case, Observation 1 has a probability of 0.424 of belonging to Segment 1, which is lower than the 0.576 probability of belonging to Segment 2. The segment memberships of the first 17 respondents are shown in Figure 3.

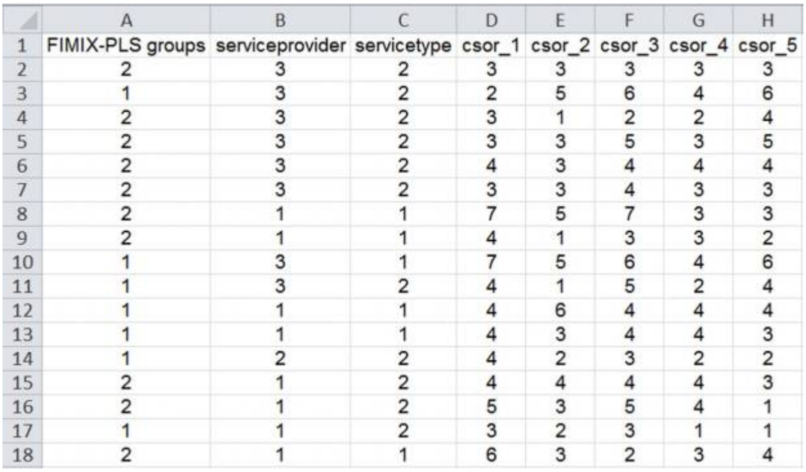
In the next step, we have to transfer the partition as indicated in the *FIMIX-PLS groups* column (Figure 3) to the original data set. To do so, open the original data set with 336 observations and create a new column in which you copy and paste the FIMIX-PLS partition (Figure 4). Next, save the data set with the additional grouping variable in the comma separated value (.csv) file format under a new name (e.g., *FIMIX-PLS for 2 groups.csv*).

We can now use this data set to compare the FIMIX-PLS partition with those indicated by other observable variables in the data set. Unfortunately, the corporate reputation data set only has two such variables, which indicate each respondent's service provider (service provider 1-4) and the type of service the respondent uses (prepaid or contract). Therefore, the chances of reproducing the FIMIX-PLS partition adequately are relatively low. As described in Hair *et al.* (2016b), we strongly recommend that in developing your questionnaire, you include as many potential explanatory variables as reasonable to increase your options when attempting to describe segments identified when using the FIMIX-PLS approach.

Given the limited number of explanatory variables, we only use simple cross tabs to compare the FIMIX-PLS partitions with those produced by the service type and service provider variables. Table III shows the cross tab for the combination of the FIMIX-PLS partition and the service provider. Comparing the cell counts, we find that the best match is achieved when assigning respondents who use service provider 1 or 2 to FIMIX-PLS group 1, and those who use service provider 3 or 4 to FIMIX-PLS group 2 (see the bold numbers in Table III). Using this grouping, (78 + 76 + 26 +21) / 336 = 59.8

	A	B	C	D	E	F	G	H	I	J
1		Segment 1	Segment 2		Max		Group 1	Group 2		FIMIX-PLS groups
2	1	0.424	0.576		0.576		0	2		2
3	2	0.949	0.051		0.949		1	0		1
4	3	0.074	0.926		0.926		0	2		2
5	4	0.387	0.613		0.613		0	2		2
6	5	0.496	0.504		0.504		0	2		2
7	6	0.000	1.000		1.000		0	2		2
8	7	0.136	0.864		0.864		0	2		2
9	8	0.432	0.568		0.568		0	2		2
10	9	0.822	0.178		0.822		1	0		1
11	10	0.798	0.202		0.798		1	0		1
12	11	0.918	0.082		0.918		1	0		1
13	12	0.719	0.281		0.719		1	0		1
14	13	0.860	0.140		0.860		1	0		1
15	14	0.063	0.937		0.937		0	2		2
16	15	0.000	1.000		1.000		0	2		2
17	16	0.888	0.112		0.888		1	0		1
18	17	0.000	1.000		1.000		0	2		2

Figure 3.
Assignment of respondents to groups



	A	B	C	D	E	F	G	H
1	FIMIX-PLS groups	serviceprovider	servicetype	csor_1	csor_2	csor_3	csor_4	csor_5
2	2	3	2	3	3	3	3	3
3	1	3	2	2	5	6	4	6
4	2	3	2	3	1	2	2	4
5	2	3	2	3	3	5	3	5
6	2	3	2	4	3	4	4	4
7	2	3	2	3	3	4	3	3
8	2	1	1	7	5	7	3	3
9	2	1	1	4	1	3	3	2
10	1	3	1	7	5	6	4	6
11	1	3	2	4	1	5	2	4
12	1	1	1	4	6	4	4	4
13	1	1	1	4	3	4	4	3
14	1	2	2	4	2	3	2	2
15	2	1	2	4	4	4	4	3
16	2	1	2	5	3	5	4	1
17	1	1	2	3	2	3	1	1
18	2	1	1	6	3	2	3	4

Figure 4.
Insertion of the new grouping variable into the data set

per cent of the respondents match the FIMIX-PLS partition. Even though the overlap is only slightly below the cut-off value of 60 per cent, this result is not very satisfactory. Alternatively, when contrasting the service type variable with the FIMIX-PLS partition, the resulting overlap is merely 51.8 per cent. As no further explanatory variables are available in the data set, we continue the illustration, using the service provider as an explanatory variable.

Step 4: estimate segment-specific models

To estimate the segment-specific PLS path models, we must first import the newly generated data set *FIMIX-PLS for 2 groups.csv* into SmartPLS [2]. To do so, go to the Project window and right-click on the Corporate Reputation project, which will open the menu shown in Figure 5. In the menu, click on Import Data File, locate the .csv (*FIMIX-PLS for 2 groups.csv*) file and select Open. The dialog box that follows allows you to modify the name of the data set. Continue by clicking on OK.

After importing the data file, the SmartPLS data view – as shown in Figure 6 – opens. Alternatively, you can access the data view by double-clicking on the newly added data set in the SmartPLS project window. Even though we treated the missing values by casewise deletion for the indicators included in the corporate reputation model, additional indicators may still have missing values. For this reason, you should left-click on None next to Missing Value Marker and insert -99.

	FIMIX-PLS groups		
Service provider	1	2	Sum
1	**78**	50	128
2	**76**	47	123
3	17	**26**	43
4	21	**21**	42
Sum	192	144	336

Table III.
Cross tab of FIMIX-PLS partition and service provider

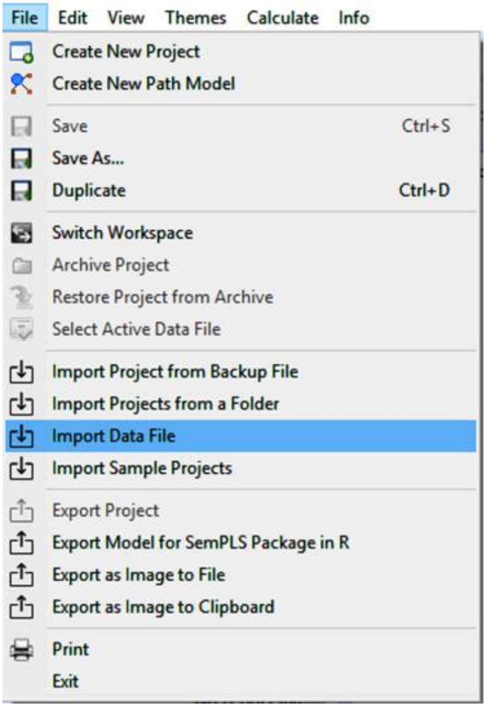


Figure 5.
Import the new data file into SmartPLS

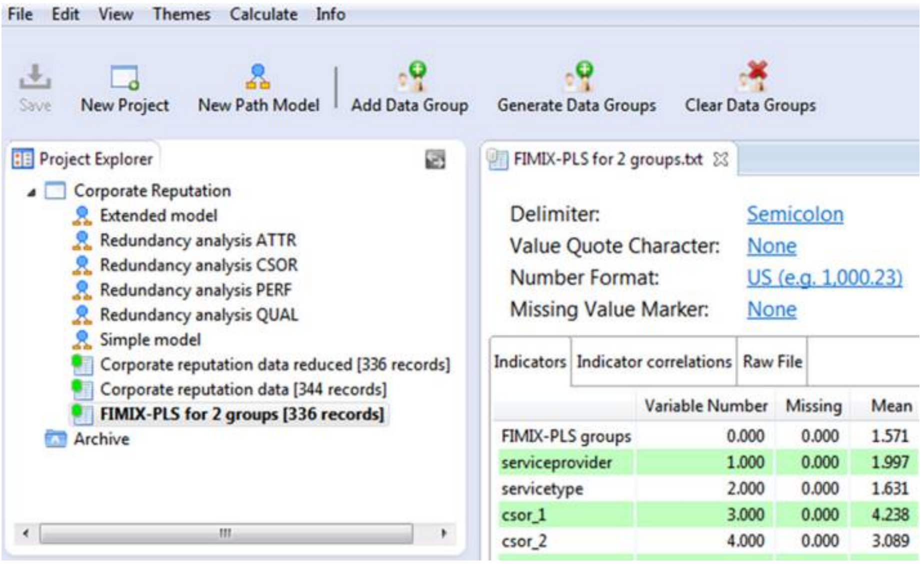


Figure 6.
The SmartPLS data view

The next step is to define the grouping variables that indicate the FIMIX-PLS partition and the corresponding partition produced by the service provider variable. For this purpose, click on Generate Data Groups in the menu bar, which will open the dialog box shown in Figure 7. Click on the pull-down menu next to Group column 0 and select the variable *FIMX-PLS groups*, followed by the OK button. SmartPLS now generates two groups of data based on the *FIMIX-PLS groups* variable: Group 1 with 192 respondents and Group 2 with 144 respondents.

Next, we need to define *service provider* as a second grouping variable. However, contrary to the previous step, *service provider* has four unique values, which we need to condense into two groups. In our case, the service provider values 1 and 2 correspond to the first FIMIX-PLS segment, while the values 3 and 4 correspond to the second FIMX-PLS segment. The Add Data Groups option in the menu bar allows us to likewise define the two service provider groups. Clicking on the Add Data Group button (note that you need to be in the data view) opens a dialog box in which we can define new grouping variables. Under Group Name, specify the new grouping variable's name (e.g., *service provider 1 + 2*) and define the values under Group Terms, as shown in the upper part of Figure 8. After clicking on OK, SmartPLS will create a group that includes those observations where the service provider is less than 3 (i.e., 1 and 2). In a similar manner, we create a new grouping variable (e.g., *service provider 3 + 4*) with all observations where the service provider is higher than 2 (i.e., 3 and 4), as shown in the lower part of Figure 8.

Having defined the groups, we can separately estimate the PLS path model for each data group. Therefore, run the PLS-SEM algorithm by going to Calculate → PLS Algorithm in the menu bar. Alternatively, you can go to the Modeling window, left-click on the wheel symbol in the tool bar and select the corresponding option in the combo box that opens. The dialog box that opens offers the standard options for running the PLS-SEM algorithm, but has a new tab labeled Data Groups. Retain the default settings for the PLS-SEM algorithm, the missing values and the weighting and click on the Data Groups tab. As shown in Figure 9, a menu will open, from which you can select one or more groups to be analyzed separately. Check the boxes next to *service provider 1+2*

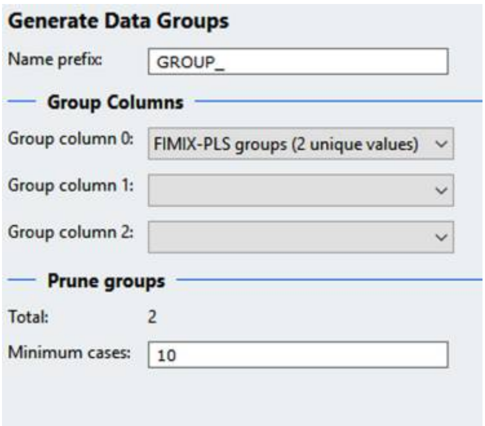


Figure 7.
Generate data groups

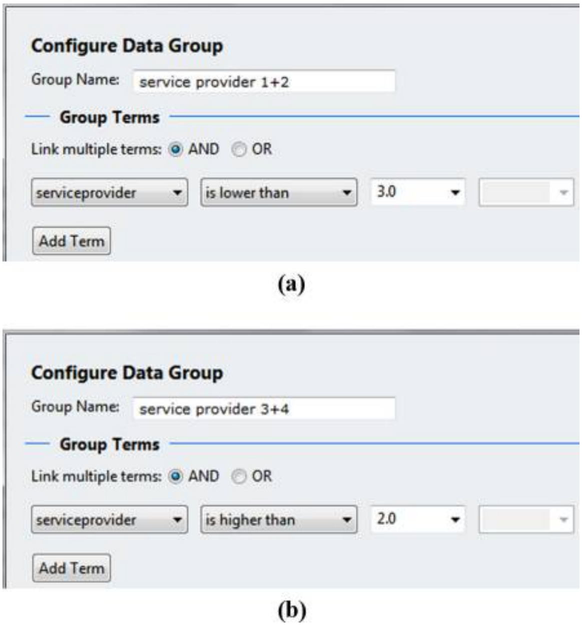


Figure 8.
Generate service provider groups in SmartPLS

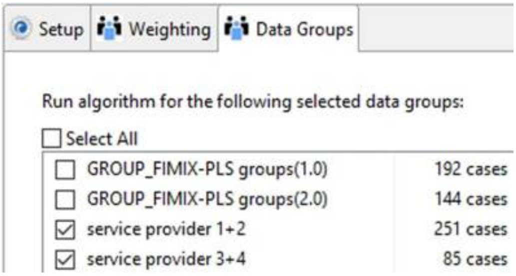


Figure 9.
Data groups in the PLS-SEM Algorithm

and *service provider 3 + 4* to select these two groups for the group-specific PLS path model estimations, followed by Start Calculation.

After completing the analyses of the aggregate data set and each of the groups, SmartPLS will open the results report. Initially, SmartPLS shows the results of the aggregate data analysis, but you can easily scan through the other groups by clicking on the pull-down menu next to Data Group or on the Next button below the results tables. We also need to separately run bootstrapping for each group by going to Calculate → Bootstrapping or using the wheel symbol in the Modeling window. Use the default settings, but make certain that you again select all the groups in the Data Groups tab.

Table IV provides an overview of the aggregate data and the group-specific results, including all the reflective and formative measurement model evaluation criteria as documented in, for example, Hair *et al.* (2016a, 2016b). The measurement model evaluation results support the measures' reliability and validity, with one exception. Discriminant validity assessment, using the heterotrait-monotrait ratio inference

	Original sample	FIMIX-PLS group 1	FIMIX-PLS group 2	Service provider 1 and 2	Service provider 3 and 4
N	336	192	144	251	85
Relative segment size (%)	100.0	57.10	42.90	74.70	25.30
Path					
ATTR → *COMP*	0.085	0.366***	−0.136**	0.047	0.275***
ATTR → *LIKE*	0.159**	0.259***	0.097	0.176**	0.001
COMP → *CUSA*	0.135**	0.463***	−0.190**	0.143*	0.191
COMP → *CUSL*	0.011	0.082	−0.045	0.097	−0.128*
CSOR → *COMP*	0.057	−0.138**	0.184**	0.011	0.215**
CSOR → *LIKE*	0.190***	0.289***	0.071	0.192***	0.105
CUSA → *CUSL*	0.510***	0.594***	0.403***	0.461***	0.611***
LIKE → *CUSA*	0.445***	0.327***	0.451***	0.501***	0.235**
LIKE → *CUSL*	0.334***	0.312***	0.329***	0.300***	0.414***
PERF → *COMP*	0.295***	0.403***	0.130	0.357***	0.162
PERF → *LIKE*	0.116*	0.218***	0.003	0.129	−0.017
QUAL → *COMP*	0.431***	0.321***	0.570***	0.454***	0.281**
QUAL → *LIKE*	0.378***	0.251***	0.442***	0.360***	0.664***
Reflective measurement model assessment (i.e., COMP, CUSL *and* LIKE*)*					
Convergent validity (AVE)	+	+	+	+	+
Reliability (composite reliability, Cronbach's α)	+	+	+	+	+
Discriminant validity ($HTMT_{inference}$)	+	−	+	+	+
Formative measurement model assessment (i.e., ATTR, CSOR, PERF *and* QUAL*)*					
Convergent validity	+	+	+	+	+
Collinearity	+	+	+	+	+
Significance and relevance of the indicators	+	+	+	+	+
R^2					
COMP	0.629	0.811	0.517	0.646	0.654
CUSA	0.293	0.556	0.159	0.365	0.141
CUSL	0.562	0.821	0.354	0.557	0.625
LIKE	0.557	0.824	0.304	0.570	0.525
Weighted R^2					
COMP	0.629	0.685		0.648	
CUSA	0.293	0.386		0.308	
CUSL	0.562	0.621		0.574	
LIKE	0.557	0.601		0.559	

Notes: ***$p \leq 0.01$; **$p \leq 0.05$; *$p \leq 0.10$; +/− = measurement model evaluation criterion fulfilled/not fulfilled in accordance with Hair *et al.* (2014a, 2016a)

Table IV. FIMIX-PLS results

($HTMT_{inference}$) criterion (Henseler *et al.*, 2015; Voorhees *et al.*, 2015) reveals inflated HTMT values between *COMP* and *LIKE* in the *FIMIX-PLS Group 1*, suggesting that the construct measures do not discriminate well. However, our main focus is on the analysis of the two service provider groups, as the FIMIX-PLS segments are by definition latent and therefore offer only limited information on how the observed effects relate to actual consumers.

Comparing the parameters of the two service provider groups shows clear differences in the structural model effects. For example, whereas the effect of *COMP* on *CUSL* is not significant in the first service provider group, it is significantly negative in the second group. Similarly, the antecedents' impact on *COMP* and *LIKE* varies substantially between the groups. For example, whereas in the first group, *CSOR* has a significant effect on *LIKE*, but not on *COMP*, the opposite holds for the second group. Further analyses may involve testing whether these differences in path coefficients are significant – see Sarstedt *et al.* (2011) for further details. Such analyses require prior establishment of at least partial measurement invariance by using Henseler *et al's.* (2016) MICOM approach. However, comparing the path coefficients from the two service provider groups with those from FIMIX-PLS shows that the results do not align well. For example, the −0.136 relationship between *ATTR* and *COMP* in the *FIMIX-PLS Group 2* is not mirrored in any of the service provider groups' results. The same holds for most of the other structural model relationships, which suggests that the FIMIX-PLS results cannot be adequately reproduced by using the service provider variable. This result is not surprising, given the limited overlap between the FIMIX-PLS and service provider partitions. By comparing the weighted R^2, calculated as the sum of the segment-specific R^2 values, weighted by the relative segment size (e.g., for *COMP* in the service provider grouping, the weighted R^2 is $0.747 \times 0.646 + 0.253 \times 0.654 = 0.648$), we can compare the overall R^2, produced by the FIMIX-PLS data grouping, with the overall R^2 resulting from the aggregate data level analysis. This analysis shows that the grouping using the service provider variables increases the model's in-sample predictive power compared to the aggregate level analysis. The increase in (weighted) R^2 is not substantial, providing support for heterogeneity not significantly affecting the data. We can therefore conclude that the results of the overall data analysis are not substantially biased by unobserved heterogeneity. However, this finding does not rule out significant differences in some structural model relationships between the groups as defined by the service provider and service type variables.

Observations and conclusions

The impact of unobserved heterogeneity on PLS-SEM results can be considerable and, if not carefully taken into account, may entail misleading interpretations (Jedidi *et al.*, 1997). As a consequence, PLS-SEM analyses require the use of complementary latent class techniques that allow testing for and dealing with unobserved heterogeneity (Hair *et al.*, 2014b). Originally introduced by Hahn *et al.* (2002) and later extended by Sarstedt *et al.* (2011), FIMIX-PLS is the first and best understood latent class approach to PLS-SEM (Sarstedt, 2008). Based on the mixture regression concept, FIMIX-PLS simultaneously estimates structural model parameters and ascertains the data structure's heterogeneity by calculating the probability that the observations will belong to a certain segment so that they fit into a predetermined number of segments. Thereby, FIMIX-PLS enables researchers to identify and treat unobserved heterogeneity. Alternatively, the results of a FIMIX-PLS analysis may suggest that there is no substantial level of heterogeneity in the data. In this case, researchers can analyze the data on the aggregate level without having to fear substantial biases being introduced by unobserved moderating factors. Therefore, FIMIX-PLS should become part of researchers' methodological toolbox and be routinely used in every PLS-SEM study, even if the research first focuses on the aggregate level results – for an example, see Fiedler and Sarstedt (2014).

Since FIMIX-PLS' introduction, a range of alternative latent class techniques has appeared that address some of the approach's limitations. For example, Squillacciotti (2005, 2010) introduced the PLS typological path modeling procedure, which Esposito Vinzi *et al.* (2007, 2008) extended by presenting the response-based procedure for detecting unit segments (REBUS-PLS). REBUS-PLS gradually re-allocates observations from one segment to the other, with the goal of minimizing the model residuals. As the goal criterion is based on the residuals of the measurement models and the structural model (i.e., the goodness of fit, GoF, index; Tenenhaus *et al.*, 2005), REBUS-PLS takes the heterogeneity in the entire model into account – not just in the structural model, as FIMIX-PLS does. However, the goal criterion does not specifically uncover heterogeneity in formatively measured latent variables (Esposito Vinzi *et al.*, 2008). With formative measurement's generally increasing importance (e.g., Cenfetelli and Bassellier, 2009; Coltman *et al.*, 2008; Diamantopoulos, 2011; Diamantopoulos *et al.*, 2008), this REBUS-PLS limitation is critical. Furthermore, Henseler and Sarstedt (2013) have challenged the usefulness of the GoF by showing that the index is largely unsuitable for judging the quality of a PLS path model.

Apart from these conceptual concerns, REBUS-PLS re-assigns many observations per iteration and thus conducts a random walk without systematically advancing toward the goal criterion (Ringle *et al.*, 2014, 2013). Becker *et al.* (2013) addressed these limitations by presenting the PLS-SEM prediction-oriented segmentation approach (PLS-POS), which is applicable to all kinds of PLS path models, regardless of whether the latent variables draw on reflective or formative measurement models. Their simulation study shows that PLS-POS performs well for segmentation purposes and provides favorable outcomes compared with alternative segmentation techniques. Genetic algorithm segmentation in PLS-SEM (PLS-GAS; Ringle *et al.*, 2014, 2013) is another versatile approach to uncover and treat heterogeneity in measurement and structural models. This approach consists of two stages. The first stage uses a genetic algorithm that aims at revealing the partition, which minimizes the endogenous latent variables' unexplained variance. The advantage of implementing a genetic algorithm is that it has the capability to avoid local optimum solutions and covers a wide area of the potential search space before delivering a final, best solution. In the second stage, a deterministic hill-climbing approach aims at delivering an even better fitting solution. The PLS-GAS method provides excellent results that usually outperform FIMIX-PLS outcomes and particularly those of REBUS-PLS. However, the downside is that it is computationally very demanding. Such run-times poses a serious problem for PLS-GAS's application in studies, hindering its dissemination in research practice. As a remedy, researchers have recently introduced the iterative reweighted regressions segmentation method (PLS-IRRS; Schlittgen *et al.*, 2015).

The new PLS-IRRS approach builds on Schlittgen's (2011) clusterwise regression. Robust regression reduces the weighting of observations with extreme values, which mitigates the influence of outliers in the data set. In the adaption of this concept for PLS-based segmentation, outliers are not treated as such, but as their own segment. Hence, when robust regression identifies a group of similar outliers, they may become a data group of their own and represent a segment-specific PLS-SEM solution. At the same time, PLS-IRRS reduces the impact of inhomogeneous observations in the computation of segment-specific PLS-SEM solutions. Like PLS-POS and PLS-GAS, PLS-IRRS is applicable to all kinds of PLS path models. Moreover, initial simulation results show that PLS-IRRS performs well in terms of parameter recovery and predictive power (Schlittgen *et al.*, 2015). However, the key advantage of PLS-IRRS is its speed. PLS-IRRS is much faster than PLS-GAS and provides very similar results.

Despite some limitations *vis-à-vis* more recently proposed approaches, such as PLS-POS and PLS-GAS, FIMIX-PLS plays an important role in research, as it indicates whether a significant level of unobserved heterogeneity is present in the data. Furthermore, FIMIX-PLS is unique, as it is the only latent class technique that offers concrete guidance regarding the number of segments to extract from the data (i.e., by using information retention and entropy measures). Therefore, combining FIMIX-PLS with another latent class technique is particularly promising. When FIMIX-PLS suggests that unobserved heterogeneity is an issue, researchers should carefully evaluate how many segments possibly underlie the data and use this information as input for further analysis, using another latent class technique. Because of its performance and implementation in the SmartPLS 3 software, PLS-POS appears particularly promising in this regard. On the contrary, if FIMIX-PLS suggests that no substantial level of heterogeneity is present, researchers can interpret the aggregate level results with confidence.

Notes

1. The SmartPLS 3 project file can be downloaded from www.pls-sem.com/files/corporate_reputation_fimix-pls.zip; the reduced data set (after casewise deletion; $n = 336$) can be downloaded from www.pls-sem.com/files/corporate_reputation_data_reduced.csv
2. The file can also be downloaded from www.pls-sem.com/files/ FIMIX-PLS for 2 groups.csv

References

Becker, J.-M., Rai, A., Ringle, C.M. and Völckner, F. (2013), "Discovering unobserved heterogeneity in structural equation models to avert validity threats", *MIS Quarterly*, Vol. 37 No. 3, pp. 665-694.

Cenfetelli, R.T. and Bassellier, G. (2009), "Interpretation of formative measurement in information systems research", *MIS Quarterly*, Vol. 33 No. 4, pp. 689-708.

Coltman, T., Devinney, T.M., Midgley, D.F. and Venaik, S. (2008), "Formative versus reflective measurement models: two applications of formative measurement", *Journal of Business Research*, Vol. 61 No. 12, pp. 1250-1262.

Diamantopoulos, A. (2011), "Incorporating formative measures into covariance-based structural equation models", *MIS Quarterly*, Vol. 35 No. 2, pp. 335-358.

Diamantopoulos, A., Riefler, P. and Roth, K.P. (2008), "Advancing formative measurement models", *Journal of Business Research*, Vol. 61 No. 12, pp. 1203-1218.

Eberl, M. (2010), "An application of PLS in multi-group analysis: the need for differentiated corporate-level marketing in the mobile communications industry", in Esposito Vinzi, V., Chin, W.W., Henseler, J. and Wang, H. (Eds), *Handbook of Partial Least Squares: Concepts, Methods and Applications (Springer Handbooks of Computational Statistics Series, Vol. II)*. Springer, Heidelberg, Dordrecht, London, New York, NY, pp. 487-514.

Eberl, M. and Schwaiger, M. (2005), "Corporate reputation: disentangling the effects on financial performance", *European Journal of Marketing*, Vol. 39 Nos 7/8, pp. 838-854.

Esposito Vinzi, V., Ringle, C.M., Squillacciotti, S. and Trinchera, L. (2007), "Capturing and treating unobserved heterogeneity by response based segmentation in PLS path modeling: a comparison of alternative methods by computational experiments", ESSEC Research Center, Working Paper No. 07019, Cergy Pontoise Cedex.

Esposito Vinzi, V., Trinchera, L., Squillacciotti, S. and Tenenhaus, M. (2008), "REBUS-PLS: a response-based procedure for detecting unit segments in PLS path modeling", *Applied Stochastic Models in Business and Industry*, Vol. 24 No. 5, pp. 439-458.

Fiedler, M. and Sarstedt, M. (2014), "Influence of community design on user behaviors in online communities", *Journal of Business Research*, Vol. 67 No. 11, pp. 2258-2268.

Hahn, C., Johnson, M.D., Herrmann, A. and Huber, F. (2002), "Capturing customer heterogeneity using a finite mixture PLS approach", *Schmalenbach Business Review*, Vol. 54 No. 3, pp. 243-269.

Hair, J.F., Hult, G.T.M., Ringle, C.M. and Sarstedt, M. (2014a), *A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM)*, Sage, Thousand Oaks, CA.

Hair, J.F., Hult, G.T.M., Ringle, C.M. and Sarstedt, M. (2016a), *A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM)*, Sage, Thousand Oaks, CA.

Hair, J.F., Sarstedt, M., Hopkins, L. and Kuppelwieser, V.G. (2014b), "Partial least squares structural equation modeling (PLS-SEM): an emerging tool in business research", *European Business Review*, Vol. 26 No. 2, pp. 106-121.

Hair, J.F., Sarstedt, M., Matthews, L. and Ringle, C.M. (2016b), "Identifying and treating unobserved heterogeneity with FIMIX-PLS: part I - method", *European Business Review*, Vol. 28 No. 1, pp. 63-76.

Helm, S., Eggert, A. and Garnefeld, I. (2010), "Modelling the impact of corporate reputation on customer satisfaction and loyalty using PLS", in Esposito Vinzi, V., Chin, W.W., Henseler, J. and Wang, H. (Eds), *Handbook of Partial Least Squares: Concepts, Methods and Applications (Springer Handbooks of Computational Statistics Series, Vol. II)*. Springer, Heidelberg, Dordrecht, London, New York, NY, pp. 515-534.

Henseler, J., Ringle, C.M. and Sarstedt, M. (2015), "A new criterion for assessing discriminant validity in variance-based structural equation modeling", *Journal of the Academy of Marketing Science*, Vol. 43 No. 1, pp. 115-135.

Henseler, J., Ringle, C.M. and Sarstedt, M. (2016), "Testing measurement invariance of composites using partial least squares", *International Marketing Review*, in press.

Henseler, J. and Sarstedt, M. (2013), "Goodness-of-fit indices for partial least squares path modeling", *Computational Statistics*, Vol. 28 No. 2, pp. 565-580.

Jedidi, K., Jagpal, H.S. and DeSarbo, W.S. (1997), "Finite-mixture structural equation models for response-based segmentation and unobserved heterogeneity", *Marketing Science*, Vol. 16 No. 1, pp. 39-59.

McLachlan, G.J. and Peel, D. (2000), *Finite Mixture Models*, Wiley, New York, NY.

Raithel, S. and Schwaiger, M. (2015), "The effects of corporate reputation perceptions of the general public on shareholder value", *Strategic Management Journal*, Vol. 36 No. 6, pp. 945-956.

Raithel, S., Wilczynski, P., Schloderer, M.P. and Schwaiger, M. (2010), "The value-relevance of corporate reputation during the financial crisis", *Journal of Product and Brand Management*, Vol. 19 No. 6, pp. 389-400.

Ringle, C.M., Sarstedt, M., Schlittgen, R. and Taylor, C.R. (2013), "PLS path modeling and evolutionary segmentation", *Journal of Business Research*, Vol. 66 No. 9, pp. 1318-1324.

Ringle, C.M., Sarstedt, M. and Schlittgen, R. (2014), "Genetic algorithm segmentation in partial least squares structural equation modeling", *OR Spectrum*, Vol. 36 No. 1, pp. 251-276.

Ringle, C.M., Wende, S. and Becker, J.-M. (2015), *SmartPLS 3*, SmartPLS GmbH, Bönningstedt.

Ringle, C.M., Wende, S. and Will, A. (2005), *SmartPLS 2*, SmartPLS, Hamburg.

Sarstedt, M. (2008), "A review of recent approaches for capturing heterogeneity in partial least squares path modeling", *Journal of Modelling in Management*, Vol. 3 No. 2, pp. 140-161.

Sarstedt, M., Becker, J.-M., Ringle, C.M. and Schwaiger, M. (2011), "Uncovering and treating unobserved heterogeneity with FIMIX-PLS: which model selection criterion provides an appropriate number of segments?", *Schmalenbach Business Review*, Vol. 63 No. 1, pp. 34-62.

Sarstedt, M., Wilczynski, P. and Melewar, T.C. (2013), "Measuring reputation in global markets - A comparison of reputation measures' convergent and criterion validities ", *Journal of World Business*, Vol. 48 No. 3, pp. 329-339.

Schlittgen, R. (2011), "A weighted least-squares approach to clusterwise regression", *Advances in Statistical Analysis*, Vol. 95 No. 2, pp. 205-217.

Schlittgen, R., Ringle, C.M., Sarstedt, M. and Becker, J.-M. (2015), "Segmentation of PLS Path models by iterative reweighted regressions", in Henseler, J., Ringle, C.M., Roldán, J.L. and Cepeda, G. (Eds), *2nd International Symposium on Partial Least Squares Path Modeling: The Conference for PLS Users*, *Seville*.

Schloderer, M.P., Sarstedt, M. and Ringle, C.M. (2014), "The relevance of reputation in the nonprofit sector: the moderating effect of socio-demographic characteristics", *International Journal of Nonprofit and Voluntary Sector Marketing*, Vol. 19 No. 2, pp. 110-126.

Schwaiger, M. (2004), "Components and. parameters of corporate. reputation: an empirical study", *Schmalenbach Business Review*, Vol. 56 No. 1, pp. 46-71.

Squillacciotti, S. (2005), "Prediction oriented classification in PLS path modeling", in Aluja, T., Casanovas, J., Esposito Vinzi, V. and Tenenhaus, M. (Eds), *PLS & Marketing: Proceedings of the 4th International Symposium on PLS and Related Methods (PLS'05)*, *DECISIA, Paris*, pp. 499-506.

Squillacciotti, S. (2010), "Prediction-oriented classification in PLS path modeling", in Esposito Vinzi, V., Chin, W. W., Henseler, J. and Wang, H. (Eds), *Handbook of Partial Least Squares: Concepts, Methods and Applications (Springer Handbooks of Computational Statistics Series, Vol. II)*. Springer, Heidelberg, Dordrecht, London, New York, NY, pp. 219-233.

Tenenhaus, M., Esposito Vinzi, V., Chatelin, Y.-M. and Lauro, C. (2005), "PLS path modeling", *Computational Statistics & Data Analysis*, Vol. 48 No. 1, pp. 159-205.

Voorhees, C.M., Brady, M.K., Calantone, R. and Ramirez, E. (2015), "Discriminant validity testing in marketing: an analysis, causes for concern, and proposed remedies", *Journal of the Academy of Marketing Science*, Vol. 44 No. 1, pp. 119-134.

Corresponding author
Marko Sarstedt can be contacted at: marko.sarstedt@ovgu.de

This article has been cited by:

1. Araceli Picón-Berjoyo, Carolina Ruiz-Moreno, Ignacio Castro. 2016. A mediating and multigroup analysis of customer loyalty. *European Management Journal* **34**:6, 701-713. [CrossRef]
2. Nadine Schirmer, Christian M. Ringle, Siegfried P. Gudergan, Matthias S. G. Feistel. 2016. The link between customer satisfaction and loyalty: the moderating role of customer characteristics. *Journal of Strategic Marketing* 1-20. [CrossRef]
3. Rainer Schlittgen, Christian M. Ringle, Marko Sarstedt, Jan-Michael Becker. 2016. Segmentation of PLS path models by iterative reweighted regressions. *Journal of Business Research* **69**:10, 4583-4592. [CrossRef]
4. Marko Sarstedt, Christian M. Ringle, Siegfried P. Gudergan. 2016. Guidelines for treating unobserved heterogeneity in tourism research: A comment on Marques and Reis (2015). *Annals of Tourism Research* **57**, 279-284. [CrossRef]
5. Joe F. Hair, Jr. Kennesaw State University, Kennesaw, Georgia, USA Marko Sarstedt Faculty of Economics and Management, Otto-von-Guericke-University Magdeburg, Magdeburg, Germany and Faculty of Business and Law, University of Newcastle, Newcastle, Australia Lucy M Matthews Department of Marketing, Middle Tennessee State University, Murfreesboro, Tennessee, USA Christian M Ringle Institute of Human Resource Management and Organizations, Hamburg University of Technology (TUHH), Hamburg, Germany and Faculty of Business and Law, University of Newcastle, Newcastle, Australia . 2016. Identifying and treating unobserved heterogeneity with FIMIX-PLS: part I – method. *European Business Review* **28**:1, 63-76. [Abstract] [Full Text] [PDF]