Identifying and treating unobserved heterogeneity with FIMIX-PLS: part I – method

Joe F. Hair
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, 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

Received 14 September 2015
Revised 14 September 2015
Accepted 20 September 2015

Abstract

Purpose – The purpose of this paper is to provide an overview of unobserved heterogeneity in the context of partial least squares structural equation modeling (PLS-SEM), its prevalence and challenges for social science researchers. Part II – in the next issue (*European Business Review*, Vol. 28 No. 2) – presents a case study, which illustrates how to identify and treat unobserved heterogeneity in PLS-SEM using the finite mixture PLS (FIMIX-PLS) module in the SmartPLS 3 software.

Design/methodology/approach – The paper merges literatures from various disciplines, such as management information systems, marketing and statistics, to present a state-of-the-art review of FIMIX-PLS. Based on this review, the paper offers guidelines on how to apply the technique to specific research problems.

Findings – FIMIX-PLS offers a means to identify and treat unobserved heterogeneity in PLS-SEM and is particularly useful for determining the number of segments to extract from the data. In the latter respect, prior applications of FIMIX-PLS restricted their focus to a very limited set of criteria, but future studies should broaden the scope by considering information criteria, theory and logic.

Research limitations/implications – Since the introduction of FIMIX-PLS, a range of alternative latent class techniques have emerged to address some of the limitations of the approach relating, for example, to the technique's inability to handle heterogeneity in the measurement models and its distributional assumptions. The second part of this article (Part II) discusses alternative latent class techniques in greater detail and calls for the joint use of FIMIX-PLS and PLS prediction-oriented segmentation.

European Business Review
Vol. 28 No. 1, 2016
pp. 63-76
© Emerald Group Publishing Limited
0955-534X
DOI 10.1108/EBR-09-2015-0094

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 paper 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 in Part I, Part II follows up by offering a step-by-step tutorial on how to use FIMIX-PLS in SmartPLS 3.

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

Paper type General review

Introduction

Advances in technology and computing power have improved academics' ability to test complex theoretical models. Nevertheless, to fully utilize the capabilities of recently developed softwares, researchers must also develop their skills and embrace the technology. The SmartPLS software offers academics an opportunity to enhance their capabilities (Ringle *et al.*, 2015, 2005b). This software has rapidly expanded the application of partial least squares structural equation modeling (PLS-SEM) in the social sciences in recent years (Hair *et al.*, 2011, 2014b), including information systems (Rapp *et al.*, 2010; Ringle *et al.*, 2012), accounting (Lee *et al.*, 2011), marketing (Hair *et al.*, 2012b), strategic management (Hair *et al.*, 2012a) and related disciplines (Kaufmann and Gaeckler, 2015; Peng and Lai, 2012; Sarstedt *et al.*, 2014). Although advanced training workshops are often offered, it is not always feasible for academics to improve their skills by attending such workshops. Alternatively, the timing of the workshop may not be aligned with a particular project of interest. Other approaches, such as scholarly articles, are therefore very useful to expand academics' knowledge of PLS-SEM and related methods.

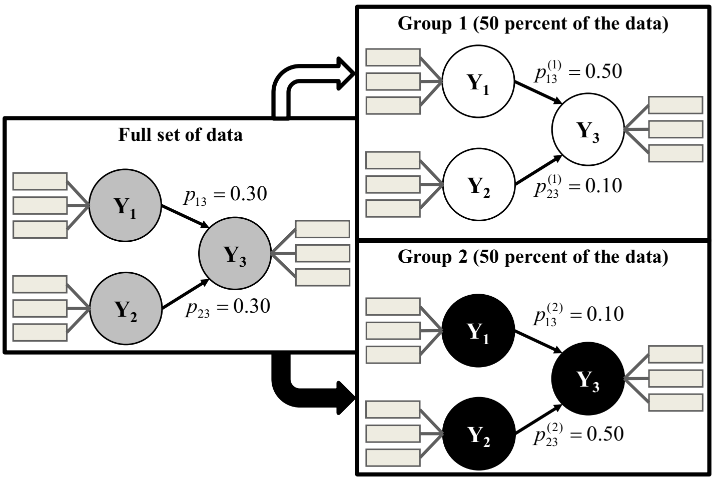
The purpose of this paper is to explain and illustrate the use of finite mixture PLS (FIMIX-PLS); a useful analysis approach in PLS-SEM that allows for dealing with unobserved heterogeneity. Unobserved heterogeneity occurs when there are significant differences in model relationships between groups of data and the sources of these differences cannot be traced back to any observable characteristics such as gender, age or income. Specifically, this paper provides an overview of unobserved heterogeneity, its prevalence and its challenges for social science researchers. It also introduces FIMIX-PLS, which facilitates identifying and treating unobserved heterogeneity by offering guidelines on how to apply the technique to specific research problems. As a follow-up to this paper, Part II features an example explaining how to identify and interpret unobserved heterogeneity in PLS-SEM using the FIMIX-PLS module in the SmartPLS 3 software (Matthews *et al.*, 2016).

What is unobserved heterogeneity and why is it important?

PLS-SEM applications usually analyze a full set of data, implicitly assuming that the data stems from a single homogeneous population (Jedidi *et al.*, 1997). This assumption of relatively homogeneous data characteristics is often unrealistic. Individuals (e.g. in their behavior) or companies (e.g. in their structure) are different, and pooling data across observations is likely to produce misleading results (Sarstedt *et al.*, 2009). Failure to consider such heterogeneity can be a threat to the validity of PLS-SEM results, leading to incorrect conclusions (Becker *et al.*, 2013; Rigdon *et al.*, 2010, 2011; Sarstedt and Ringle, 2010).

The model shown in Figure 1, in which customer satisfaction with a product (Y_3) depends on the two perceptual dimensions – satisfaction with the quality (Y_1) and satisfaction with the price (Y_2) –, illustrates the problems stemming from failure to treat heterogeneity in the context of PLS-SEM. Suppose that there are two segments of equal size: group 1 is quality conscious, whereas group 2 is price conscious, as indicated by the different segment-specific path coefficients. More specifically, the path of quality (Y_1) on satisfaction (Y_3) is much higher in group 1 ($p_{13}^{(1)}$) than in group 2 ($p_{13}^{(2)}$); the superscript in round brackets indicates the group. Similarly, with an absolute difference of 0.40, the path from price (Y_2) to satisfaction (Y_3) is much higher in group 2 ($p_{23}^{(2)}$) compared to group 1 ($p_{23}^{(1)}$). In this example, heterogeneity is the result of one group of customers exhibiting a price preference, while the preference of the second group is quality over price. From a technical perspective, there is a categorical moderator variable that splits the data set into two groups (quality conscious and price conscious customer) and thus requires estimating two separate models, as indicated in Figure 1. Importantly, if we fail to recognize the heterogeneity between the groups and analyze the model using the full set of data, the path coefficients will be substantially biased. That is, both estimates would equal approximately 0.30 when using the full set of data, thus, leading the researcher to conclude that price and quality are equally important for customer satisfaction, although they are not. Consequently, it is important to identify, assess and, if present, treat heterogeneity in the data.

Heterogeneity in data can be observed or unobserved. When differences between two or more groups of data relate to observable characteristics, such as gender, age or country of origin, heterogeneity is observed. Researchers can use these observable characteristics to partition the data into separate groups of observations and carry out group-specific PLS-SEM analyses, as illustrated in Figure 1, with regard to customers' price versus quality consciousness. On the contrary, unobserved heterogeneity emerges when differences between two or more groups of data do not depend on a specific observable characteristic or combinations of several characteristics. To account for unobserved heterogeneity, researchers have routinely used clustering techniques, such as k-means on the indicator data, or latent variable scores derived from a preceding analysis of the aggregate data set.



Source: Hair *et al*. (2014a, 2014b, 2016)

Figure 1. Effects of heterogeneity in PLS-SEM

The partition that this analysis produces is then used as an input for group-specific PLS-SEM estimations. While easy to apply, such an approach is conceptually flawed because traditional clustering techniques ignore the path model relationships that researchers have specified prior to the analysis. However, it is exactly these relationships that are likely responsible for some of the group differences. At the same time, prior research has shown that traditional clustering approaches perform very poorly regarding identifying group differences (Sarstedt and Ringle, 2010). Acknowledging this limitation of sequential approaches, methodological research in PLS-SEM has proposed a multitude of specific methods to identify and treat unobserved heterogeneity, commonly referred to as latent class techniques. These techniques have proven very useful to identify unobserved heterogeneity and partition the data accordingly. Also, latent class techniques may ascertain that unobserved heterogeneity does not influence the results, supporting an analysis of a single model based on the aggregate data level.

FIMIX-PLS

Originally introduced by Hahn *et al.* (2002) and later extended by Sarstedt *et al.* (2011a), FIMIX-PLS is the first and best understood latent class approach to PLS-SEM (Sarstedt, 2008b). As indicated by its name, the approach relies on the finite mixture models concept, which assumes that the overall population is a mixture of group-specific density functions. The aim of FIMIX-PLS is to disentangle the overall mixture distribution and estimate parameters (e.g. the path coefficients) of each group in a regression framework (i.e. mixture regressions; Wedel and Kamakura, 2000). Figure 2 shows an example of a mixture distribution that FIMIX-PLS aims to separate.

To do so, the FIMIX-PLS approach follows two steps. In the first step, the standard PLS-SEM algorithm is run on the full set of data to obtain the scores of all the latent variables in the model. This analysis is done automatically in software programs such as SmartPLS 3, and the user does not have to initiate it manually. The resulting latent variable scores then serve as input for a series of mixture regression analyses in the second step. The mixture regressions allow for the simultaneous probabilistic classification of observations into groups and the estimation of regression models explaining the means and variances of the endogenous latent variables within each of these groups (Wedel and Kamakura, 2000). As such, FIMIX-PLS assumes that

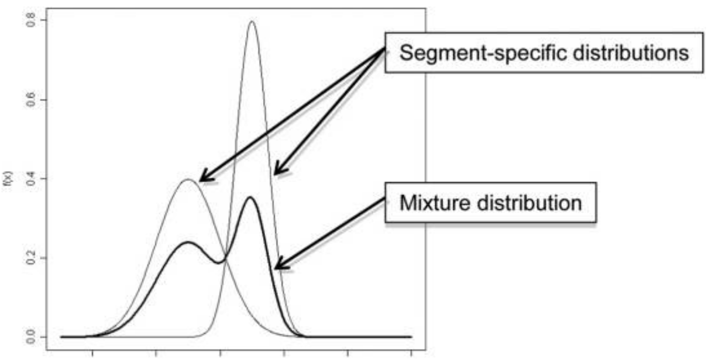


Figure 2.
Mixture distribution example

heterogeneity only occurs in the structural model and that measurement models are invariant across groups (Henseler *et al.*, 2016).

The systematic application of FIMIX-PLS follows a four-step approach, as illustrated in Figure 3. In the following text, we discuss each step in greater detail.

Step 1: run the FIMIX-PLS procedure

Running the FIMIX-PLS procedure requires the researcher to make several choices regarding the algorithm settings. The model estimation in FIMIX-PLS follows the likelihood principle, which asserts that all the evidence in a sample that is relevant for the model parameters is contained in the likelihood function. This likelihood function is maximized by using the expectation-maximization (EM) algorithm. The EM algorithm alternates between performing an expectation (E) step and a maximization (M) step. The E step creates a function for the expectation of the log-likelihood, which is evaluated using the current estimate of the parameters. The M step computes parameters by maximizing the expected log-likelihood found in the E step. The E and M steps are successively applied until the results stabilize. Stabilization is reached when there is no substantial improvement in the log-likelihood value from one iteration to the next. A threshold value of $1 \cdot 10^{-10}$ is recommended as a stop criterion to ensure that the algorithm converges at reasonably low levels of iterative changes in the log-likelihood values. When the stop criterion is set very low, the FIMIX-PLS algorithm may not converge within a reasonable time. Therefore, the researcher also needs to specify a maximum number of iterations after which the algorithm will automatically terminate. Specifying a maximum number of 5,000 iterations ensures a sound balance between warranting acceptable computational running time and obtaining results that are precise enough.

Using the EM algorithm for model estimation is attractive because it is very efficient and always converges to a pre-defined number of segments. However, this convergence may occur in a local optimum, which means that the solution is only optimal compared to similar solutions, but not globally (Steinley, 2003). To investigate the possible occurrence of a local optimum, researchers should run FIMIX-PLS multiple times. The initialization in FIMIX-PLS occurs randomly, which means that each time it is initiated the algorithm uses different randomly selected starting values for parameter estimation. Generally, the results of multiple FIMIX-PLS computations will be very similar. However, if the results are not very similar, then a local optimum has occurred and the

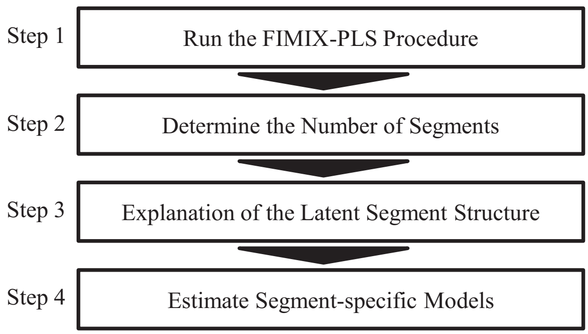


Figure 3.
A systematic procedure for applying FIMIX-PLS

solution should be discarded. In line with simulation study results of the technique (Sarstedt *et al.*, 2011a, 2011b), we suggest using ten repetitions of the FIMIX-PLS algorithm and choosing the solution with the best log-likelihood value. Another consequence of the FIMIX-PLS algorithm's random nature is that the numbering of the segments is not determinate. That is, the results of a certain segment might appear in a segment with a different number when FIMIX-PLS is run again. This characteristic is commonly referred to as label switching (McLachlan and Peel, 2000). Software programs such as SmartPLS 3 address this issue by sorting and labeling the segments based on their relative sizes.

A further important consideration when running FIMIX-PLS involves the treatment of missing values. Kessel *et al.* (2010) have shown that just 5 per cent missing values in one variable causes severe problems in a FIMIX-PLS analysis when they are replaced with the mean of that indicator's valid values (i.e. mean value replacement). In this case, the missing values treatment option creates a set of common scores, which FIMIX-PLS identifies as a distinct homogeneous segment. As a consequence, the number of segments will likely be over specified and observations that truly belong to other segments will be forced into this artificially generated one. Therefore, mean value replacement must not be used in a FIMIX-PLS context, even if there are only very few missing values in the data set. Instead, researchers should remove all cases that include missing values in any of the indicators used in the model from the analysis (i.e. casewise deletion). While this approach also has its problems, particularly when values are missing at random (Sarstedt and Mooi, 2014), it avoids the generation of an artificial segment as is the case with mean value replacement and other imputation methods, such as EM imputation and regression imputation.

Finally, the FIMIX-PLS algorithm needs to be run for alternating numbers of segments, starting with the one-segment solution. As the number of segments is a priori unknown, researchers must compare the solutions with the different segment numbers in terms of their statistical adequacy and interpretability (Henseler *et al.*, 2015; Sarstedt *et al.*, 2014). The range of possible segment numbers depends on the interplay between the sample size and the minimum sample size requirements to reliably estimate the given model. For example, when analyzing a data set with 200 observations and facing a minimum sample size of 50, it is not reasonable to run FIMIX-PLS with more than four segments. Therefore, it is imperative to consider model-specific minimum segment sample size requirements as documented in, for example, Hair *et al.* (2016), before defining a range of segment solutions to consider in the FIMIX-PLS analysis. The theoretical maximum number of segments is given by the largest integer when dividing the sample size n by the minimum sample size n_{min}: $\lfloor n/n_{\min} \rfloor$. However, because it is highly unlikely that the observations are evenly distributed across the segments, especially when the upper bound is high, considering a lower number of segments is generally preferred.

Step 2: determine the number of segments

A fundamental challenge with the application of FIMIX-PLS is determining the number of groups to retain from the data. Identifying a suitable number of groups is crucial, because many managerial decisions are based on this result. As Becker *et al.* (2015) note:

> […] a misspecified number of segments results in under- or oversegmentation, which easily leads to inaccurate management decisions regarding, for example, customer targeting, product positioning, or determining the optimal marketing mix.

Unlike other PLS-SEM-based latent class techniques, FIMIX-PLS allows researchers to compute likelihood-based information criteria, which provide an indication of how many segments to retain from the data. Information criteria simultaneously take into account the fit (i.e. the likelihood) of a model and the number of parameters used to achieve that fit. The information criteria therefore denote a penalized likelihood function. That is, the negative likelihood plus a penalty term, which increases with the number of segments (Sarstedt, 2008a). The smaller the value of a certain information criterion, the better the segmentation solution. Prominent examples of information criteria include Akaike's information criterion (AIC; Akaike, 1973), modified AIC with factor 3 (AIC_3; Bozdogan, 1994), consistent AIC (CAIC; Bozdogan, 1987) and Bayesian information criterion (BIC; Schwarz, 1978). For a formal representation of these criteria see, for example, the study by Sarstedt *et al.* (2011a).

Information criteria are not scaled within a certain range of values (e.g. between 0 and 1). Rather, the criteria may take values in the hundreds or thousands, depending on the starting point of the FIMIX-PLS algorithm, which is set randomly. Importantly, however, each criterion's values can be compared across different solutions with varying segment numbers (provided they are calculated on the same computer). Therefore, the researcher needs to examine several solutions with alternating numbers of segments and select the model that minimizes a particular information criterion.

Sarstedt *et al.* (2011a) have evaluated the efficacy of different information criteria in FIMIX-PLS across a broad range of data and model constellations. Their results demonstrate that researchers should jointly consider AIC_3 and CAIC. Whenever these two criteria indicate the same number of segments, the results likely point to the appropriate number of segments. AIC with factor 4 (AIC_4; Bozdogan, 1994) and BIC generally perform well, while other criteria exhibit a pronounced overestimation tendency. This holds especially for AIC, which often over specifies the correct number of segments by three or more segments. Still, other criteria, such as minimum description length 5 (MDL_5; Liang *et al.*, 1992), show pronounced underestimation tendencies. Researchers can use this information to determine a certain range of reasonable segment numbers. For example, when AIC indicates a five-segment solution, retaining a smaller number of segments seems warranted. Table I provides an overview of selected information criteria and highlights their performance in the context of FIMIX-PLS.

Information criteria are not a silver bullet to determine the most suitable number of segments in FIMIX-PLS, because criteria such as AIC_4 and BIC do not offer any indication of how well separated the segments are. For this reason, researchers should consider the complementary use of entropy-based measures, such as the normed entropy statistic (EN; Ramaswamy *et al.*, 1993). The EN uses the observations' segment membership probabilities to indicate whether the partition is reliable or not. The more observations exhibit high segment membership probabilities, the more clear-cut their segment affiliation is. The EN ranges between 0 and 1, with higher values indicating a better quality partition. Prior research provides evidence that EN values above 0.50 permit a clear-cut classification of data into the pre-determined number of segments (Ringle *et al.*, 2005a, 2010b).

Abbreviation	Criterion name	Performance in FIMIX-PLS
AIC	Akaike's information criterion	Weak performance Very strong tendency to overestimate the number of segments Can be used to determine the upper limit of reasonable segmentation solutions
AIC_3	Modified Akaike's information criterion with factor 3	Fair to good performance Tends to overestimate the number of segments Works well in combination with CAIC and BIC
AIC_4	Modified Akaike's information criterion with factor 4	Good performance Tends to over- and under-estimate the number of segments
BIC	Bayesian information criterion	Good performance Tends to underestimate the number of segments Should be considered jointly with AIC_3
CAIC	Consistent Akaike's information criterion	Good performance Tends to underestimate the number of segments Should be considered jointly with AIC_3
MDL_5	Minimum description length with factor 5	Weak performance Very strong tendency to underestimate the number of segments Can be used to determine the lower limit of reasonable segmentation solutions

Table I. Selected information criteria and their performance in FIMIX-PLS

When deciding on the number of segments to retain, it is particularly important to keep in mind that the EM algorithm always converges to the pre-specified number of segments. However, the result may be that FIMIX-PLS forces a small subset of data into extraneous segments, simply because the researcher specified too high a number of segments in the analysis. Such extraneous segments account for only a marginal portion of heterogeneity in the overall data set and are usually too small to ensure valid group-specific results (Rigdon *et al.*, 2010). Therefore, in addition to information criteria and the EN, the researcher should carefully consider the segment sizes produced by FIMIX-PLS. If the analysis yields an extraneous segment that is too small to warrant valid analysis, the researcher should consider reducing the number of segments or dropping this segment and focusing on the analysis and interpretation of the other, larger segments.

Finally, it is important to note that a purely data-driven approach provides only rough guidance regarding the number of segments to be selected. Heuristics such as information criteria and the EN are fallible because they are sensitive to data and model characteristics. For example, research results by Becker *et al.* (2015) suggest that even low levels of collinearity in the structural model can have adverse consequences for the information criteria's performance. FIMIX-PLS is an exploratory tool and should be treated as such. Consequently, any decision regarding the number of segments should be made on pragmatic grounds and practical considerations should be taken into account (Sarstedt *et al.*, 2009). For example, researchers might have a priori knowledge or a theory on which the choice can be based. Likewise, the number of segments must be small enough to ensure parsimony and manageability, but each segment should also be large enough to warrant strategic attention (Sarstedt and Mooi, 2014).

Step 3: explanation of the latent segment structure

Upon completion of the analysis, FIMIX-PLS provides users with group membership probabilities with respect to each observation, as well as group-specific model estimates, most notably path coefficients. These path coefficients are based on weighted least squares regressions using the segment membership probabilities as an input (Hahn *et al.*, 2002). This means that each observation contributes to the estimation of segment-specific path coefficients, which is different from a situation where observations are grouped into non-overlapping groups (i.e. a hard clustering) and each group is estimated separately. As a result, the initial path coefficient estimates produced by FIMIX-PLS are highly abstract and offer only brief orientation regarding the relationships to expect within each of the groups. Turning the initial FIMIX-PLS results into actionable understanding requires the researcher to interpret the segments in terms of observable and managerially meaningful variables. To do so, researchers need to identify one or more explanatory variable(s) that match the FIMIX-PLS partition in the best possible way (Hahn *et al.*, 2002; Ringle *et al.*, 2010a; Sarstedt and Ringle, 2010).

The analysis, also referred to as the ex post analysis (Hahn *et al.*, 2002), first involves assigning each observation to a certain segment based on the maximum segment membership probabilities. For example, if an observation in a two-segment solution has a 71 per cent probability of belonging to segment 1 and a 29 per cent probability of belonging to segment 2, one would assign this observation to the first segment. This process, also referred to as hard clustering, ensures that each observation is assigned to only one segment (i.e. the segments are disjoint). Next, the researcher needs to partition the data using an explanatory variable, or a combination of several explanatory variables, which yields a grouping of data that largely corresponds to the one produced by FIMIX-PLS.

Researchers have proposed different means to identify suitable explanatory variables. For example, in a related context, Ramaswamy *et al.* (1993) suggest regressing the following adjusted segment membership probabilities on a set of explanatory variables to identify the variable with the strongest impact on the partition solution:

$$\ln\left(p_{ij}\Big/\left(\prod_{i=1}^{s} p_{ij}\right)^{1/s}\right),$$

where, p_{ij} is the probability that observation i ($i = 1, \ldots, n$) belongs to segment j ($j = 1, \ldots, s$). Prior applications of FIMIX-PLS have also relied on cross tabs (Ringle *et al.*, 2010b), classification and regression trees (Ringle *et al.*, 2010a; Sarstedt and Ringle, 2010) and logistic regressions (Money *et al.*, 2012; Wilden and Gudergan, 2015), among others.

Importantly, to successfully run an ex post analysis, researchers must be able to consider a wide range of observable characteristics that can serve as a possible input. Examining a too limited number of possible observable characteristics limits the researcher's ability to reproduce the FIMIX-PLS partition. With this in mind, researchers should assess whether a single explanatory variable, or a set of variables, has theoretical meaning regarding elucidating possible differences in path coefficients across identified segments. Therefore, assessing the explanatory role of possible variables so that FIMIX-PLS can be implemented more completely must already be considered in the research design stage when collecting descriptive or other variables that may matter (Sarstedt *et al.*, 2016).

Nevertheless, reproducing the FIMIX-PLS partition remains a very challenging task, as observable characteristics often do not match the latent segment structures well. Against this background, an overlap of 60 per cent between the FIMIX-PLS partition and the one produced by the explanatory variable(s) can be considered satisfactory. The following summarizes the rules of thumb for FIMIX-PLS analyses:

FIMIX-PLS algorithm settings

- Use a stop criterion of $1 \cdot 10^{-10}$ and a maximum number of 5,000 iterations.
- Use ten repetitions to avoid convergence in a local optimum.
- Use the indicators have missing values. Do not use mean value replacement or any other imputation method.
- To define a range of reasonable segment numbers, use one segment as the lower bound and the largest integer when dividing the sample size by the minimum segment sample size as the upper bound.

Determining the number of segments to retain

- *Information criteria*: If AIC_3 and CAIC indicate the same number of segments, choose this solution. Alternatively, jointly consider AIC_3 and BIC. Also, consider the segment number as indicated by AIC_4 and BIC. Generally, choose fewer segments than indicated by AIC and more segments than indicated by MDL_5.
- *Entropy criterion*: The entropy criterion should preferably be higher than 0.50.
- Ensure that the segment sizes meet the minimum sample size requirements. If not met, reduce the number of segments or discard the extraneous segments and focus on the remaining larger ones.
- Take practical considerations into account. If possible, let a priori information and theory guide the choice. Ensure that the solution is managerially relevant.

Ex post analysis

- Assign each observation to a single segment using the maximum segment membership probabilities.
- Partition the data using an explanatory variable, or a combination of several explanatory variables, which yields a grouping of data that largely corresponds to the one produced by FIMIX-PLS.
- A 60 per cent overlap between the FIMIX-PLS partition and the one produced by the explanatory variable(s) is considered satisfactory.

Step 4: estimate segment-specific models

Once the researcher has identified one or more explanatory variables that match the FIMIX-PLS partition well, the final step is to estimate segment-specific models as indicated by the explanatory variable(s). In doing so, the researcher must ensure that all the model measures meet common quality standards as documented in, for example, the study by Hair *et al.* (2014a, 2016).

These analyses complete the basic FIMIX-PLS application. However, further analyses may involve testing whether the numerical differences between segment-specific path coefficients are also significantly different using a multigroup analysis. Research has provided several approaches to multigroup analysis, which Sarstedt *et al.* (2011b) and Hair *et al.* (2016) discuss in greater detail. Hair *et al.* (2016) recommend using the permutation approach (Chin and Dibbern, 2010; Dibbern and Chin, 2005), which has also been implemented in the SmartPLS 3 software. However, before interpreting the results from a multigroup analysis, researchers must ensure that the measurement models are invariant across the groups. By establishing measurement invariance, researchers can be confident that the group differences in model estimates are not due to the distinctive content and/or meanings of the latent variables across the groups. For example, variations in the structural relationships between the latent variables could stem from the different meanings that the groups' respondents attribute to the phenomena being measured, rather than from true differences in the structural relationships. To test measurement invariance in a PLS-SEM context, researchers should execute the measurement invariance of composite models (MICOM) procedure described by Henseler *et al.* (2016). The MICOM procedure involves three steps, which address the following:

(1) the equality of model parameterization and estimation (configural invariance);

(2) the equality of indicator weights (compositional invariance); and

(3) the equality of composite mean values and variances.

MICOM has also been implemented in SmartPLS 3 software.

Summary and conclusion

Checking for unobserved heterogeneity is important to ensure that the aggregate data level analysis is not substantially biased as a result of two or more unidentified, dissimilar groups within the data set. Unobserved heterogeneity, if present in the data, needs to be identified, assessed and treated, for example, using FIMIX-PLS. The method is applied via a four-step approach in which researchers run the FIMIX-PLS procedure, determine the number of segments, explain the latent segment structure and, finally, estimate the segment-specific models.

Identifying a suitable number of groups to retain from the data is one of the most difficult tasks in the application of FIMIX-PLS. Unlike other PLS-SEM-based latent class techniques, FIMIX-PLS provides the researcher with, for example, likelihood-based information criteria to assist with the selection of the number of segments to avoid under- or over-segmenting the data set. An example in the second article "Identifying and Treating Unobserved Heterogeneity with FIMIX-PLS: Part II – A Case Study" explains how to identify and treat unobserved heterogeneity in PLS-SEM using the FIMIX-PLS module in SmartPLS 3 software (Matthews *et al.*, 2016).

References

Akaike, H. (1973), "Information theory and an extension of the maximum likelihood principle", in Petrov, B.N. and Csáki, F. (Eds), *Second International Symposium on Information Theory*, Académiai Kiadó, Budapest, pp. 267-281.

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.

Becker, J.-M., Ringle, C.M., Sarstedt, M. and Völckner, F. (2015), "How collinearity affects mixture regression results", *Marketing Letters*, Vol. 26 No. 4, pp. 643-659.

Bozdogan, H. (1987), "Model selection and Akaike's Information Criterion (AIC): the general theory and its analytical extensions", *Psychometrika*, Vol. 52 No. 3, pp. 345-370.

Bozdogan, H. (1994), "Mixture-model cluster analysis using model selection criteria in a new information measure of complexity", in Bozdogan, H. (Ed.), *Proceedings of the First US/Japan Conference on Frontiers of Statistical Modelling: An Information Approach*, Kluwer Academic Publishers, Dordrecht, Boston, London, pp. 69-113.

Chin, W.W. and Dibbern, J. (2010), "A permutation based procedure for multi-group PLS analysis: results of tests of differences on simulated data and a cross cultural analysis of the sourcing of information system services between Germany and the USA", 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. 2, Springer, Heidelberg, Dordrecht, London, New York, NY, pp. 171-193.

Dibbern, J. and Chin, W.W. (2005), "Multi-group comparison: testing a PLS model on the sourcing of application software services across Germany and the USA using a permutation based algorithm", in Bliemel, F.W., Eggert, A., Fassott, G. and Henseler, J. (Eds), *Handbuch PLS-Pfadmodellierung, Methode, Anwendung, Praxisbeispiele*, Schäffer-Poeschel, Stuttgart, pp. 135-160.

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., Ringle, C.M. and Sarstedt, M. (2011), "PLS-SEM: indeed a silver bullet", *The Journal of Marketing Theory and Practice*, Vol. 19 No. 2, pp. 139-152.

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

Hair, J.F., Sarstedt, M., Ringle, C.M. and Mena, J.A. (2012b), "An assessment of the use of partial least squares structural equation modeling in marketing research", *Journal of the Academy of Marketing Science*, Vol. 40 No. 3, pp. 414-433.

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. (2016), *A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM)*, 2nd edition, 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.

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.

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.

Kaufmann, L. and Gaeckler, J. (2015), "A structured review of partial least squares in supply chain management research", *Journal of Purchasing and Supply Management*, Vol. 21 No. 4, pp. 259-272.

Kessel, F., Ringle, C.M. and Sarstedt, M. (2010), "On the impact of missing values on model selection in FIMIX-PLS", *Proceedings of the 2010 INFORMS Marketing Science Conference*, *Cologne*.

Lee, L., Petter, S., Fayard, D. and Robinson, S. (2011), "On the use of partial least squares path modeling in accounting research", *International Journal of Accounting Information Systems*, Vol. 12 No. 4, pp. 305-328.

Liang, Z., Jaszak, R.J. and Coleman, R.E. (1992), "Parameter estimation of finite mixtures using the EM algorithm and information criteria with applications to medical image processing", *IEEE Transactions on Nuclear Science*, Vol. 39 No. 4, pp. 1126-1133.

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

Matthews, L., Sarstedt, M., Hair, J.F. and Ringle, C.M. (2016), "Identifying and treating unobserved heterogeneity with FIMIX-PLS: part II – a case study", *European Business Review*, Vol. 28 No. 2.

Money, K.G., Hillenbrand, C., Henseler, J. and Da Camara, N. (2012), "Exploring unanticipated consequences of strategy amongst stakeholder segments: the case of a European revenue service", *Long Range Planning*, Vol. 45 Nos 5/6, pp. 395-423.

Peng, D.X. and Lai, F. (2012), "Using partial least squares in operations management research: a practical guideline and summary of past research", *Journal of Operations Management*, Vol. 30 No. 6, pp. 467-480.

Ramaswamy, V., DeSarbo, W.S., Reibstein, D.J. and Robinson, W.T. (1993), "An empirical pooling approach for estimating marketing mix elasticities with PIMS data", *Marketing Science*, Vol. 12 No. 1, pp. 103-124.

Rapp, A., Trainor, K.J. and Agnihotri, R. (2010), "Performance implications of customer-linking capabilities: examining the complementary role of customer orientation and CRM technology", *Journal of Business Research*, Vol. 63 No. 11, pp. 1229-1236.

Rigdon, E.E., Ringle, C.M. and Sarstedt, M. (2010), "Structural modeling of heterogeneous data with partial least squares", in Malhotra, N.K. (Ed.), *Review of Marketing Research*, Sharpe, Armonk, NY, pp. 255-296.

Rigdon, E.E., Ringle, C.M., Sarstedt, M. and Gudergan, S.P. (2011), "Assessing heterogeneity in customer satisfaction studies: across industry similarities and within industry differences", *Advances in International Marketing*, Vol. 22, pp. 169-194.

Ringle, C.M., Sarstedt, M. and Mooi, E.A. (2010a), "Response-based segmentation using finite mixture partial least squares: theoretical foundations and an application to American customer satisfaction index data", *Annals of Information Systems*, Vol. 8, pp. 19-49.

Ringle, C.M., Sarstedt, M. and Straub, D.W. (2012), "A critical look at the use of PLS-SEM in MIS quarterly", *MIS Quarterly*, Vol. 36 No. 1, pp. 3-15.

Ringle, C.M., Wende, S. and Will, A. (2005a), "Customer segmentation with FIMIX-PLS", in Aluja, T., Casanovas, J., Esposito Vinzi, V., Morineau, A. and Tenenhaus, M. (Eds), *PLS and Related Methods: Proceedings of the PLS'05 International Symposium*. *Decisia*, *Paris*, pp. 507-514.

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

Ringle, C.M., Wende, S. and Will, A. (2010b), "Finite mixture partial least squares analysis: methodology and numerical examples", 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. 2), Springer, Heidelberg, Dordrecht, London, NY, pp. 195-218.

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

Sarstedt, M. (2008a), "Market segmentation with mixture regression models", *Journal of Targeting, Measurement and Analysis for Marketing*, Vol. 16 No. 3, pp. 228-246.

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

Sarstedt, M., Becker, J.-M., Ringle, C.M. and Schwaiger, M. (2011a), "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., Henseler, J. and Ringle, C.M. (2011b), "Multi-group analysis in partial least squares (PLS) path modeling: alternative methods and empirical results", in Sarstedt, M., Schwaiger, M. and Taylor, C.R. (Eds), *Advances in International Marketing*, Emerald, Bingley, Vol. 22, pp. 195-218.

Sarstedt, M. and Mooi, E.A. (2014), *A Concise Guide to Market Research: The Process, Data, and Methods Using IBM SPSS Statistics*, Springer, Berlin.

Sarstedt, M., Ringle, C.M. and Gudergan, S.P. (2016), "Guidelines for treating unobserved heterogeneity in tourism research: a comment on Marques and Reis (2015)", *Annals of Tourism Research*, in press.

Sarstedt, M. and Ringle, C.M. (2010), "Treating unobserved heterogeneity in PLS path modelling: a comparison of FIMIX-PLS with different data analysis strategies", *Journal of Applied Statistics*, Vol. 37 No. 8, pp. 1299-1318.

Sarstedt, M., Ringle, C.M., Smith, D., Reams, R. and Hair, J.F. (2014), "Partial least squares structural equation modeling (PLS-SEM): a useful tool for family business researchers", *Journal of Family Business Strategy*, Vol. 5 No. 1, pp. 105-115.

Sarstedt, M., Schwaiger, M. and Ringle, C.M. (2009), "Do we fully understand the critical success factors of customer satisfaction with industrial goods? - Extending Festge and Schwaiger's model to account for unobserved heterogeneity", *Journal of Business Market Management*, Vol. 3 No. 3, pp. 185-206.

Schwarz, G. (1978), "Estimating the dimensions of a model", *Annals of Statistics*, Vol. 6 No. 2, pp. 461-464.

Steinley, D. (2003), "Local optima in K-means clustering: what you don't know may hurt you", *Psychological Methods*, Vol. 8 No. 3, pp. 294-304.

Wedel, M. and Kamakura, W.A. (2000), *Market Segmentation: Conceptual and Methodological Foundations*, Kluwer, Boston, MA.

Wilden, R. and Gudergan, S. (2015), "The impact of dynamic capabilities on operational marketing and technological capabilities: investigating the role of environmental turbulence", *Journal of the Academy of Marketing Science*, Vol. 43 No. 2, pp. 181-199.

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. Siham Mourad, Pierre Valette-Florence. 2016. Improving prediction with POS and PLS consistent estimations: An illustration. *Journal of Business Research* **69**:10, 4675-4684. [CrossRef]
5. MatthewsLucy M. Lucy M. Matthews SarstedtMarko Marko Sarstedt HairJoseph F. Joseph F. Hair RingleChristian M. Christian M. Ringle Department of Marketing, Middle Tennessee State University, Murfreesboro, Tennessee, USA Faculty of Economics and Management, Otto-von-Guericke-University, Magdeburg, Germany, and Faculty of Business and Law, University of Newcastle, Newcastle, Australia Coles College of Business, Kennesaw State University, Kennesaw, Georgia, USA 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. *European Business Review* **28**:2, 208-224. [Abstract] [Full Text] [PDF]