Georgia State University

ScholarWorks @ Georgia State University

Computer Information Systems Faculty Publications | Department of Computer Information Systems

9-2013

Discovering Unobserved Heterogeneity in Structural Equation Models to Avert Validity Threats

Jan-Michael Becker
University of Cologne, j.becker@wiso.uni-koeln.de

Arun Rai
Georgia State University, arunrai@gsu.edu

Christian M. Ringle
University of Newcastle - Australia, christian.ringle@newcastle.edu.au

Franziska Völckner
University of Cologne, voelckner@wiso.uni-koeln.de

Follow this and additional works at: https://scholarworks.gsu.edu/cis_facpub

Part of the Management Information Systems Commons

Recommended Citation

Becker, J-M., Rai, A., Voleckner, F., and Ringle, C., Discovering Unobserved Heterogeneity in Structural Equation Models, MIS Quarterly, September 2013, 37(3), 665-694. http://misq.org/discovering-unobserved-heterogeneity-in-structural-equation-models-to-avert-validity-threats.html.

This Article is brought to you for free and open access by the Department of Computer Information Systems at ScholarWorks @ Georgia State University. It has been accepted for inclusion in Computer Information Systems Faculty Publications by an authorized administrator of ScholarWorks @ Georgia State University. For more information, please contact scholarworks@gsu.edu.

MIS Quarterly

RESEARCH ESSAY

DISCOVERING UNOBSERVED HETEROGENEITY IN STRUCTURAL EQUATION MODELS TO AVERT VALIDITY THREATS[1]

Jan-Michael Becker
Department of Marketing and Brand Management, University of Cologne,
Cologne, 50923, GERMANY {j.becker@wiso.uni-koeln.de}

Arun Rai
Center for Process Innovation and Department of Computer Information Systems, Robinson College of Business,
Georgia State University, Atlanta, GA 30303 U.S.A. {arunrai@gsu.edu}

Christian M. Ringle
Institute for Human Resource Management and Organizations, Hamburg University of Technology (TUHH),
Hamburg, 21073, GERMANY {ringle@tuhh.de} and
Faculty of Business and Law, University of Newcastle, Callaghan, NSW 2308 AUSTRALIA {christian.ringle@newcastle.edu.au}

Franziska Völckner
Department of Marketing and Brand Management, University of Cologne,
Cologne, 50923, GERMANY {voelckner@wiso.uni-koeln.de}

A large proportion of information systems research is concerned with developing and testing models pertaining to complex cognition, behaviors, and outcomes of individuals, teams, organizations, and other social systems that are involved in the development, implementation, and utilization of information technology. Given the complexity of these social and behavioral phenomena, heterogeneity is likely to exist in the samples used in IS studies. While researchers now routinely address observed heterogeneity by introducing moderators, a priori groupings, and contextual factors in their research models, they have not examined how unobserved heterogeneity may affect their findings. We describe why unobserved heterogeneity threatens different types of validity and use simulations to demonstrate that unobserved heterogeneity biases parameter estimates, thereby leading to Type I and Type II errors. We also review different methods that can be used to uncover unobserved heterogeneity in structural equation models. While methods to uncover unobserved heterogeneity in covariance-based structural equation models (CB-SEM) are relatively advanced, the methods for partial least squares (PLS) path models are limited and have relied on an extension of mixture regression—finite mixture partial least squares (FIMIX-PLS) and distance measure-based methods—that have mismatches with some characteristics of PLS path modeling. We propose a new method—prediction-oriented segmentation (PLS-POS)—to overcome the limitations of FIMIX-PLS and other distance measure-based methods and conduct extensive simulations to evaluate the ability of PLS-POS and FIMIX-PLS to discover unobserved heterogeneity in both structural and measurement models. Our results show that both PLS-POS and FIMIX-PLS perform

[1]Ron Thompson was the accepting senior editor for this paper. Ron Cenfetelli served as the associate editor.

The appendices for this paper are located in the "Online Supplements" section of the *MIS Quarterly*'s website (http://www.misq.org).

well in discovering unobserved heterogeneity in structural paths when the measures are reflective and that PLS-POS also performs well in discovering unobserved heterogeneity in formative measures. We propose an unobserved heterogeneity discovery (UHD) process that researchers can apply to (1) avert validity threats by uncovering unobserved heterogeneity and (2) elaborate on theory by turning unobserved heterogeneity into observed heterogeneity, thereby expanding theory through the integration of new moderator or contextual variables.

Keywords: Unobserved heterogeneity, validity, structural equation modeling, partial least squares, formative measures, prediction-oriented segmentation

Introduction

Assuming that data in empirical studies are homogeneous and represent a single population is often unrealistic in the social and behavioral sciences, such as information systems, management, and marketing (Rust and Verhoef 2005; Wedel and Kamakura 2000). There may be significant heterogeneity in the data across unobserved groups, and it can bias parameter estimates, lead to Type I and Type II errors, and result in invalid conclusions (Jedidi et al. 1997). Consider the following technology acceptance model (TAM) example: A researcher is interested in individuals' *intention to use* an IT system or service (Davis et al. 1989; Venkatesh 2000; Venkatesh and Davis 2000; Venkatesh et al. 2003). Informed by existing theory, the researcher proposes a model in which *perceived usefulness* (PU) and *perceived ease of use* (PEOU) of the IT system explain *intention to use the system* (IU) (Figure 1). The empirical results reveal that PU and PEOU are equally important in explaining IU. However, the theory and model overlook the two underlying groups: experienced IT users (Figure 1a, segment 1) and inexperienced IT users (Figure 1a, segment 2). Experienced users show a strong positive relationship between PU and IU and a weak, or nonsignificant, relationship between PEOU and IU. In contrast, inexperienced users show a strong positive relationship between PEOU and IU and a weak, or nonsignificant, relationship between PU and IU (Figure 1a). In this scenario, drawing inferences based on results from the overall sample would lead to Type I errors as we would be overgeneralizing the significant findings from the overall sample to the underlying user groups, one with a nonsignificant estimate for PEOU→IU and the other with a nonsignificant estimate for PU→IU. If the model is not refined to accommodate this unobserved heterogeneity, a system that is unsuitable for either user group (i.e., one with average usefulness and average ease of use) may be provided to all users.

In addition, a study may not find PEOU to be a significant predictor of IU because of unobserved heterogeneity across two groups of users (i.e., experienced versus inexperienced). If experienced users (Figure 1b, segment 1) perceive an easy-to-use system (i.e., high PEOU) as being too simple to fulfill their needs, they may show a strong negative relationship between PEOU and IU. In contrast, if inexperienced users (Figure 1b, segment 2) show a strong positive relationship between PEOU and IU, as in the first example, a sign reversal occurs between the two groups with regard to the effect of PEOU on IU, thereby leading to an overall nonsignificant effect of PEOU on IU and a Type II error.

Recent TAM models acknowledge existing heterogeneity by incorporating experience as a moderator of PEOU's effect on IU. However, before its inclusion in the theory, experienced versus inexperienced users represented unobserved heterogeneity that could lead to biased findings on the effects of PU and PEOU on IU. This illustration shows how not accounting for unobserved heterogeneity can lead to misinterpretations and invalid conclusions in IS research—a point we emphasize later in the paper based on a review of 12 meta-analysis studies on key IS phenomena (see Table A1 in Appendix A).

Despite the threats to validity from unobserved heterogeneity, there are important gaps in the IS literature about the specific threats to validity and how to safeguard against them.

(1) While IS studies now routinely address observed heterogeneity by introducing moderators, *a priori* groupings, contextual factors, and control variables in their research models, they have not considered unobserved heterogeneity in their data. In fact, none of the papers appearing in the field's two most widely recognized journals (*MIS Quarterly* and *Information Systems Research*) over the last 20 years that have developed and tested structural equation models have examined unobserved heterogeneity. Our first research objective is to introduce the concept of unobserved heterogeneity in the IS literature and to show how IS researchers can safeguard against biases and facilitate theory development.

(2) While research in some fields notes that unobserved heterogeneity threatens empirical results and their interpretation, a systematic analysis of the threats to specific

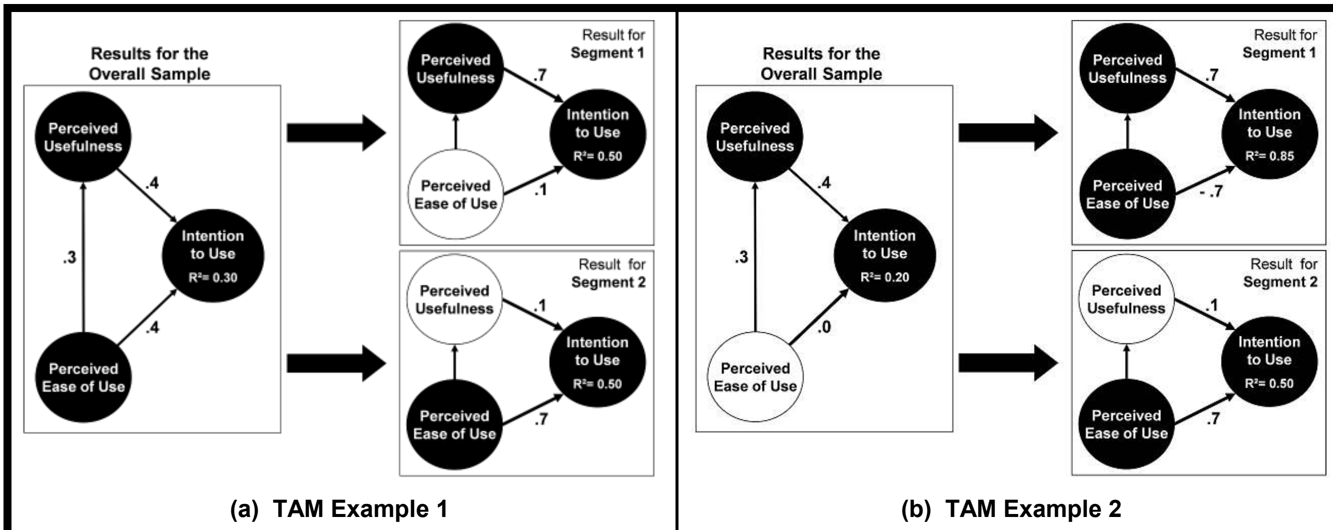


Figure 1. Examples for Unobserved Heterogeneity in TAM

types of validity is missing in the literature. Our second research objective is to evaluate the implications of unobserved heterogeneity for four types of validity (i.e., instrument, internal, statistical conclusion, and external validity; Cook and Campbell 1976, 1979; Straub 1989), thereby broadening our understanding of the specific validity threats that arise from unobserved heterogeneity.

(3) In structural equation modeling (SEM), unobserved heterogeneity is not only a validity threat for the structural model but also for the measurement model regardless of whether the measures are reflective or formative. While heterogeneity in reflective measures has been discussed in terms of measurement equivalence or invariance (ME/I) (e.g., Steenkamp and Baumgartner 1998; Vandenberg and Lance 2000), the implications of unobserved heterogeneity for formative measures have not been examined. Our third research objective is to evaluate the implications of unobserved heterogeneity for formative measures.

(4) In contrast to covariance-based SEM (CB-SEM; e.g., Jöreskog 1978, 1982), research on partial least squares (PLS) path modeling (e.g., Chin 1998; Lohmöller 1989; Wold 1982) has paid limited attention to unobserved heterogeneity. Only recently has a method been proposed to detect unobserved heterogeneity in PLS path models: finite mixture partial least squares (FIMIX-PLS; Hahn et al. 2002; Sarstedt and Ringle 2010). However, FIMIX-PLS does not account for heterogeneity in the measurement model and assumes multivariate normal distributions for latent variables. Furthermore, there is limited evidence of this method's performance in discovering unobserved heterogeneity. Our fourth research objective is to propose and evaluate a new method, PLS prediction-oriented segmentation (PLS-POS), which does not follow distributional assumptions and uncovers unobserved heterogeneity not only in the structural model but also in the measurement model.

(5) Researchers facing the problem of unobserved heterogeneity in their empirical work lack guidelines on how to apply methods systematically to uncover unobserved heterogeneity. Therefore, our fifth research objective is to develop an unobserved heterogeneity discovery (UHD) process to guide researchers in applying methods to ensure the validity of findings and to elaborate theory by turning unobserved heterogeneity into observed heterogeneity.

By addressing the above research objectives, we make six contributions. First, we provide evidence and reasoning for why unobserved heterogeneity is an important issue in IS research. Second, we demonstrate that unobserved heterogeneity in SEM has implications not only for the structural model but also for measurement models. Third, we identify the implications of unobserved heterogeneity for different types of validity and surface the importance of uncovering unobserved heterogeneity to avoid validity threats. Fourth, we introduce the new PLS-POS method for detecting unobserved heterogeneity. This method is specifically developed to fit PLS path modeling, as it employs a prediction-oriented and nonparametric approach and uncovers heterogeneity in both the structural model and the (formative) measurement models and thereby overcomes the limitations of FIMIX-PLS and other distance measure-based methods. Fifth, we evaluate FIMIX-PLS and PLS-POS using an extensive simulation study and generate important insights into the performance of the two methods in uncovering unobserved heterogeneity in PLS path models. Sixth, we provide a UHD process to guide researchers in discovering and addressing unobserved heterogeneity in structural equation models.

Concept of Heterogeneity and its Treatment in IS Research

Researchers can obtain different parameter estimates when they consider differences among observations relative to when they overlook them. However, heterogeneity among observations is not necessarily captured by variables that are preconceived by the researcher and specified by existing theory, as it can exist beyond these previously identified variables (Jedidi et al. 1997). As a consequence, it is necessary to differentiate between the following two types of heterogeneity: (1) *observed heterogeneity* when subpopulations are defined *a priori* based on known variables and (2) *unobserved heterogeneity* when the subpopulations in the data are unknown (Lubke and Muthén 2005).

Observed Heterogeneity

Observed heterogeneity occurs when differences in parameter estimates between groups are expected *a priori* for the phenomenon—that is, when group differences are explained by existing theory that incorporates moderators or contextual factors. Examples of such moderators or contextual factors considered in IS research include individual cultural differences (e.g., individualism versus collectivism; Srite and Karahanna 2006), individual demographic differences (e.g., gender, income levels, and education; Hsieh et al. 2008; Venkatesh et al. 2003), and organizational demographic differences (e.g., large versus small firms; Rai et al. 2006). In our TAM example from earlier, existing theory expects gender-based heterogeneity in structural paths (i.e., men are expected to have a stronger relationship between PU and IU, and women are expected to have a stronger relationship between PEOU and IU) (e.g., Venkatesh and Morris 2000). Moreover, existing theory expects contextual variables, such as voluntariness or task type (e.g., Venkatesh and Davis 2000), or psychographic variables, such as personal innovativeness and computer attitude, to cause heterogeneity in the relationships among the TAM constructs (e.g., Venkatesh and Bala 2008).

Unobserved Heterogeneity

When theory does not assume heterogeneity even though it exists or when theory indicates heterogeneity but the specified group variables do not sufficiently capture it in the population, unobserved heterogeneity occurs. In such situations, researchers need to uncover unobserved heterogeneity by segmenting data to form homogenous groups. If the differences uncovered by segmentation can be explained *post hoc* using contextual or demographic variables (e.g., culture, gender, experience, etc.) making the groups accessible, theory can be expanded accordingly, and unobserved heterogeneity is turned into observed heterogeneity for future studies. If the differences cannot be explained by well-known contextual variables, the researcher has to consider complementary theoretical explanations for the phenomenon.

Treatment of Heterogeneity in IS Research

Given the complexity of the social and behavioral phenomena tackled in IS research, heterogeneity is likely to exist in samples that are used to develop, test, and refine models. If this heterogeneity is not uncovered and controlled, the (unobserved) heterogeneity can bias results and conclusions (e.g., Ansari et al. 2000; Johns 2006). Consequently, unobserved heterogeneity is receiving increasing attention in related disciplines (e.g., marketing, where scholars study similar complex phenomena pertaining to consumer choices and preferences, the alignment of firm-level marketing strategies, interorganizational relationships, and the business value of tangible and intangible resources) to safeguard against biases and probe the underlying reasons for unobserved heterogeneity (e.g., Rigdon et al. 2010). This enhances the likelihood of obtaining valid results as well as of generating greater theoretical contributions. Methodologists in marketing, econometrics, and psychology have proposed advances to uncover unobserved heterogeneity in various approaches—for instance, regression analysis (DeSarbo and Cron 1988; Späth 1979; Wedel and DeSarbo 1994), CB-SEM (e.g., Ansari et al. 2000; Jedidi et al. 1997; Muthén 1989), panel data models (e.g., Allenby and Rossi 1998; Popkowski Leszczyc and Bass 1998), and conjoint analysis (e.g., DeSarbo et al. 1995; Gilbride et al. 2006; Lenk et al. 1996).

While IS studies now routinely address observed heterogeneity by introducing moderators, *a priori* groupings, contextual factors, and control variables in their research models, they have not examined threats to validity due to unobserved heterogeneity. Our review of 12 meta-analysis studies that synthesize the findings of empirical research across various IS phenomena (e.g., technology acceptance, IT investment payoff, IT innovation adoption, IS implementation success, and group support systems) reveals that all of them identify inconsistent, conflicting, or mixed findings; "heterogeneity of effect sizes" (Wang and Keil 2007, p. 9); "wide variation in the predicted effects" (King and He 2006, p. 740); and "correlations that vary across studies more than would be produced by sampling error" (Wu and Lederer 2009, p. A6) (see Table A1 in Appendix A). Most of these 12 meta-analysis studies note that these inconsistencies may be caused by the omission of key contextual variables or moderators. However, investigating the known moderators or contextual variables controls for observed heterogeneity (Haenlein and Kaplan 2011), but as long as these moderators and contextual variables are not specified in theory, population heterogeneity will remain unobserved and threatens model validity. (In the next section, we discuss how unobserved heterogeneity biases estimates and causes Type I and II errors.) Furthermore, uncovering unobserved heterogeneity at the study level accelerates the theory-development cycle by generating insights into relationships among constructs (Edmondson and McManus 2007). In a later section, we describe a UHD process where uncovering unobserved heterogeneity facilitates abduction (by raising the possibilities of rival explanations not previously considered; Van de Ven 2007), directing researchers to identify variables that account for unobserved heterogeneity and, through this process, make segments accessible and turn unobserved heterogeneity into observed heterogeneity (e.g., by discovering moderators and grouping variables). This introduction of constructs to capture formerly unobserved heterogeneity revises models and theoretical explanations, making it possible for the revised models to be tested in future research.

Effects of Heterogeneity on Structural Equation Models

Unobserved Heterogeneity in the Structural Model

In the context of SEM, heterogeneity can affect the structural model, the measurement model (formative and reflective), or both (e.g., Ansari et al. 2000; Qureshi and Compeau 2009). Unobserved heterogeneity can influence path coefficients in the structural model because the parameter estimates are determined based on the overall sample, which pools observations across the underlying (unobserved) groups. As a result, researchers may encounter the following biases: (1) biased parameter estimates of structural paths, (2) nonsignificant estimates at the group level becoming significant at the overall sample level that combines (unobserved) groups, (3) sign differences in the parameter estimates across (unobserved) groups being masked as nonsignificant results at the overall sample level that combines (unobserved) groups, and (4) decreased predictive power of the model (R^2 of the endogenous variables). These biases can lead to Type I and Type II errors and invalid inferences.

To substantiate that these biases occur due to unobserved heterogeneity, we conducted a simulation of a PLS path model with the following three situations with two unobserved groups: (1) the parameter estimates across the groups have the same sign but differ in absolute values, (2) the parameter estimates across the groups have opposite signs, and (3) the parameter estimates are nonsignificant for one group but significant for the other. Table 1 summarizes the findings (see Appendix D for details).

The results show that unobserved heterogeneity biases the parameter estimates, decreases the R^2, and increases the risk of Type I and Type II errors. Specifically, in all three simulated situations, biases in the parameter estimates distort effect sizes and cause misinterpretation of the parameter values, which is especially problematic for comparative hypotheses (e.g., path coefficient 1 > path coefficient 2). When the group-specific parameters show inconsistent signs (i.e., situation 2 in which signs are reversed across the groups) and when one of the groups involves nonsignificant parameters, while the other does not (i.e., situation 3), Type I and Type II errors are exacerbated by the following: (1) If a researcher overlooks unobserved heterogeneity and there is a significant non-zero relationship between the constructs as the overall sample estimate, this researcher is incorrectly overgeneralizing the significant relationship that exists in the first segment, thereby leading to a Type I error with respect to the second segment.[2] (2) If a researcher overlooks unobserved heterogeneity and obtains a nonsignificant relationship between the constructs as the overall sample estimate, this researcher may overgeneralize the nonsignificant finding, which exists only in the second segment, thereby leading to a Type II error with respect to the first segment. In contrast, when all parameters are significant and show the same sign (situation 1), it is unlikely that Type II errors will occur: in this situation the occurrence of Type II errors depends on the effect size and the degree to which the increase in standard errors due to unobserved heterogeneity is compensated by the increased power of the larger sample size due to combining the groups. The R^2 decreases in all situations, implying an

[2]This does not mean that there will be a Type I error in general (i.e., for both segments) but only with respect to segment 2 where the true effect is zero. To be specific, the overall sample estimate cannot show a significant non-zero relationship because of unobserved heterogeneity when all segments have a true zero relationship.

Table 1. Conclusions from the Simulation Study on Heterogeneity Effects

True Group Parameters (heterogeneity is uncovered)				Overall Parameter Estimates (heterogeneity is not uncovered)				
	Situation	**Group 1**	**Group 2**	**Biased**	**Type I Error**	**Type II Error**	**Lower R^2**	**Explanation for Type I and Type II Errors**
1.	Significant in all groups with consistent signs	+	+	**Yes**	×	*Depends*	**Yes**	Increase in standard errors vs. increased sample size
		–	–	**Yes**	×	*Depends*	**Yes**	
2.	Significant in all groups with inconsistent signs	–	+	**Yes**	×	**Likely**	Yes	Effects cancel each other
3.	Significant in some groups but not in others	+ / –	0	**Yes**	**Likely**	**Likely**	Yes	Depends on the effect size

Notes: + = significantly positive; – = significantly negative; 0 = nonsignificant; ×= not possible.

inferior model fit to the overall sample: the decrease in R^2 is greater when group-specific effect sizes are high; however, R^2 is almost unaffected when the group-specific effects are low.

Unobserved Heterogeneity in the Measurement Model

Measurement model specification requires the consideration of the nature of the relationship between constructs and measures. There are two types of measurement models: reflective and formative measures (Diamantopoulos and Winklhofer 2001; Jarvis et al. 2003). In reflective measures, changes in the construct are reflected in changes in all of its indicators, and the direction of causality is from the construct to the indicators. Reflective indictors are assessed in terms of their loadings, which entails the simple correlation between the indicator and the construct. In formative measures, the indicators do not reflect the underlying construct but are combined to form it without any assumptions about the intercorrelation patterns among them. The direction of causality is from the indicators to the construct and the weights of formative indicators represent the importance of each indicator in explaining the variance of the construct (Edwards and Lambert 2007; Petter et al. 2007; Wetzels et al. 2009).

Unobserved heterogeneity can lead to differences between measurement model weights and loadings across groups. If the construct's measures are reflective, unobserved heterogeneity may result in different loadings when respondents across groups interpret and respond to measures differently or when they provide information with different degrees of accuracy (Ansari et al. 2000). Thus, when reflective measures are not equivalent across groups, ME/I is not established (e.g., Steenkamp and Baumgartner 1998; Vandenberg and Lance 2000). In this case, the construct does not capture the same theoretical meaning across groups, implying that differences in the construct's relationships with other constructs cannot be compared across groups. That is, the group-specific parameters are only interpretable at the group level, and the data should not be pooled across groups. For example, when considering reflective measures of PU, users' understanding of usefulness can differ significantly across groups. If this is the case, one cannot combine the groups into an overall sample because the construct measured does not capture the same meaning across groups. The relationship between PU and other constructs would be biased as a result of the absence of invariant measurement. However, the lack of ME/I arising from heterogeneity provides valuable information that structural parameters should not be compared between groups and that the data across the groups should not be combined. As such, ignoring the heterogeneity and interpreting results based on the overall sample would lead to invalid conclusions.

In contrast, when a construct's measures are formative, unobserved heterogeneity can lead to differences in the formative indicators' weights across groups. While recent research has discussed ME/I in formative measures (Diamantopoulos and Papadopoulos 2010), it is important to uncover formative indicator weight differences due to unobserved heterogeneity in order to avoid ambiguous interpretations. Formative indicators cause variance in the construct and can be interpreted as actionable attributes of a construct. The weights of formative indicators represent the relative importance of the construct's different facets. Therefore, the problems associated with unobserved heterogeneity in formative measures are similar to those that occur in the structural model. Consequently, ignoring differences in formative indicator weights due to unobserved heterogeneity can bias parameter estimates and lead to Type I and Type II errors. Thus, when researchers find formative indicator weights to be unstable and nonsignificant, in addition to exploring multicollinearity (Cenfetelli and Bassellier 2009), they should also explore unobserved heterogeneity.

As an example, assume that service quality (SERVQUAL) is measured using the following five formative indicators: (1) tangibles, (2) reliability, (3) assurance, (4) empathy, and (5) responsiveness (e.g., Cenfetelli and Bassellier 2009; Collier and Bienstock 2009; Parasuraman et al. 1988). Some customers might favor the communication facets (e.g., empathy and responsiveness) when they evaluate service quality, while others might favor the trust facets (e.g., assurance and reliability) in their evaluation. These differences in customer perceptions result in different measurement weights across the groups although the underlying theoretical construct of service quality remains the same. For example, two equally sized groups have measurement weights of w_{g1} = [.6, .6, .6, .0, .0] for a certain formative construct in one group and w_{g2} = [.2, .2, .2, .6, .6] in the other group. Combining these two groups in the overall sample results in equal relative importance (weights) for all indicators with measurement weights of w = [.4, .4, .4, .3, .3] for the overall sample. As a consequence, the interpretation of the weights estimated using the overall sample is misleading, and the formative measures based on the overall sample represent neither the first group nor the second. Given this bias in the formative measures for service quality, the relationship between service quality and other constructs (e.g., customer satisfaction) is also likely to be biased.

Implications of Unobserved Heterogeneity for Model Validity

If unobserved heterogeneity characterizes the data and results are based on the overall sample, the estimated model lacks validity because it will not uncover the true effects of the underlying groups. In a broad sense, validity is the extent to which a method (i.e., the design, the model, or the construct) measures what it claims to measure. We elaborate on why unobserved heterogeneity affects the major types of validity—(1) *internal*, (2) *instrumental* (including *content*, *construct*, and *criterion validity* and *reliability*), (3) *statistical conclusion*, and (4) *external* (e.g., Cook and Campbell 1976, 1979; Heeler and Ray 1972; Straub 1989). See Table 2 for definitions of each type of validity and explanations of how unobserved heterogeneity threatens it.

Unobserved heterogeneity is a threat to *internal validity* because contextual or group variables that affect results are overlooked, thereby resulting in an incomplete model. The observations across the 12 meta-analyses that we discussed earlier show that inconsistent findings arise when contextual or group variables are omitted. Uncovering these variables and improving theory through the discovery of unobserved heterogeneity safeguards against internal validity threats.

In addition, unobserved heterogeneity threatens *statistical conclusion validity*. Analyzing the overall sample without accounting for heterogeneity increases standard errors and reduces (averages) effect sizes, thereby biasing estimates and leading to Type I and Type II errors. (The simulations in the previous section show how statistical conclusion validity is threatened by unobserved heterogeneity.)

Our earlier discussion of unobserved heterogeneity shows that it can bias the measurement model estimates of constructs, thereby adversely affecting *instrument validity*. There is a particular threat to *reliability* (internal consistency) when measures show different correlation patterns or error variances between groups. For example, experienced users might have a different understanding of a system's usefulness compared to inexperienced users, thereby leading to different correlation patterns for the PU construct's indicators. The respondents' experience can also affect PU's error variance between groups, as inexperienced users might have higher variability in their responses than experienced users who have a clearer understanding of the system's usefulness.

Unobserved heterogeneity can also threaten *construct validity* because differences in indicator loadings and weights across groups will not be detected. As such, an evaluation of construct validity based on the overall sample while overlooking unobserved heterogeneity will not reveal the true group-specific measures of the constructs, thereby risking not detecting if the construct captures a different phenomenon for each group. Moreover, if the measures derived based on the overall sample do not represent the true construct (e.g., PU), the biased construct can lead to invalid inferences on relationships with other constructs, thereby threatening *criterion validity*. Both threats are regularly addressed when testing for ME/I in multigroup models (i.e., observed heterogeneity) (see Steenkamp and Baumgartner 1998; Vandenberg and Lance 2000) but are usually overlooked in the context of unobserved heterogeneity.

In contrast, unobserved heterogeneity typically does not affect *content validity* because the constructs' measures are normally the same across groups and are grounded in theory. However, an increase in the value of a formative measure's error term due to unobserved heterogeneity can lead to misinterpretations, as a high error term is typically associated with the construct measure's incompleteness (Diamantopoulos et al. 2008).

Table 2. Implications of Unobserved Heterogeneity for Model Validity

Type of Validity		What is It?	Threats Due to Unobserved Heterogeneity	Why Is It a Threat?
Internal Validity		• Is the effect due to unhypothesized variables? • Are there rival explanations for the findings or just one single explanation?	• There are other viable explanations for the findings, namely group differences that are not accounted for.	• The observed effects are a result of unhypothesized and/or unmeasured variables (i.e., the groups and corresponding explanatory variables). • Example: the underlying theory does not include differences in the technology acceptance between experienced and inexperienced users.
Instrumental Validity	**Content Validity**	• Do the indicators accurately reflect the theoretical domain?	**Formative & Reflective** • In general, heterogeneity does not affect content validity, as content validity is grounded in theory. **Formative** • The error term of the formative construct likely increases due to unobserved heterogeneity, which can be mistakenly interpreted as lack of content validity (Type II Error).	• The empirically relevant (i.e., significant) set of indicators may vary across groups. • Varying nonsignificant indicators across groups indicate problems with ME/I, but this is a problem of construct validity in the sense of (not) capturing the right phenomenon. • Nonsignificant indicators should remain in the model if theoretically relevant. • Following Diamantopoulos et al. (2008), the error term in formative constructs represents those "aspects of the construct domain not represented by the indicators." Understanding the error term in this way and assessing it without capturing unobserved heterogeneity may indicate insufficient content validity although all important indicators are included in the formative construct.
	Construct Validity	• Are the chosen measures repre-senting the true con-struct of the phenomenon? • Are the operationali-zations of the constructs correct?	**Formative & Reflective** • Indicator weights/loadings estimated with the assumption that no underlying groups exist are biased if groups actually exist.	• For formative measures, differences in the importance of indicators across groups lead to different measurement weights although the phenomenon is still the same. • For reflective measures, when ME/I is established across groups (i.e., there are no differences in the weights/ loadings), there is no threat of unobserved heterogeneity to construct validity. Otherwise, the construct captures a different phenomenon for each group. Combining the measures at the overall sample level is not allowed.
	Criterion Validity	• Are inferences from the construct to a related behavioral criterion of interest accurate?	**Formative & Reflective** • Differences in construct perceptions across groups (i.e., different weights/ loadings) lead to biased construct scores, which, in turn, influence (bias) the estimated relationship with other constructs.	• The measures based on the overall sample do not represent the true group-specific measures of the constructs. This causes problems when interpreting the construct scores or their relationships with other constructs in the model. • For reflective measures, when there is no ME/I established across groups, the apparently different phenomena across groups have varying and incomparable relationships with other constructs.
	Reliability	• Are the measures accurate? • Are the measures consistent?	**Test-Retest Reliability (Formative & Reflective)** • Not affected **Internal Consistency (Reflective)** • Reliability (e.g., Cronbach's alpha) at the overall sample level is negatively influenced by the lack of ME/I across groups.	• Repeating the measurement with the same observations under the same conditions should lead to the same results on the overall and group levels. • Different correlation patterns across groups for a reflective perceived usefulness construct can lead to an average correlation pattern on the overall sample level, which does not show appropriate internal consistency.
Statistical Conclusion Validity		• Have adequate sampling procedures, appropriate statistical tests, and reliable measurements been used?	• Heterogeneous samples may lead to higher standard errors or lower effect sizes, thereby influencing the power of tests. • Biased estimates, Type I, and Type II errors.	• Path coefficients for relationships between constructs (e.g., ease of use and intention to use) might have higher standard errors on the overall sample than in their underlying groups, indicating a variety of different coefficients across user groups. • This also applies to formative measurement weights.
External Validity		• Are findings generalizable to other populations and conditions?	• Interpretations of the overall sample may be ambiguous and misleading. • Results cannot be generalized easily, as they are valid for only a special condition of the model.	• Analyzing population differences reveals more general conclusions about the model than those from the overall sample. • Example: Based on the overall sample level, usefulness has the same importance as ease of use. However, there are no users who value usefulness and ease of use equally; rather, there are two distinct groups of experienced and inexperienced users.

Finally, if unobserved heterogeneity is not uncovered, there is a threat to *external validity* (i.e., the ability to generalize findings beyond the current population and context) because the overall sample results are not representative of the underlying groups. As findings are averaged across groups, results obtained using the overall sample cannot be generalized to different groups. The observation of inconsistent, conflicting, or mixed findings in the 12 meta-analyses in Table A1 (Appendix A) also show that the results of one study often cannot be generalized to other studies (indicating low external validity) with unobserved heterogeneity being one of the plausible reasons.

Because of these threats to the different types of validity, it is important to uncover heterogeneity in data that may otherwise lead to invalid conclusions. Next, we present an overview of methods to uncover unobserved heterogeneity in structural equation models that researchers can apply to overcome threats to validity due to unobserved heterogeneity.

Uncovering Heterogeneity in Structural Equation Models

In this section, we first synthesize and compare different methods in SEM (i.e., CB-SEM and PLS path modeling) to uncover observed and unobserved heterogeneity. Given the objectives of our paper, we focus primarily on methods in SEM to uncover unobserved heterogeneity.[3] We also introduce a new method to address some of the limitations of existing methods to uncover unobserved heterogeneity in PLS path models.

Existing Methods to Uncover Observed Heterogeneity in SEM

SEM methods to address observed heterogeneity are now commonly applied in the social and behavioral sciences, including information systems. The first category of methods identifies homogenous groups of observations (e.g., individuals) *a priori* based on grouping variables (e.g., psychographic or socio-demographic). A multigroup analysis reveals the heterogeneity between the groups by testing for differences across group-specific parameter estimates. Examples of these methods for PLS path modeling can be found in Chin and Dibbern (2010), Sarstedt et al. (2011b), and Qureshi and Compeau (2009) and for CB-SEM in Jöreskog (1971) and Sörbom (1974). The second category of methods aims at identifying moderating factors that explain heterogeneity in specific structural model relationships. Examples of these methods in PLS path modeling can be found in Chin et al. (2003), Goodhue et al. (2007) and Henseler and Chin (2010) and for CB-SEM in Jaccard and Wan (1995), Jöreskog and Yang (1996), and Klein and Moosbrugger (2000). Uncovering observed heterogeneity with both types of methods requires *a priori* knowledge about differences across groups. Consequently, these two types of methods do not account for unobserved heterogeneity—that is, differences across groups that are not informed by existing theory and are unknown *a priori*.

Existing Methods to Uncover Unobserved Heterogeneity in SEM

The next sections present methods in CB-SEM and PLS path modeling to uncover unobserved heterogeneity.

CB-SEM Methods to Uncover Unobserved Heterogeneity

In CB-SEM, the following two primary methods have been developed to uncover unobserved heterogeneity: (1) finite mixture models that extend multigroup CB-SEM (Arminger et al. 1999; Dolan and van der Maas 1998; Jedidi et al. 1997) and (2) hierarchical Bayesian models that extend multilevel CB-SEM (Ansari et al. 2000; Cai and Song 2010; Lee and Song 2003). Table 3 presents a summary of these CB-SEM methods.

Finite mixture models for CB-SEM were developed by Jedidi et al. (1997), Arminger et al. (1999), and Dolan and van der Maas (1998). These models (1) assume that data originate from subpopulations (groups) in the overall population that is a mixture of them and (2) generalize multigroup CB-SEM (Jöreskog 1971; Sörbom 1974) to unobserved latent groups assuming the structural parameters (covariance) and factor means to be mixtures of components. The method used for finite mixture models assigns the observations to a prespecified number of groups by means of fuzzy (probabilistic) clustering, thereby permitting the simultaneous estimation of group-specific parameters (Jedidi et al. 1997). Consequently, finite mixture models address unobserved heterogeneity in the data by grouping observations and estimating group-specific

[3]There are several methods to uncover both observed and unobserved heterogeneity in other methodological contexts—for example, regression analysis (DeSarbo and Cron 1988; Späth 1979; Wedel and DeSarbo1994), panel data models (Allenby and Rossi 1998; Popkowski Leszczyc and Bass 1998), and conjoint analysis (DeSarbo et al. 1995; Gilbride et al. 2006; Lenk et al. 1996). Given the objectives of our paper and for reasons of scope, we do not review these methods.

Table 3. Overview of CB-SEM Methods to Uncover Unobserved Heterogeneity in SEM

Method	Description	Parameter Estimates	Limitations	Illustrative Applications
Finite Mixture Models for CB-SEM Jedidi et al. 1997	Generalizes the multigroup SEM for unobserved group-specific differences in the following: • Structural parameters (covariance) • Factor means	For a defined number of groups	• Number of groups is unknown to the researcher • Does not account for heterogeneity in the covariance of the measures • Requires large number of observations (large sample sizes)	Bart et al. 2005 DeSarbo et al. 2006 Reinecke 2006 Tueller and Lubke 2010
Hierarchical Bayesian CB-SEM Ansari et al. 2000	Generalizes the multilevel SEM for unobserved individual-specific differences in the following: • The covariance structure (i.e., structural parameters, measurement error variance, and factor covariance) • Factor means	Specific estimates for individuals	• Needs continuous data with multiple observations per individual • Only works for recursive structural equation models • Not available in standard software packages	Luo et al. 2008

parameters simultaneously, thus avoiding well-known biases that occur when group-specific models are estimated separately (Fraley and Raftery 2002). Several applications and simulation studies (e.g., Arminger et al. 1999; Henson et al. 2007; Jedidi et al. 1997; Tueller and Lubke 2010) illustrate the usefulness of finite mixture models by showing how structural relationships among factors differ across unobserved groups.

In contrast to finite mixture models, hierarchical Bayesian models for CB-SEM, which were developed by Ansari et al. (2000), do not assume heterogeneity among a defined number of groups of individuals but estimate unobserved heterogeneity at the individual[4] level using a random coefficients model. Specifically, they uncover unobserved heterogeneity in the factor means and covariance structure (i.e., structural parameters, measurement error variance, and factor covariance), thereby generalizing multilevel SEM models (Muthén 1994; Rabe-Hesketh et al. 2004) that only account for heterogeneity in the mean structure. Hierarchical Bayesian CB-SEM provides individual-specific estimates for the factor scores, structural coefficients, and other model parameters (Ansari et al. 2000). However, this method requires continuous data with multiple observations per individual to estimate individual-level heterogeneity, and the method is limited to recursive structural equation models. There has been some work (e.g., Cai and Song 2010; Lee and Song 2003) to extend the method to dichotomous variables and missing data and evaluate the performance of these methods.

[4]An individual can be a person, group, team, or company that is the object of investigation in a study and has provided several observations (e.g., over time or within a group).

While both the finite mixture and the hierarchical Bayesian CB-SEM models have been the subject of extensive methodological research, finite mixture models have been applied in empirical CB-SEM research to a greater extent. An increasing number of applications, especially in the marketing, econometrics, and sociology literatures, have utilized finite mixture models to uncover unobserved heterogeneity, thereby improving theoretical and practical implications (e.g., Bart et al. 2005; DeSarbo et al. 2006; Reinecke 2006; Tueller and Lubke 2010).

PLS Path Modeling Methods to Uncover Unobserved Heterogeneity

Although PLS path modeling research has paid limited attention to unobserved heterogeneity in comparison to CB-SEM research, multiple PLS segmentation methods have been proposed. We draw on Sarstedt's (2008) review of these methods to identify the following key PLS segmentation methods:

1. The PATHMOX (path modeling segmentation tree) algorithm (Sánchez 2009; Sánchez and Aluja 2006).[5] This algorithm requires the *a priori* specification of explanatory variables that are not used as indicators in the PLS path model to discover segments. While this feature can be advantageous for interpreting discovered segments, it limits the heterogeneity discovery process to the selected explanatory variables (and their specified

[5]PATHMOX is available in the "pathmox" package of the statistical software R (Sánchez and Aluja 2012).

order) that are provided as inputs to the PATHMOX algorithm (Sarstedt 2008).

2. Distance measure-based methods. These methods determine the distance of an observation to its current group and all other given groups in order to decide on this observation's group membership. PLS typological path modeling (PLS-TPM; Squillacciotti 2005; Squillacciotti 2010) and its enhancement—response-based detection of respondent segments in PLS (REBUS-PLS; Esposito Vinzi et al. 2010; Esposito Vinzi et al. 2008)—are the key methods in this class.[6] Both PLS-TPM and REBUS-PLS[7] can only uncover unobserved heterogeneity in PLS path models with reflective measures (i.e., they cannot be applied to path models that include formative measures) (Esposito Vinzi et al. 2010; Esposito Vinzi et al. 2008).

3. The finite mixture partial least squares method (FIMIX-PLS) (Hahn et al. 2002).[8] This method assumes that each endogenous latent variable is distributed as a finite mixture of conditional multivariate normal densities. It captures heterogeneity by estimating the probabilities of segment memberships for each observation in order to optimize the likelihood function. Consequently, it implicitly maximizes the segment-specific explained variance (i.e., the R^2 value), which is part of the likelihood function. While FIMIX-PLS is generally applicable to PLS path models regardless of whether the latent variables are measured reflectively or formatively, it does not account for the heterogeneity in the measurement models. Moreover, the assumption that the endogenous latent variables have multivariate normal distribution is inconsistent with the nonparametric PLS path modeling which does not impose distributional assumption.

We select FIMIX-PLS to benchmark the performance of the new PLS-POS method for two reasons. First, based on an assessment of the benefits and limitations of these methods, Sarstedt (2008, p. 152) concludes: "To sum up, FIMIX-PLS can presently be viewed as the most comprehensive and commonly used approach to capture heterogeneity in PLS path modeling." Second, as our research objectives include developing/evaluating a method (i.e., PLS-POS) that detects unobserved heterogeneity in both the structural model and formative measures, we conduct simulations with both formative and reflective models. While PLS-TPM and REBUS-PLS are not applicable to PLS path models that include formative measures, FIMIX-PLS is applicable to PLS path models regardless of the use of reflective/formative measurement. We next elaborate briefly on FIMIX-PLS' assumptions, procedure, and limitations.

FIMIX-PLS follows the assumption that heterogeneity is concentrated in the parameters of the estimated relationships among latent variables (i.e., the path coefficients in the structural model). Based on this concept, FIMIX-PLS assigns observations to a prespecified number of groups by means of probabilistic clustering to optimize the likelihood function (which implicitly maximizes the segment-specific explained variance as part of the likelihood function), thereby simultaneously estimating the model parameters for the groups and ascertaining the heterogeneity of the data for the PLS path model. It adapts a finite mixture regression model that, in contrast to conventional mixture regression models, can be comprised of a multitude of interrelated endogenous latent variables (Hahn et al. 2002).

Compared to the finite mixture and hierarchical Bayesian CB-SEM, FIMIX-PLS does not account for group-specific mean differences of latent variables because it is based on the standardized results of an overall sample PLS path model. In addition, FIMIX-PLS builds on the latent variable scores of the PLS path model estimation using the full set of data and, thus, only focuses on the relationships among latent variables. Consequently, it is generally applicable to PLS path models (regardless of the latent variables being measured reflectively or formatively) but does not account for the heterogeneity in the measurement models (e.g., the factor covariance or the measurement error variance) (Hahn et al. 2002; Sarstedt and Ringle 2010).

FIMIX-PLS has been applied recently to uncover unobserved heterogeneity in PLS path models for success factors in industrial goods (Sarstedt et al. 2009), intention to adopt new movie distribution services on the Internet (Papies and Clement) 2008), the American customer satisfaction index model (Ringle et al. 2010a), and unanticipated reactions to organizational strategy among stakeholder segments (Money et al. 2012). The advantage of applying the parametric finite mixture regression concept to PLS path models is that it offers segment retention criteria (e.g., AIC, BIC, and CAIC; Hahn et al. 2002; Sarstedt et al. 2011a) for model selection (i.e., to

[6]Other distance-based methods, which are in earlier stages of development and currently not available as software packages, include fuzzy PLS path modeling for latent class detection (FPLS-LCD; Palumbo et al. 2008) and partial least squares genetic algorithm segmentation (PLS-GAS) (Ringle et al. 2010b; Ringle et al. 2013).

[7]The REBUS-PLS method is included in the XLSTAT software as well as in the "plspm" package (Sánchez and Trinchera 2013) of the statistical software R (R Core Team 2013).

[8]The FIMIX-PLS method is included in the PLS path modeling software SmartPLS (Ringle et al. 2005).

decide on an appropriate number of segments). However, FIMIX-PLS has some limitations in that it (1) assumes that the endogenous latent variables in the structural model have a multivariate normal distribution (which is inconsistent with PLS' distribution-free assumption) and (2) uses latent variable scores in the structural model based on the measurement model for the overall sample and ignores plausible heterogeneity in the measurement model's weights. Consequently, it not only ignores heterogeneity in the measurement model but may also fail to detect heterogeneity in the structural model that results from unobserved heterogeneity in the measurement model.

Partial Least Squares–Prediction-Oriented Segmentation (PLS-POS)

To overcome the identified methodological limitations of FIMIX-PLS and of existing distance measure-based PLS segmentation methods for uncovering unobserved heterogeneity, we introduce the PLS prediction-oriented segmentation (PLS-POS) method that offers three novel and distinctive features: (1) it uses a PLS-specific *objective criterion* to form homogeneous groups that maximize the explained variance (R^2) of all endogenous latent variables in the PLS path model and, thereby, takes the entire path model's structure into account;[9] (2) it includes a new *distance measure* that is appropriate for formative measures (and heterogeneity within them); and (3) it reassigns observations only if *reassigning observations improves the objective criterion*. The latter feature of PLS-POS ensures continuous improvement of the objective criterion throughout the iterations of the algorithm (hill-climbing approach) and provides the ability to uncover very *small niche segments*. However, like the expectation–maximization (EM) algorithm in FIMIX-PLS, PLS-POS can face the problem of ending in local optima due to its use of a hill-climbing approach. Thus, a repeated application of PLS-POS with different starting partitions is advisable.

PLS-POS follows a clustering approach with a deterministic assignment of observations to groups and uses a distance measure for the reassignment of observations; as such, it has no distributional assumptions. The segmentation objective in a PLS path model is to form homogenous groups of observations with increased predictive power (R^2 of the endogenous latent variables) of the group-specific path model estimates (compared to the overall sample model). In accordance with Anderberg's (1973, p. 195) notion of "clustering for maximum prediction," a fitting objective criterion for PLS segmentation is to maximize the sum of the endogenous latent variables' explained variance (R^2) across all groups.

A key challenge of this approach is the indeterminacy of the data assignment task, as it is unknown how the group-specific PLS results will change when an observation is reassigned to a different group. For this purpose, the PLS-POS method uses a distance measure to identify appropriate observations for reassignment that serve as candidates to improve the PLS-POS objective criterion. Using a distance measure (i.e., calculating each observation's distance from its current group and from each of the other groups) for segmentation builds on an idea of earlier work on distance-measure-based segmentation in PLS path modeling (i.e., PLS-TPM and its later improvement REBUS-PLS).

Appendix B provides the details of PLS-POS' algorithm, objective criterion, and distance measure. It also includes a detailed comparison of the technical differences between FIMIX-PLS, PLS-TPM, REBUS-PLS, and PLS-POS (Table B1). We implement the PLS-POS algorithm as an extension of the SmartPLS software (Ringle et al. 2005) to evaluate its performance in our simulation study. The extension will be made available with the next release of SmartPLS.

In summary, the PLS-POS method complies with the most important objectives in PLS path modeling. It (1) improves the objective criterion by nonparametric means; (2) accounts for heterogeneity in the structural model, as well as in the formative measurement model; and (3) is applicable to all path models regardless of the type of measurement model, the distribution of the data, or the complexity of the structural model. Table 4 compares the key properties of PLS-POS and FIMIX-PLS, which we use as the benchmark method in this study as depicted in the previous section, in terms of five desired criteria for a PLS segmentation method.

In the next section, we detail the comprehensive simulation experiments we conducted to evaluate whether the differences in the capabilities of FIMIX-PLS and PLS-POS noted in Table 4 hold empirically. Specifically, we focused our simulations on the criteria in columns 2 through 5 because our goal

[9]While PLS-TPM only focuses on a single target construct, REBUS-PLS accounts for this limitation by replacing PLS-TPM's distance measure with the goodness-of-fit criterion-based (GoF; Tenenhaus et al. 2005) closeness measure: "The aim of REBUS-PLS is to detect sources of heterogeneity in both the structural and the outer model for all exogenous and endogenous latent variables." (Esposito Vinzi et al. 2008, p. 444). As in PLS-TPM, REBUS-PLS requires reflective measurement models (Esposito Vinzi et al. 2008). In contrast, by focusing on the R^2 of all the endogenous latent variables as an explicit objective criterion, PLS-POS stresses the prediction-oriented character of PLS path modeling, and allows the general application of this method to PLS path models with both reflective and formative measurement models.

Table 4. Conceptual Capabilities of FIMIX-PLS and PLS-POS

Segmentation Methods	Desired Criteria for a PLS Segmentation Method: Ability to detect heterogeneity in reflective measures	Ability to detect heterogeneity in formative measures	Ability to detect heterogeneity in the structural model	Maximizes group-specific R^2 of endogenous latent variables (prediction orientation)	Ability to handle non-normal data
FIMIX-PLS Hahn et al. 2002	–	–	✓	✓	–
PLS-POS	✓*	✓	✓	✓	✓

*The method can detect heterogeneity in the reflective model if there is heterogeneity in the structural model (i.e., if heterogeneity in the reflective measurement model is the source of heterogeneity in the structural model).

is to discover heterogeneity in the structural model and in formative measures while assuming measurement invariance in the reflective measures.

Simulations of PLS-POS and FIMIX-PLS Performance

We conducted experiments with simulated data that define the true group-specific PLS parameters *a priori*. We assessed the performance of PLS-POS and FIMIX-PLS based on the differences between the true parameters and those estimated by each method. Subsequently, we compared the performance of PLS-POS and FIMIX-PLS in recovering the true parameter estimates.

Model Specification

Consistent with most simulation studies on PLS path models (e.g., Chin et al. 2003), we specified a direct effects path model that includes four exogenous latent variables and one endogenous variable. We specified two versions of the path model: model 1 uses reflective measures for the exogenous and endogenous latent variables (Figure 2a), while model 2 uses formative measures for the exogenous latent variables and reflective measures for the endogenous latent variables (Figure 2b). While we limit the results reported in this paper to those obtained from the simulations of a direct effects path model, we also evaluated more complex path models with multiple endogenous variables and mediation paths between the latent variables. Our results were generally stable for these more complex models as well.

We generated the simulated data so each of the two groups has one particularly strong relationship in the structural model, while all other path coefficients are at lower levels of magnitude. For example, for group 1, the structural path p_1 has a high true parameter value, while the structural paths p_2 to p_4 have lower true parameter values. Conversely, for group 2, p_4 has a high true parameter value, while the path coefficients p_1 to p_3 have lower true values. The mean differences in the coefficients for path p_1 to p_4 between group 1 and group 2 reflect the heterogeneity in the model (i.e., the differences between the groups). The same principle applies to the measurement weights in the formative measures. We used four formative indicators per construct. For group 1, the measurement weights w_1 and w_3 have high true values, while weights w_2 and w_4 have low true values. Conversely, for group 2, w_2 and w_4 have high true values, and w_1 and w_3 have low true values. The mean differences between the weights for group 1 and group 2 reflect the amount of heterogeneity in the measurement model.

Factor Design of the Simulations

Our selection of experimental factors and their levels was informed by criteria that were shown to influence PLS path modeling or segmentation results in prior simulation studies. Specifically, we manipulated the following factors:

(1) *Explained variance* (R^2) of the endogenous latent variable per group (1.00, .95, .90, .85)[10] (e.g., Reinartz et al. 2009).

(2) *Structural model heterogeneity*—that is, the group-specific differences in structural model path coefficients (.25, .50, .75, 1.00) (e.g., Andrews and Currim 2003b).

[10]This manipulation results in R^2 values of .425 to .5 in the overall sample that combines groups. For example, when the R^2 value in both groups is .85, the overall sample that combines the two groups has a R^2 value of .425 because of unobserved heterogeneity.

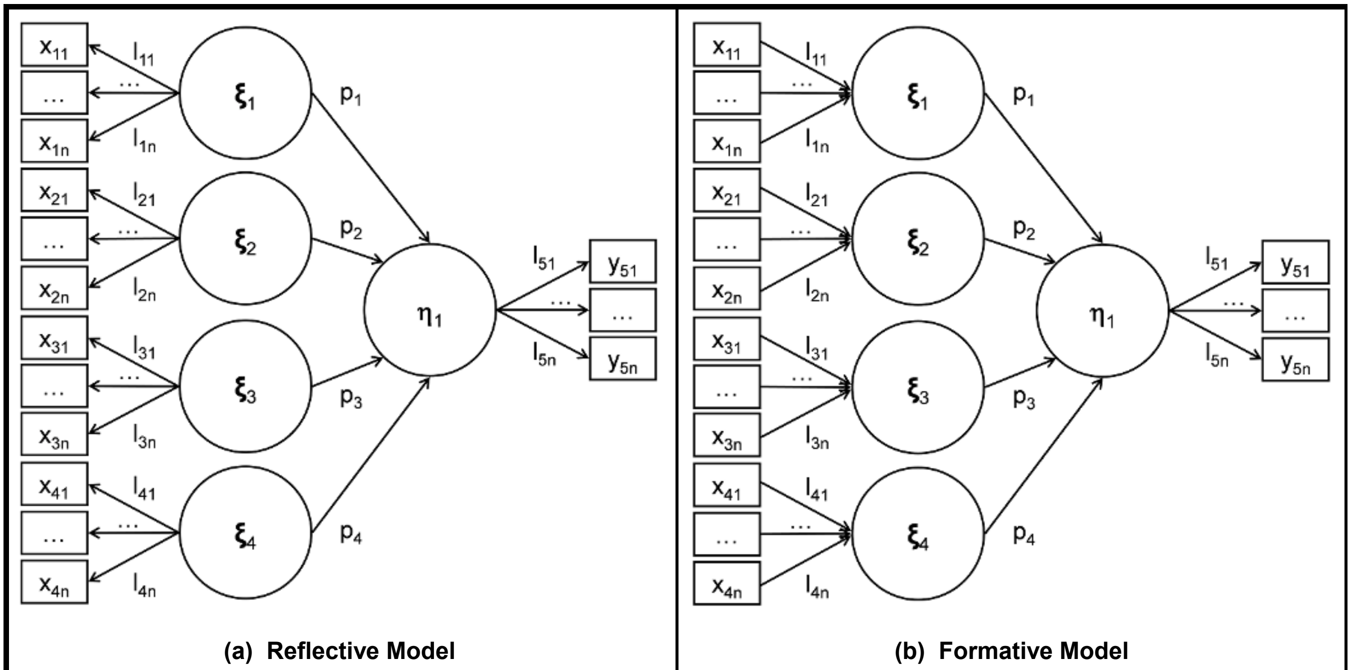


Figure 2. The Models

(3) *Sample size* per group (100, 200, 400) (e.g., Chin et al. 2003).

(4) *Data distribution* (normal, non-normal[11]) (e.g., Reinartz et al. 2009).

(5) *Relative segment sizes* (equal, unequal[12]) (e.g., Andrews and Currim 2003b).

In addition, we manipulated the following factors related to the measurement model:

(6) *Reliability* of reflective measures (perfect versus normal; loadings of 1.00 and ~.85) (e.g., Chin et al. 2003).

(7) *Measurement model heterogeneity*—that is, the group-specific differences in formative measurement weights (.25, .50, .75). (We note that to the best of our knowledge, this particular factor has not been examined in prior simulation research on PLS path models.)

(8) *Multicollinearity between formative indicators* (none, level 1, level 2)[13] (Mason and Perreault 1991).

The number of factors and the number of factor levels systematically increase the complexity of the PLS segmentation task. The full factorial design for the study results in $4^2 \times 3 \times 2^3 =$ 384 different combinations for the reflective model (model 1) and $4^2 \times 3^3 \times 2^2 = 1{,}728$ different combinations for the formative model (model 2). To ensure stability of the results, all factor combinations include 30 data-generation and segmentation runs for each segmentation method, so in total, (384 + 1,728) × 2 × 30 = 126,720 segmentation runs were performed.

Data Generation

Simulation studies in PLS path modeling require that data generated for the indicators (manifest variables) match the true values of the model. Previous studies on PLS path modeling (e.g., Chin et al. 2003; Henseler and Chin 2010; Reinartz et al. 2009) first generated data by extracting latent variable scores to match the true relationships in the structural model and then generated data for the indicators by adding measurement errors to match the indicators' true parameters

[11]For the non-normal data, we use a log-transformation of the normal data to get a skewness of about 2 and a kurtosis of about 5 for the indicators.

[12]The unequal condition has one segment with 80% and one with 20% of the total sample size.

[13]For a detailed explanation of this factor, see Appendix C.

in the measurement model. This procedure does not allow for generating data for formative indicators, as the direction of causality in formative measures is from the indicators to the construct (in contrast to reflective measures, where the indicators cause the construct). Data for the formative indicators must first be generated to compute the latent variable scores for formative constructs. We address this requirement by generating random variables for the formative indicators such that the generated formative indicators match a prespecified correlation matrix (for modeling multicollinearity in the simulation design), the true values of the formative measurement weights, as well as the true values for the structural model parameters.

Performance Assessment

The objectives of our simulation experiments were to (1) assess PLS-POS and FIMIX-PLS in terms of their respective abilities to recover true group-specific parameters, (2) compare PLS-POS and FIMIX-PLS based on the assessment of their parameter recovery, and (3) identify the relative effects of the design factors on the parameter recovery of PLS-POS and FIMIX-PLS.

We knew the true parameters of each factorial combination (i.e., the R^2, path coefficients, outer weights, and loadings) *a priori* based on the parameter settings for the data generation. The smaller the differences between the true values and the segmentation method's parameter estimates, the better the parameter recovery. As FIMIX-PLS cannot provide segmentation results for the measurement model—because parameters are fixed to those resulting from the overall sample—we assessed each segmentation method by comparing the structural model's path coefficients from the two segmentation methods with the *a priori* known values. Consistent with prior studies (e.g., Henseler and Chin 2010; Reinartz et al. 2002), we evaluated parameter recovery using the mean absolute bias (MAB), which is the average of the simple absolute deviations between the true parameter and the parameter estimated by the segmentation method. MAB values close to zero indicate near perfect parameter recovery. To assess PLS-POS and FIMIX-PLS, we compared each method's MAB with the MAB when the overall sample was analyzed without uncovering unobserved heterogeneity (i.e., without using a segmentation method). Finally, to understand the relative importance of the design factors, we evaluated parameter recovery (i.e., the path coefficient's MAB) using a mixed-effects ANOVA model with the two segmentation methods (PLS-POS and FIMIX-PLS; within-subjects factor) and the eight design factors (between-subjects factors).

Results of the Simulation Experiments

We discuss the findings for both model 1 (reflective measures) and model 2 (formative measures) below starting with the results for model 1.

Results for Model 1: Reflective Measures

Table 5 presents the results for the ANOVA with MAB as the dependent variable. Our extensive simulations enabled us to detect even very small effects, indicating high power. For the sake of space and simplicity, Table 5 shows only the direct effects, all two-way interactions with the method factor, and all other interactions having a significant and substantial effect (i.e., explaining more than 2% of the total variance in MAB, implying a partial η^2 of more than .02 (Reinartz et al. 2009)). The partial η^2 represents the contribution of each factor or interaction as if it is the only variable, so its effect is not masked by other variables. See Appendix E for the complete results.

The ANOVA results for model 1 show that parameter recovery is unaffected by the measurement model's reliability. The direct effect and all of the interaction effects of reliability are nonsignificant. As the reliability has neither a between-subjects nor a within-subjects effect, we find no evidence that the accuracy of either segmentation method is affected by the reliability of the measurement model.

The between-subjects effects identify the factors that influenced MAB for both segmentation methods. All of the direct effects are significant with two notable findings: (1) sample size (partial η^2 = .013) and relative segment size (partial η^2 = .002) have a partial eta-square below .02, so their influence on MAB is not substantial, and (2) R^2 has the strongest impact on parameter recovery both as a direct effect and as an interaction effect with structural model heterogeneity. This result is not surprising, as an increasing error in the model distorts group differences. As PLS-POS capitalizes on the model's predictive power of the model (i.e., the explained variance), the method is better at uncovering heterogeneity when the predictive power is high.

The within-subjects effects identify the differential influence of the design factors on MAB across the segmentation methods. In general, the method has a significant and substantial impact on the parameter recovery for the reflective model. Furthermore, the method's two interaction effects with structural model heterogeneity and R^2 are significant and substantial. All other interaction effects with the method are nonsignificant or are not substantial.

Table 5. Model 1 (Reflective Measures) ANOVA Explaining MAB by Method (PLS-POS/FIMIX-PLS) and Design Factors

	Source of Variance in MAB	df	F-value	p-value	Partial η^2
Between-Subjects Effects	Intercept	1	14,658.62	.000	.568
	Structural Model Heterogeneity	3	1,121.71	.000	.232
	R^2	3	1,948.85	.000	.344
	Sample Size	2	70.77	.000	.013
	Reliability	1	1.88	.170	.000
	Data Distribution	1	497.52	.000	.043
	Relative Segment Size	1	22.62	.000	.002
	Structural Model Heterogeneity × R^2	9	178.96	.000	.126
	Error	11,136			
Within-Subjects Effects	Method	1	952.31	.000	.079
	Method × Structural Model Heterogeneity	3	217.47	.000	.055
	Method × R^2	3	137.14	.000	.036
	Method × Sample Size	2	4.66	.009	.001
	Method × Reliability	1	.01	.974	.000
	Method × Data Distribution	1	87.97	.000	.008
	Method × Relative Segment Size	1	104.01	.000	.009
	Error (Method)	11,136			

Note: df = degrees of freedom

Table 6 shows the MAB for each factor level when PLS-POS or FIMIX-PLS is applied to uncover heterogeneity or the overall sample was analyzed without the use of a segmentation method to uncover heterogeneity. A detailed examination of the significant interaction effects of the method with the structural model heterogeneity and the R^2 shows that the MAB for PLS-POS increases more than the MAB for FIMIX-PLS when the structural model heterogeneity or the R^2 is lower (Figures 3a and 3b). However, using PLS-POS results in a MAB that is still very low compared to the MAB when the overall sample was analyzed without the use of a segmentation method.

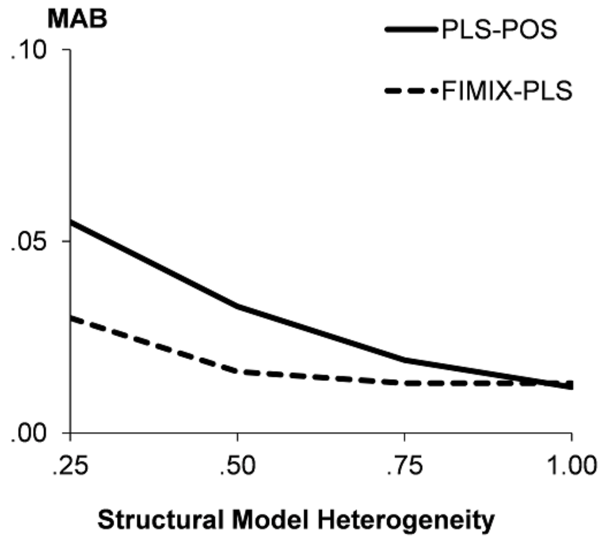
Overall, the results reveal that for model 1 (reflective measures), both methods perform equally well in almost all conditions; FIMIX-PLS is slightly better than PLS-POS when the R^2 or the structural model heterogeneity is low, and the bias from using either of the two methods (FIMIX-PLS or PLS-POS) is much lower than the bias from analyzing the overall sample without uncovering heterogeneity.

Results for Model 2: Formative Measures

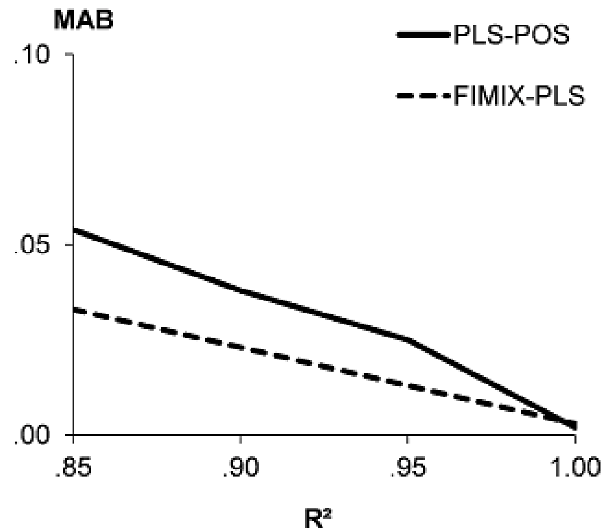
Table 7 presents the results for the ANOVA in model 2 (formative measures) with MAB as the dependent variable. Again, for the sake of space and simplicity, Table 7 presents the direct effects, all two-way interactions with the method, and all other interactions that have significant and substantial effects (partial η^2 of more than .02). See Appendix F for the complete results.

For the between-subjects effects, all of the direct effects on MAB are significant, but again, the effect of relative segment size (partial η^2 = .012) on MAB is not substantial. Interestingly, the relative segment size and sample size have a substantial interaction in this model (partial η^2 = .054). The MAB decreases for increased sample sizes in groups of equal size but stays constant for increased sample sizes in unequal groups.

The MAB for both segmentation methods is influenced by the heterogeneity in the structural model, the heterogeneity in the measurement model, the R^2 of the model, the sample size, the data distribution, and the multicollinearity. In contrast to the results for model 1 (reflective measures), it is not the R^2 (partial η^2 = 0.204) but the structural model heterogeneity that has the highest impact (partial η^2 = .313) on parameter recovery for model 2 (formative measures). The impact of the measurement model heterogeneity (this factor is only relevant for formative measures) on MAB is the third most important factor and explains about 10 percent of the MAB variance (partial η^2 = .104). Moreover, the interaction effects between the structural model and measurement model heterogeneity as



(a) MAB of Both Segmentation Methods for Different Structural Model Heterogeneity



(b) MAB of Both Segmentation Methods for Different R² Values

Figure 3. MAB of Both Segmentation Methods for Model 1 (Reflective Measures)

Table 6. MAB in Model 1 (Reflective Measures) for Each Method

Design Factor Level		**POS** Mean Absolute Bias	**FIMIX** Mean Absolute Bias	**No Segmentation Method** Mean Absolute Bias
Structural Model Heterogeneity	.25	.055	.030	.125
	.50	.033	.016	.250
	.75	.019	.013	.375
	1.00	.012	.013	.500
R^2	.85	.054	.033	.312
	.90	.038	.023	
	.95	.025	.013	
	1.00	.002	.003	
Sample Size	100	.032	.021	.312
	200	.031	.018	
	400	.026	.015	
Reliability	Perfect	.030	.018	.312
	Normal	.029	.018	
Data Distribution	Normal	.024	.015	.312
	Non-Normal	.036	.021	
Relative Segment Size	Equal	.027	.019	.312
	Unequal	.033	.017	
Overall		.030	.018	.312

Table 7. Model 2 (Formative Measures) ANOVA Explaining MAB by Method (PLS-POS/FIMIX-PLS) and Design Factors

	Source of Variance in MAB	df	F-value	p-value	Partial η^2
Between-Subjects Effects	Intercept	1	142696.80	.00	.740
	Structural Model Heterogeneity	3	7605.33	.00	.313
	Measurement Model Heterogeneity	2	2912.99	.00	.104
	R^2	3	4286.31	.00	.204
	Sample Size	2	864.77	.00	.033
	Relative Segment Size	1	629.83	.00	.012
	Data Distribution	1	1465.75	.00	.028
	Multicollinearity	2	848.18	.00	.033
	Structural Model Heterogeneity × Measurement Model Heterogeneity	6	298.09	.00	.034
	Sample Size × Relative Segment Size	2	1426.86	.00	.054
	Measurement Model Heterogeneity × Multicollinearity	4	287.84	.00	.022
	Error	50,112			
Within-Subjects Effects	Method	1	3938.52	.00	.073
	Method × Structural Model Het.	3	3987.98	.00	.193
	Method × Measurement Model Het.	2	6771.05	.00	.213
	Method × R^2	3	826.32	.00	.047
	Method × Sample Size	2	227.55	.00	.009
	Method × Relative Segment Size	1	171.66	.00	.003
	Method × Data Distribution	1	2.97	.08	.000
	Method × Multicollinearity	2	1739.12	.00	.065
	Method × Structural Model Het. × Measurement Model Het.	6	976.49	.00	.105
	Method × Structural Model Het. × Multicollinearity	6	372.96	.00	.043
	Method × Measurement Model Het. × Multicollinearity	4	257.24	.00	.020
	Error (Method)	50,112			

Note: df = degrees of freedom

well as between measurement model heterogeneity and multicollinearity are significant and substantial but have very little impact compared to the factors discussed earlier.

For the within-subjects effects, the method's effect on MAB is significant and substantial. The method also significantly and substantially interacts with heterogeneity in both the structural model and the measurement model. Looking at these interaction effects in more detail reveals that PLS-POS performs consistently well across all of the factor levels, while the performance of FIMIX-PLS deteriorates with decreasing structural model heterogeneity or increasing measurement model heterogeneity. Interestingly, the three-way interaction of method with structural and measurement model heterogeneity is also significant and substantial (partial $\eta^2 = .105$) (Figures 4a and 4b). While the MAB for PLS-POS is always below .05, thereby indicating good parameter recovery, the MAB for FIMIX-PLS increases when measurement model heterogeneity becomes higher and structural model heterogeneity becomes lower.

Table 8 shows the MAB for each factor level in model 2 (formative measures) and reveals that the level of structural or measurement model heterogeneity only slightly affects parameter recovery for PLS-POS. In contrast, parameter recovery for FIMIX-PLS decreases with decreasing structural model heterogeneity or increasing measurement model heterogeneity. Thus, FIMIX-PLS is as good as PLS-POS in

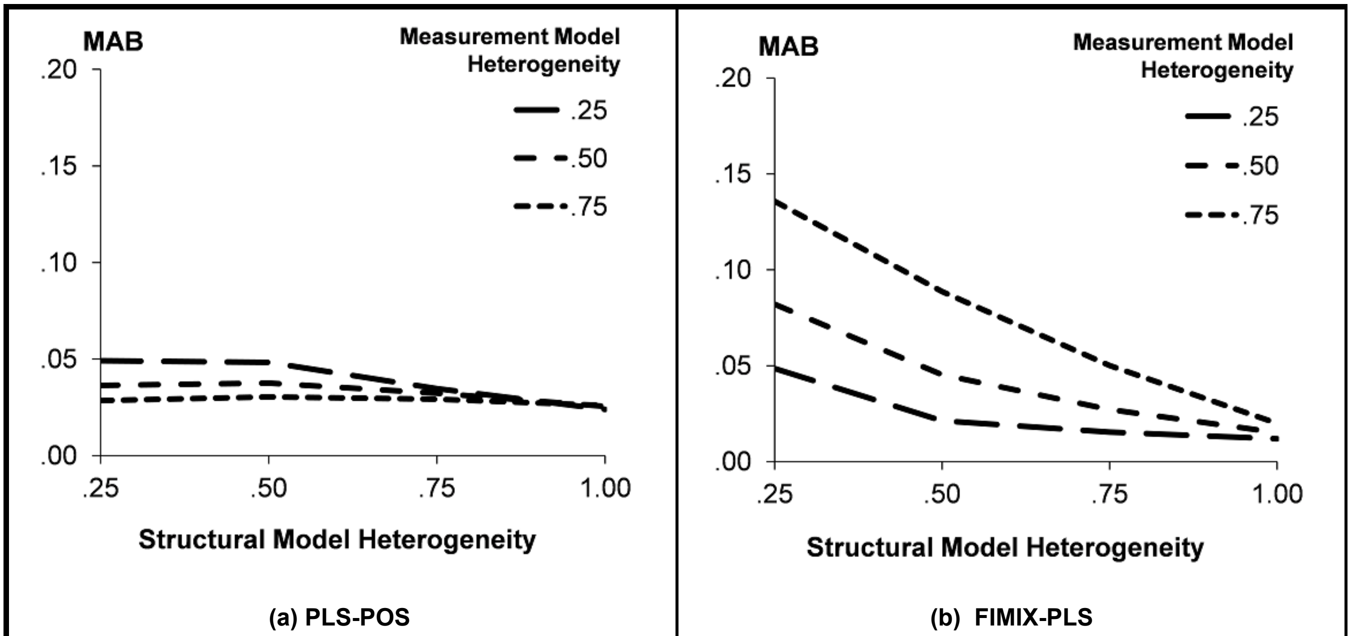


Figure 4. MAB of Both Methods for Different Structural and Measurement Model Heterogeneity

Table 8. MAB in Model 2 (Formative Measures) for Each Method

Design Factor Level		**POS** Mean Absolute Bias	**FIMIX** Mean Absolute Bias	**No Segmentation Method** Mean Absolute Bias
Structural Model Heterogeneity	.25	.038	.089	.132
	.50	.039	.052	.250
	.75	.032	.031	.375
	1.00	.025	.016	.500
Measurement Model Heterogeneity	.25	.039	.024	.312
	.50	.033	.042	.312
	.75	.029	.074	.318
R^2	.85	.057	.056	.314
	.90	.041	.050	
	.95	.025	.043	
	1.00	.011	.038	
Sample Size	100	.043	.050	.314
	200	.030	.047	
	400	.028	.043	
Data Distribution	Normal	.030	.043	.314
	Non-Normal	.037	.051	
Relative Segment Size	Equal	.029	.046	.314
	Unequal	.038	.048	
Multicollinearity	none	.031	.062	.314
	Level 1	.034	.041	
	Level 2	.036	.037	
Overall		.034	.047	.314

Table 9. Empirical Evaluation Summary of FIMIX-PLS and PLS-POS

Segmentation Method	Desired Criteria for a PLS Segmentation Method: Ability to detect heterogeneity in reflective measures	Ability to detect heterogeneity in formative measures	Ability to detect heterogeneity in the structural model	Maximizes group-specific R^2 of endogenous latent variables (prediction orientation)	Ability to handle non-normal data
FIMIX-PLS Hahn et al. 2002	*Not tested*	–	✓	✓	✓
PLS-POS	*Not tested*	✓	✓	✓	✓

Note: ✓ indicates support by the simulation experiments; – indicates that the criterion is not associated with the method.

situations with very high structural model heterogeneity regardless of the measurement model heterogeneity and also in situations where the measurement model heterogeneity is low and the structural model heterogeneity is at moderate levels. Therefore, as the results in Figures 4a and 4b reveal, the parameter recovery ability of a segmentation method cannot be assessed independently for these two types of heterogeneity.

It is worth noting that the interaction effect between method and data distribution is not substantial for either model 1 (reflective measures) or model 2 (formative measures). In addition, data distribution only has a small impact on parameter recovery in both model 1 and model 2 (direct effects of partial η^2 = .043 and partial η^2 = .028). Accordingly, we conclude that both methods perform equally well with both normal and non-normal distributions. This finding is especially interesting, as FIMIX-PLS assumes multivariate normal distributions of the endogenous latent variables, which should theoretically result in unfavorable performance with non-normal data compared to PLS-POS. However, with several indicators for each construct, the composite latent variable scores might become essentially normal even if the indicators are not. This might explain this initially surprising result.

Summary of Results

Overall, we can conclude that the use of either PLS-POS or FIMIX-PLS is better for reducing biases in parameter estimates and avoiding inferential errors than ignoring unobserved heterogeneity in PLS path models. A notable exception is when there is low structural model heterogeneity and high formative measurement model heterogeneity; in this condition, FIMIX-PLS produces results that are even more biased than those resulting from ignoring heterogeneity and estimating the model at the overall sample level. PLS-POS shows very good performance in uncovering heterogeneity for path models involving formative measures and is significantly better than FIMIX-PLS, which shows unfavorable performance when there is heterogeneity in formative measures. However, FIMIX-PLS becomes more effective when there is high multicollinearity in the formative measures, while PLS-POS consistently performs well. There are two interrelated reasons for this result: (1) multicollinearity masks heterogeneity in the measurement model, making the measures more similar (i.e., homogenous) across groups, and (2) FIMIX-PLS ignores heterogeneity in the measurement model and therefore the multicollinearity problems in formative indicators. The strongly correlated formative measures become closer to a homogenous reflective measurement of the construct. Therefore, the performance of PLS-POS and FIMIX-PLS converges in situations with high multicollinearity because FIMIX-PLS performs marginally better in purely reflective models (model 1) regardless of the distribution being normal or non-normal. However, the performance differences between FIMIX-PLS and PLS-POS are much smaller in the case of a reflective model than in the case of a formative model. Therefore, PLS-POS is more generally applicable than FIMIX-PLS to discover heterogeneity in PLS path models.

Thus, the simulation experiments provide an empirical assessment of the segmentation criteria associated with PLS-POS and FIMIX-PLS (Table 9). All criteria associated with each of these methods are supported by our findings with the exception that FIMIX-PLS does not degrade in performance with non-normal data.

A Process for Unobserved Heterogeneity Discovery

Given the availability of methods to uncover unobserved heterogeneity, as discussed in the two previous sections, researchers working with SEM face the following two major questions: when to investigate unobserved heterogeneity and how to apply methods for uncovering unobserved heterogeneity and defining segments. We address these questions by proposing a UHD process (Figure 5) and also by identifying how this process can be applied given the research objective (i.e., purely testing a model or testing and elaborating a model; Colquitt and Zapata-Phelan 2007).

How to Apply the UHD Process

When selecting an appropriate UHD method, researchers have to determine whether they are interested in evaluating unobserved heterogeneity associated with latent segments or individual-level estimates (e.g., hierarchical Bayesian approach, fixed effects, and random effects). As our focus is on the discovery of latent segments, we propose a UHD process for defining the segments in this context. In contrast, if the objective is to examine unobserved heterogeneity for individual-level estimates, the described UHD process does not apply because the methods have different assumptions and objectives and require different data (i.e., several observations per individual). The UHD process for the discovery of latent segments consists of the following three stages:

1. Selecting an appropriate UHD method

2. Applying the segmentation method to define the segments
   a. Using heuristics to narrow the range of statistically well-fitting segments
   b. Separating relevant from irrelevant segments (Are the segments *substantial*?)
   c. Testing the significance of the differences between segments (Are the segments *differentiable*?)
   d. Characterizing segments using constructs in the model/theory (Are the segments *plausible*?)
   e. Turning unobserved heterogeneity into observed heterogeneity (Are the segments *accessible*?)

3. Validating the segmentation results

Selecting an Appropriate UHD Method (Stage 1 of the UHD Process)

As discussed earlier, the methodological options for analyzing unobserved heterogeneity involving CB-SEM cover two conceptually different approaches (i.e., latent segment analysis and individual-level estimate correction). For latent segment analysis, the appropriate UHD choice is the finite mixture model as no model-based clustering alternative is available.

For analyses involving PLS path modeling, there are no methods available that address unobserved heterogeneity associated with individual-level estimates. Latent segments in PLS path modeling can be uncovered using one of the two methods we present in this paper (i.e., FIMIX-PLS and PLS-POS). Our simulation results show that FIMIX-PLS is restricted to uncovering unobserved heterogeneity in the structural model, while PLS-POS can uncover unobserved heterogeneity in both the measurement and structural models. Therefore, researchers should choose FIMIX-PLS if their models include only reflective measures and heterogeneity is expected to affect only the structural model and not the measurement model. In contrast, PLS-POS should be applied for discovering unobserved heterogeneity when PLS path models include formative measures and heterogeneity can affect both the structural and measurement models.

Applying the UHD Method to Define Segments (Stage 2 of the UHD Process)

After choosing the appropriate method for uncovering unobserved heterogeneity, the researcher has to apply the method to evaluate whether significant unobserved heterogeneity is present in the model and to define the number of segments to retain from the data. Determining the correct number of segments is important as under- or over-segmentation leads to biased results and misinterpretations. The second stage of the UHD process focuses on (1) defining with heuristics a range of statistically well-fitting segments and (2) evaluating the segments based on theoretical considerations. The steps in this stage emphasize that researchers (1) evaluate the plausibility of segments by connecting the segmentation solution to theory and (2) avoid capitalizing on data idiosyncrasies to improve the explained variance or significance of parameters.

Stage 2, Step 1: Narrow the range of statistically well-fitting segments. To determine the best fitting number of segments, the researcher has to apply the selected segmentation method for a consecutive number of segments (e.g., 1 to 10) and assess the method-specific heuristics to generate information on the number of segments that result in good model fit. Researchers have to rely on heuristics to determine a well-fitting number of segments as there is no exact statistical test to accomplish this task (McLachlan and Peel 2000). In mixture models, these heuristics include model-selection criteria that are well known from the model-selection literature (e.g., AIC, BIC, and CAIC) and can also be used to approximate the best fitting number of segments (Andrews and Currim 2003a; Sarstedt et al. 2011a).

In contrast, model-based clustering methods, such as PLS-POS, are not based on the mixture model concept and do not

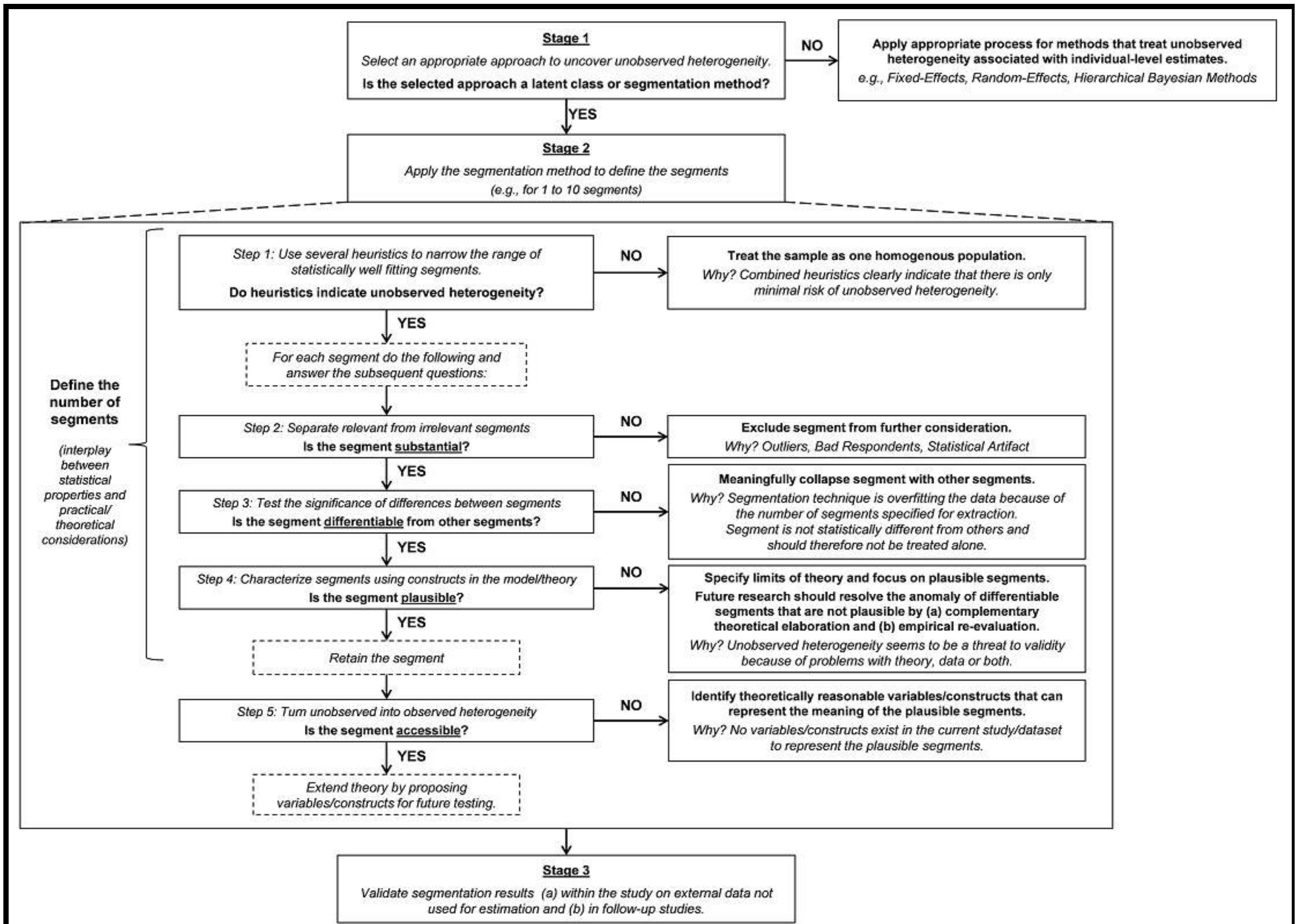


Figure 5. Unobserved Heterogeneity Discovery (UHD) Process

provide model-selection criteria. These methods require other model-specific heuristics to compare the results across different numbers of groups, for example, in terms of their average explained variance (R^2) or the increase in predictive relevance (Q^2). However, researchers should not rely purely on heuristics (e.g., model-selection criteria in finite-mixture modeling or the explained variance per segment in PLS-POS) to retain the best fitting number of segments because past studies have shown heuristics to have a low probability of finding the true number of segments. There is some empirical evidence that the best information criteria in mixture models only have about a 60 percent chance of identifying the true number of segments (Andrews and Currim 2003a, 2003b; Sarstedt et al. 2011a). Consequently, relying on heuristics can lead to strongly data-driven outcomes if the researcher fits the number of segments to the data without considering the theoretical or practical meaning of the segments. Therefore, these heuristics should only be used to narrow the range of segments for further theoretical assessment.

Regardless of whether mixture models or model-based clustering is used, if multiple heuristics clearly point to a one-segment solution, the researcher might conclude that the threat to validity from unobserved heterogeneity is low and the overall sample represents a homogenous population. This will occur when (1) the average variance explained in PLS path models for the "multisegment solution" is substantially lower than the overall sample and (2) the model-selection criteria in the mixture models collectively indicate a one-segment solution as showing the best fit and a large deterioration in fit for the best multisegment solution.

Stage 2, Step 2: Are the segments substantial? The next step after defining a range of well-fitting segments is to separate relevant from irrelevant segments. Often, segmentation methods produce very small but well-fitting segments that are likely to represent data idiosyncrasies (e.g., outliers and bad respondents). However, the problem with these very small segments is that they may (1) be irrelevant for theory or practice (e.g., outliers), (2) represent statistical artifacts or data-collection problems (e.g., bad respondents), (3) yield unreliable parameter estimates because of the small sample size, and (4) not be usable in the next step of the UHD process (i.e., multigroup difference testing). Therefore, each segment has to be large enough to represent a "real" segment; however, one also needs to be cautious when contrasting *niche* and *irrelevant* segments. Each segment should, therefore, be carefully assessed if it represents a *substantial* segment. A guideline for this analysis might be to take the average expected segment size to evaluate a segment's relevance (i.e., five segments would suggest an average expected segment size of 20%). If the segment size is considerably lower in proportion (e.g., a 2% segment size), it is a candidate for exclusion as an irrelevant segment. In addition, the total segment size should meet the minimum standards for reliable parameter estimates for the given SEM estimation method (i.e., CB-SEM and PLS path modeling). The researcher will need to determine if the segment may be a niche segment that is *substantial* and needs to be evaluated further in the next steps of the UHD process.

Stage 2, Step 3: Are the segments differentiable? To determine whether heterogeneity significantly affects the results, the *substantial* segments from the previous step need to be tested to determine the significance of group differences, assessing if a given segment is *differentiable* from others. Therefore, researchers should perform multigroup structural equation modeling or multigroup PLS analysis and assess (1) the measurement invariance/equivalence and (2) the significance of differences in path coefficients between segments. If a segment is not significantly different from other segments, researchers should consider either combining the segment meaningfully with other segments that are not significantly different from it or reducing the number of segments in the segmentation method. A reason for nonsignificant segment differences might be that the prespecified number of segments for extraction in the segmentation method has caused overfitting of the data. If no significant differences are detected among any of the segments, researchers should conclude they have a homogenous population and low validity threats due to unobserved heterogeneity.

Stage 2, Step 4: Are the segments plausible? Given a set of *differentiable* segments, the next step is to evaluate whether the segments are *plausible*. This plausibility assessment is to be conducted by characterizing the segments with the constructs in the model/theory. Each segment's theoretical plausibility should be assessed by considering the (1) segment-specific characteristics based on constructs in the model/theory; (2) the conceptual differences between the segment and other segments; and (3) the segment's theoretical or managerial relevance. If it is plausible within the specific research domain that segments can change the explanatory role of the constructs (e.g., certain types of IS users emphasize different IS characteristics, which changes the role of the constructs in predicting usage), researchers should include user type segments in their theoretical implications to avoid the premature invalidation or overgeneralization of theoretical claims based on results from the overall sample. If a segment is not theoretically plausible, it should also be considered a limitation of the theory. One possible reason for an implausible segment could be that it was mistaken as *substantial* when it actually represented outliers. Future research should solve the anomaly of *differentiable* segments that cannot be explained by (1) complementary theoretical elaboration and/or (2) empirical reevaluation. However, because unobserved heterogeneity can threaten the validity of conclusions based on the overall sample due to significant segment differences, *differentiable* segments that are not *plausible* should not be part of a combined sample used to test the model/hypotheses.

Stage 2, Step 5: Are the segments accessible? The last step in applying the segmentation methods is to turn unobserved heterogeneity into observed heterogeneity by making the segments *accessible*. Researchers can further elaborate on the theoretical meaning of the *plausible* segments by identifying additional variables (e.g., demographic, psychographic, contextual, etc.) beyond the original model that (1) help distinguish the segments by explaining the differences between retained segments and (2) determine to which segment responses belong. Statistical techniques to support this step include (1) discriminant analysis, (2) exhaustive CHAID, and (3) contingency tables where potential variables are tested for their ability to explain segment differences. However, instead of applying an *ad hoc* approach, complementary theoretical considerations should guide the process of identifying external variables. It should not be a process in which the best discriminating "left-over" variable in the dataset (that is not part of the model) is used to explain segment differences. If it is not possible to identify theoretically reasonable variables within the given dataset/study that have sufficient explanatory power to differentiate between segments, suggestions for additional variables based on complementary theoretical perspectives should guide future research.

Validating the Segmentation Results (Stage 3 of the UHD Process)

In the final stage of the UHD process, researchers should validate the segmentation results, including the number of segments, with external data not used in the estimation process. Researchers may (1) apply holdout sample validation techniques using data that are already available (Andrews et al. 2010; Bapna et al. 2011), (2) use cross-validation/

random splits to compare the stability of segmentation results (Jedidi et al. 1997), or (3) collect additional data (e.g., in a follow-up study) to evaluate the results and find new explanatory variables that match segments better to explain heterogeneity (i.e., make them *accessible*). Furthermore, repeating the segmentation study on a different population (i.e., sample) and testing the proposed explanatory variables (i.e., moderators or grouping variables) in follow-up studies increases the generalizability of the results.

When to Apply Methods to Uncover Unobserved Heterogeneity

Given a model that is grounded in substantive theory, the complexity of the social and behavioral phenomena examined in IS research makes it plausible there will be heterogeneity in any sample that is used to test and refine the model. Accordingly, we recommend that all empirical IS research should consider the discovery of unobserved heterogeneity following the UHD process just as we evaluate reliability and validity. However, researchers should (1) only use segmentation methods when substantive theory supports the model and (2) avoid using segmentation methods in models that are not well grounded in theory to merely improve the explained variance or the significance of parameters. As Jedidi et al. (1997, p. 57) observe, "one practice that should be avoided is that of fitting a … model which is not well grounded in substantive theory and simply adding segments until a reasonable fit is found." This rule applies to both CB-SEM and PLS path modeling regardless of the unobserved heterogeneity discovery method that is to be used.

For models grounded in substantive theory, the objectives for discovering unobserved heterogeneity can differ depending on the study's research objectives. If the research objective is theory testing (i.e., testers; Colquitt and Zapata-Phelan 2007), uncovering unobserved heterogeneity serves as a validity check to safeguard against biases and the false rejection or false confirmation of theoretical claims. When the theory tester uncovers unobserved heterogeneity in the sample (i.e., significant segment differences are detected and the segments are determined to be theoretically plausible), he/she has evidence of a theoretical breakdown given the segments. As such, the discovery of unobserved heterogeneity safeguards against (1) premature invalidation of theoretical claims (i.e., the results based on the overall sample suggest certain relationships are nonsignificant, but the significance of these relationships is actually masked by the heterogeneity) and (2) premature overgeneralization of theoretical claims (i.e., the model/theory holds in some segments and not in others, thus requiring qualifiers for support found for the theory in different segments). Hence, theory testers apply the UHD process to evaluate validity threats due to unobserved heterogeneity. If significant differences across *plausible* segments are detected, researchers should revise the boundary conditions for the theory (i.e., specify within which *plausible* segments the theory was supported and in which it was not). If unobserved heterogeneity is not uncovered in the sample (i.e., no significant differences across segments are detected; segments are not *differentiable*), the researcher can continue with the standard analysis on the overall sample, (in)validate theoretical claims, and note that the validity of the findings is not threatened by unobserved heterogeneity.

If the research objective is theory testing and elaboration (i.e., expanders; Colquitt and Zapata-Phelan 2007), uncovering unobserved heterogeneity not only serves as a validity check but can also guide researchers to identify variables explaining the uncovered segments and to integrate these variables to expand the model/theory. Hence, researchers should turn unobserved heterogeneity into observed heterogeneity by (1) advancing theoretical reasons to explain the differences between segments; (2) identifying constructs beyond the original model that explain these differences, thereby making the segments *accessible*; and (3) expanding the model/theory by integrating the constructs that make the segments accessible. Accordingly, the accessibility stage in the UHD process will be facilitated when researchers anticipate this task during the research design, identify complementary theoretical perspectives and corresponding constructs, and collect additional data for these constructs that can be instrumental in making the segments accessible. Of course, these considerations require extra effort and data-collection costs and should be accommodated in a study when the researcher expects unobserved heterogeneity (e.g., based on inconsistent results in past studies, meta-analysis, the nature of phenomena, etc.).

We note that the discovery of unobserved heterogeneity for theoretical tests and elaboration is relevant even when existing theory offers *a priori* knowledge about observed heterogeneity (e.g., age, gender, or income). There can be additional explainable and generalizable heterogeneity beyond the known heterogeneity (e.g., experienced versus inexperienced users) that threatens the theoretical validity of the test and, when discovered, can be used to elaborate theory/models.

As an illustration, assume that the research objective is to test the baseline technology acceptance model presented in the introduction. Based on the analysis of the overall sample, the researcher risks overgeneralization in that the effects of PU and PEOU are always important for IU. To avert this risk, the researcher applies the UHD process and discovers two *substantial* and *differentiable* segments. One segment shows a strong positive relationship between PU and IU and a weak, or nonsignificant, relationship between PEOU and IU. In contrast, the other segment shows a strong positive relationship between PEOU and IU and a weak, or nonsignificant, relationship between PU and IU (Figure 1a). The researcher concludes that these two identified segments (i.e., users emphasizing PU or PEOU) are theoretically plausible (i.e., within TAM, it is reasonable that there are different users who emphasize different system characteristics) and conceptually important for the theory. In contrast to the results derived from the overall sample, only one of the posited TAM constructs influences IU in each segment. As such, the researcher (1) does not overgeneralize the theory by assuming that it will always be applicable, (2) acknowledges there are user segments that determine which construct is influential for IU, and (3) specifies the need to make the segments *accessible*, thereby expanding the TAM model.

Given the study's objective (i.e., theory testing) and the limited availability of additional data (e.g., a lack of demographic or psychographic variables, such as experience), researchers might end the UHD process after concluding the segments are *plausible* (i.e., that it is plausible that the segments change the explanatory role of the constructs) without explaining which users belong to which segment (i.e., without making the segments *accessible*).

Instead, if the research objective is theory testing and elaboration, researchers should continue to find complementary theoretical explanations to make the segments *accessible* (i.e., to give additional theoretical meaning to the segments). A complementary theory could explain that users' experience influences their appreciation of system characteristics (e.g., PEOU and PU). Experience, therefore, could be an external variable/construct that, if available in the dataset, could be tested for explaining the segment membership. Other plausible theoretical considerations could suggest other variables/constructs that might explain segment membership and should be evaluated (e.g., age, income, computer anxiety, task type, subjective norms, etc.). If researchers are able to identify a variable/construct that explains the segment membership (i.e., makes segments *accessible*), the unobserved heterogeneity is turned into observed heterogeneity, thereby expanding the theory with new constructs accounting for the group differences (e.g., a moderator). If researchers are unable to assess the ability of variables/constructs to explain segment membership because of lack of data in the study, they can only theoretically identify reasonable variables/constructs for future testing.

Limitations and Future Research

In this study, we (1) discussed why unobserved heterogeneity is an important issue in IS research, (2) identified threats to validity due to unobserved heterogeneity, (3) synthesized current work on unobserved heterogeneity in CB-SEM and PLS path modeling, (4) introduced a new segmentation method (PLS-POS) for PLS path modeling, (5) assessed its performance and that of FIMIX-PLS, and (6) provided guidelines for researchers on when and how to uncover unobserved heterogeneity. While our study makes contributions, it has its limitations and opens up avenues for future research.

First, the validity and generalizability of simulation studies are limited by the choice of design factors and factor levels. We focused on eight factors based on past studies on PLS path modeling or segmentation. The analysis of all factor-level combinations of the two PLS path models entailed 126,720 simulated segmentation runs for assessing the performance of PLS-POS and FIMIX-PLS. The inclusion of additional design factors—namely, those that are theoretically less important for PLS segmentation—or additional factor levels would have increased the complexity of the simulations exponentially and is beyond the scope of a single study. Therefore, researchers should also apply PLS-POS and FIMIX-PLS in a broad range of empirical studies to find additional evidence of the methods' abilities to detect unobserved heterogeneity.

Second, heterogeneity is a special type of endogeneity problem (i.e., omitted group variables). Future studies may want to evaluate the impact of other types of endogeneity problems (e.g., reciprocal relationships) on PLS path modeling results. As PLS path modeling cannot handle non-recursive models, these issues might also threaten the consistency of parameters. In addition, researchers may want to assess the effect of unobserved heterogeneity in models that do not comply with the recursive nature of models imposed by PLS path models. If heterogeneity affects non-recursive (reciprocal) relationships, it might have a strong impact on the ability of both PLS segmentation methods (FIMIX-PLS and PLS-POS) to uncover unobserved heterogeneity.

Third, this research does not focus on the parameter settings of the methods or the time needed to arrive at the final segmentation solution. Our simulations suggest that PLS-POS is more time consuming than FIMIX-PLS.[14] Determining efficient parameter settings to reduce the computational effort of PLS-POS represents another avenue for future research.

[14]In absolute terms, PLS-POS works within acceptable timeframes. Applying both methods to the ECSI mobile phone dataset from Tenenhaus et al. (2005) with two segments, the FIMIX-PLS algorithm needs approximately 10 seconds, while PLS-POS requires about 3 minutes to arrive at a solution. (We used a Windows 7 PC with an Intel Core 2 T7300 2GHz and 2GB RAM.) We believe this should be acceptable to researchers in an advanced stage of model investigation.

Conclusion

We differentiated between observed and unobserved heterogeneity and showed why unobserved heterogeneity biases structural equation model estimates, leads to Type I and Type II errors, and is a threat to different types of validity (i.e., internal, instrumental, statistical conclusion, and external). We demonstrated that heterogeneity is present in empirical IS research across various IS phenomena by presenting evidence from 12 meta-analyses showing that inconsistent findings are prevalent across IS studies with unobserved heterogeneity being a plausible cause for these inconsistencies. We explained how researchers can avoid threats to validity due to unobserved heterogeneity in structural equation modeling by using different methods that have been proposed in the literature to uncover unobserved heterogeneity. The application of these methods not only safeguards against biases and validity threats but also facilitates theory development by promoting abduction (Van de Ven 2007). Specifically, uncovering unobserved heterogeneity and explaining segments with new constructs beyond those in the model allows researchers to develop additional theoretical descriptions that make segments *accessible*. Thereby, they can expand and further develop existing theory.

We introduced a new segmentation method for PLS path modeling—PLS-POS—that overcomes some of the restrictive assumptions associated with FIMIX-PLS and other distance measure-based methods, and we evaluated the ability of the FIMIX-PLS and PLS-POS methods to uncover unobserved heterogeneity in PLS path models. Our findings show that both FIMIX-PLS and PLS-POS alleviate threats to validity from unobserved heterogeneity by providing considerably less biased parameter estimates than those that are based on invalid assumptions of homogenous data. However, FIMIX-PLS is restricted to uncovering unobserved heterogeneity in the structural model, while PLS-POS can uncover unobserved heterogeneity in *both* the measurement and structural models. Our results show that the parameter recovery of PLS-POS and FIMIX-PLS is comparable for those PLS path models in which all measures are reflective (with measurement invariance across groups) and that heterogeneity is limited to the structural model. PLS-POS performs very well in uncovering heterogeneity across all types of PLS path models with different locations of heterogeneity in the model (structural model, measurement model, or both) and different data conditions (sample size, relative segment sizes, multicollinearity, and data distribution).

Our findings also reveal that unobserved heterogeneity in formative measures and in the structural model should be evaluated collectively. As FIMIX-PLS does not uncover heterogeneity in measurement models, PLS-POS should be applied for discovering unobserved heterogeneity if PLS path models include formative measures. This finding is particularly important because formative measurement models are often used in IS research. A comprehensive analysis of the application of PLS path models in *MIS Quarterly* over the last 20 years indicates that about 42 percent of the models use only reflective measures, about 32 percent of the models use formative measures, and about a quarter of the studies/models do not explicitly state which measurement model was used (Ringle et al. 2012). In addition, the number of studies using formative measures in IS research has increased over time.

While there is an ongoing discussion on the interpretation and use of formative measures (Aguirre-Urreta and Marakas 2012; Diamantopoulos 2011; Edwards 2010; Jarvis et al. 2012; Petter et al. 2012), there is general consensus that the theoretical meaning of a construct should correspond to its empirical meaning and that some theoretical constructs fit formative specifications better than reflective specification (Bagozzi 2011; Diamantopoulos and Winklhofer 2001; Jarvis et al. 2012; Petter et al. 2007). As Bagozzi (2011) notes, there are different ontologies underlying formative and reflective measures, which have different accompanying approaches for interpreting and assessing the construct and its relationships with other constructs. If researchers have chosen a formative ontology, the discovery of unobserved heterogeneity in formative indicator weights can assist them in evaluating plausible differences in the construct's theoretical or empirical meaning between groups, thereby safeguarding against interpretational confounds.

It is important to note that we do not recommend using segmentation methods (including FIMIX-PLS and PLS-POS) for *post hoc* data-driven improvement of results where researchers engage in "fishing expeditions" with the objective of improving the significance of an association or the predictive power of the model, as described earlier in the section on the UHD process. Instead, consistent with Jedidi et al. (1997) and Van de Ven (2007), we take the position that theory development in the social and behavioral sciences does not need to be confined to deductive reasoning. Moreover, in situations in which the researcher discovers anomalies that must be resolved through theoretical elaboration, theory development is significantly enhanced by abduction. Segmentation provides a mechanism to facilitate abduction by surfacing anomalies, which must then be confronted and resolved theoretically. Using the presented methods in PLS path modeling and CB-SEM within the UHD process is a possible way to achieve this goal.

Acknowledgments

We thank the senior editor, Ron Thomson, the associate editor, and the reviewers for their constructive comments and valuable suggestions. We also appreciate the comments from Ed Rigdon and Detmar Straub from Georgia State University on our motivation and initial ideas for this study.

References

Aguirre-Urreta, M. I., and Marakas, G. M. 2012. "Revisiting Bias Due to Construct Misspecification: Different Results from Considering Coefficients in Standardized Form," *MIS Quarterly* (36:1), pp. 123-138.

Allenby, G. M., and Rossi, P. E. 1998. "Marketing Models of Consumer Heterogeneity," *Journal of Econometrics* (89:1-2), pp. 57-78.

Anderberg, M. R. 1973. *Cluster Analysis for Applications*, New York: Academic Press.

Andrews, R. L., Brusco, M. J., and Currim, I. S. 2010. "Amalgamation of Partitions from Multiple Segmentation Bases: A Comparison of Non-Model-Based and Model-Based Methods," *European Journal of Operational Research* (201:2), pp. 608-618.

Andrews, R. L., and Currim, I. S. 2003a. "A Comparison of Segment Retention Criteria for Finite Mixture Logit Models," *Journal of Marketing Research* (40:20), pp. 235-243.

Andrews, R. L., and Currim, I. S. 2003b. "Retention of Latent Segments in Regression-Based Marketing Models," *International Journal of Research in Marketing* (20:4), pp. 315-321.

Ansari, A., Jedidi, K., and Jagpal, S. 2000. "A Hierarchical Bayesian Methodology for Treating Heterogeneity in Structural Equation Models," *Marketing Science* (19:4), pp. 328-347.

Arminger, G., Stein, P., and Wittenberg, J. 1999. "Mixtures of Conditional Mean- and Covariance-Structure Models," *Psychometrika* (64:4), pp. 475-494.

Bagozzi, R. P. 2011. "Measurement and Meaning in Information Systems and Organizational Research: Methodological and Philosophical Foundations," *MIS Quarterly* (35:2), pp. 261-292.

Bapna, R., Goes, P., Kwok Kee, W., and Zhongju, Z. 2011. "A Finite Mixture Logit Model to Segment and Predict Electronic Payments System Adoption," *Information Systems Research* (22:1), pp. 118-133.

Bart, Y., Shankar, V., Sultan, F., and Urban, G. L. 2005. "Are the Drivers and Role of Online Trust the Same for All Web Sites and Consumers? A Large-Scale Exploratory Empirical Study," *The Journal of Marketing* (69:4), pp. 133-152.

Cai, J.-H., and Song, X.-Y. 2010. "Bayesian Analysis of Mixtures in Structural Equation Models with Non-Ignorable Missing Data," *British Journal of Mathematical and Statistical Psychology* (63:3), pp. 491-508.

Cenfetelli, R. T., and Bassellier, G. 2009. "Interpretation of Formative Measurement in Information Systems Research," *MIS Quarterly* (33:4), pp. 689-707.

Chin, W. W. 1998. "The Partial Least Squares Approach to Structural Equation Modeling," in *Modern Methods for Business Research*, G. A. Marcoulides (ed.), Mahwah, NJ: Erlbaum, pp. 295-358.

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 *Handbook of Partial Least Squares: Concepts, Methods and Applications,* V. Esposito Vinzi, W. W. Chin, J. Henseler, and H. Wang (eds.), Berlin: Springer, pp. 171-193.

Chin, W. W., Marcolin, B. L., and Newsted, P. R. 2003. "A Partial Least Squares Latent Variable Modeling Approach for Measuring Interaction Effects: Results from a Monte Carlo Simulation Study and an Electronic-Mail Emotion/Adoption Study," *Information Systems Research* (14:2), pp. 189-217.

Collier, J. E., and Bienstock, C. C. 2009. "Model Misspecification: Contrasting Formative and Reflective Indicators for a Model of E-Service Quality," *Journal of Marketing Theory & Practice* (17:3), pp. 283-293.

Colquitt, J. A., and Zapata-Phelan, C. P. 2007. "Trends in Theory Building and Theory Testing: A Five-Decade Study of the Academy of Management Journal," *Academy of Management Journal* (50:6), pp. 1281-1303.

Cook, T. D., and Campbell, D. T. 1976. "The Design and Conduct of Quasi-Experiments and True Experiments in Field Settings," in *Handbook of Industrial and Organizational Psychology,* M. D. Dunnette (ed.), Chicago: Rand McNally, pp. 223-326.

Cook, T. D., and Campbell, D. T. 1979. *Quasi-Experimentation: Design and Analysis Issues for Field Settings*, Chicago: Rand-McNally.

Davis, F. D., Bagozzi, R. P., and Warshaw, P. R. 1989. "User Acceptance of Computer Technology: A Comparison of Two Theoretical Models," *Management Science* (35:8), pp. 982-1003.

DeSarbo, W. S., and Cron, W. L. 1988. "A Maximum Likelihood Methodology for Clusterwise Linear Regression," *Journal of Classification* (5:2), pp. 249-282.

DeSarbo, W. S., Di Benedetto, C. A., Jedidi, K., and Song, M. 2006. "Identifying Sources of Heterogeneity for Empirically Deriving Strategic Types: A Constrained Finite-Mixture Structural-Equation Methodology," *Management Science* (52:6), pp. 909-924.

Desarbo, W. S., Ramaswamy, V., and Cohen, S. H. 1995. "Market Segmentation with Choice-Based Conjoint Analysis," *Marketing Letters* (6:2), pp. 137-147.

Diamantopoulos, A. 2011. "Incorporating Formative Measures into Covariance-Based Structural Equation Models," *MIS Quarterly* (35:2), pp. 335-358.

Diamantopoulos, A., and Papadopoulos, N. 2010. "Assessing the Cross-National Invariance of Formative Measures: Guidelines for International Business Researchers," *Journal of International Business Studies* (41:2), pp. 360-370.

Diamantopoulos, A., Riefler, P., and Roth, K. P. 2008. "Advancing Formative Measurement Models," *Journal of Business Research* (61:12), pp. 1203-1218.

Diamantopoulos, A., and Winklhofer, H. M. 2001. "Index Construction with Formative Indicators: An Alternative to Scale Development," *Journal of Marketing Research* (38:2), pp. 269-277.

Dolan, C., and van der Maas, H. 1998. "Fitting Multivariage Normal Finite Mixtures Subject to Structural Equation Modeling," *Psychometrika* (63:3), pp. 227-253.

Edmondson, A. C., and McManus, S. E. 2007. "Methodological Fit in Management Field Research," *Academy of Management Review* (32:4), pp. 1155-1179.

Edwards, J. R. 2010. "The Fallacy of Formative Measurement," *Organizational Research Methods* (14:2), pp. 370-388.

Edwards, J. R., and Lambert, L. S. 2007. "Methods for Integrating Moderation and Mediation: A General Analytical Framework Using Moderated Path Analysis," *Psychological Methods* (12:1), pp. 1-22.

Esposito Vinzi, V., Trinchera, L., and Amato, S. 2010. "PLS Path Modeling: From Foundations to Recent Developments and Open Issues for Model Assessment and Improvement," in *Handbook of Partial Least Squares: Concepts, Methods and Applications*, V. Esposito Vinzi, W. W. Chin, J. Henseler, and H. Wang (eds.), Berlin: Springer, pp. 47-82.

Esposito Vinzi, V., Trinchera, L., Squillacciotti, S., and Tenenhaus, M. 2008. "REBUS-PLS: A Response-Based Procedure for Detecting Unit Segments in PLS Path Modelling," *Applied Stochastic Models in Business & Industry* (24:5), pp. 439-458.

Fraley, C., and Raftery, A. 2002. "Model-Based Clustering, Discriminant Analysis, and Density Estimation," *Journal of the American Statistical Association* (97:458), pp. 611-631.

Gilbride, T. J., Allenby, G. M., and Brazell, J. D. 2006. "Models for Heterogeneous Variable Selection," *Journal of Marketing Research* (43:3), pp. 420-430.

Goodhue, D., Lewis, W., and Thompson, R. 2007. "Statistical Power in Analyzing Interaction Effects: Questioning the Advantage of PLS with Product Indicators," *Information Systems Research* (18:2), pp. 211-227.

Haenlein, M., and Kaplan, A. M. 2011. "The Influence of Observed Heterogeneity on Path Coefficient Significance: Technology Acceptance Within the Marketing Discipline," *The Journal of Marketing Theory and Practice* (19:2), pp. 153-168.

Hahn, C., Johnson, M. D., Herrmann, A., and Huber, F. 2002. "Capturing Customer Heterogeneity Using a Finite Mixture PLS Approach," *Schmalenbach Business Review (SBR)* (54:3), pp. 243-269.

Heeler, R. M., and Ray, M. L. 1972. "Measure Validation in Marketing," *Journal of Marketing Research* (9:4), pp. 361-370.

Henseler, J., and Chin, W. W. 2010. "A Comparison of Approaches for the Analysis of Interaction Effects Between Latent Variables Using Partial Least Squares Path Modeling," *Structural Equation Modeling: A Multidisciplinary Journal* (17:1), pp. 82-109.

Henson, J. M., Reise, S. P., and Kim, K. H. 2007. "Detecting Mixtures from Structural Model Differences Using Latent Variable Mixture Modeling: A Comparison of Relative Model Fit Statistics," *Structural Equation Modeling* (14:2), pp. 202-226.

Hsieh, J. J. P.-A., Rai, A., and Keil, M. 2008. "Understanding Digital Inequality: Comparing Continued Use Behavioral Models of the Socio-Economically Advantaged and Disadvantaged," *MIS Quarterly* (32:1), pp. 97-126.

Jaccard, J., and Wan, C. K. 1995. "Measurement Error in the Analysis of Interaction Effects Between Continuous Predictors Using Multiple Regression: Multiple Indicator and Structural Equation Approaches," *Psychological Bulletin* (117:2), pp. 348-357.

Jarvis, C. B., MacKenzie, S. B., and Podsakoff, P. M. 2003. "A Critical Review of Construct Indicators and Measurement Model Misspecification in Marketing and Consumer Research," *Journal of Consumer Research* (30:2), pp. 199-218.

Jarvis, C. B., MacKenzie, S. B., and Podsakoff, P. M. 2012. "The Negative Consequences of Measurement Model Misspecification: A Response to Aguirre-Urreta and Marakas," *MIS Quarterly* (36:1), pp. 139-146.

Jedidi, K., Jagpal, H. S., and DeSarbo, W. S. 1997. "Finite-Mixture Structural Equation Models for Response-Based Segmentation and Unobserved Heterogeneity," *Marketing Science* (16:1), pp. 39-59.

Johns, G. 2006. "The Essential Impact of Context on Organizational Behavior," *The Academy of Management Review* (31:2), pp. 386-408.

Jöreskog, K. G. 1971. "Simultaneous Factor Analysis in Several Populations," *Psychometrika* (36:4), pp. 409-426.

Jöreskog, K. G. 1978. "Structural Analysis of Covariance and Correlation Matrices," *Psychometrika* (43:4), pp. 443-477.

Jöreskog, K. G. 1982. "The LISREL Approach to Causal Model-Building in the Social Sciences" in *Systems Under Indirect Observation, Part I*, H. Wold and K. G. Jöreskog (eds.), Amsterdam: North-Holland, pp. 81-100.

Jöreskog, K. G., and Yang, F. 1996. "Nonlinear Structural Equation Models: The Kenny-Judd Model with Interaction Effects," in *Advanced Structural Equation Modeling: Issues and Techniques,* G. A. Marcoulides and R. E. Schumacker (eds.), Mahwah, NJ: Lawrence Earlbaum Associates, pp. 57-87.

King, W. R., and He, J. 2006. "A Meta-Analysis of the Technology Acceptance Model," *Information & Management* (43:6), pp. 740-755.

Klein, A., and Moosbrugger, H. 2000. "Maximum Likelihood Estimation of Latent Interaction Effects with the LMS Method," *Psychometrika* (65:4), pp. 457-474.

Lee, S.-Y., and Song, X.-Y. 2003. "Bayesian Analysis of Structural Equation Models with Dichotomous Variables," *Statistics in Medicine* (22:19), pp. 3073-3088.

Lenk, P. J., DeSarbo, W. S., Green, P. E., and Young, M. R. 1996. "Hierarchical Bayes Conjoint Analysis: Recovery of Partworth Heterogeneity from Reduced Experimental Design," *Marketing Science* (15:2), pp. 173-191.

Lohmöller, J.-B. 1989. *Latent Variable Path Modeling with Partial Least Squares*, Heidelberg: Physica.

Lubke, G. H., and Muthén, B. 2005. "Investigating Population Heterogeneity With Factor Mixture Models," *Psychological Methods* (10:1), pp. 21-39.

Luo, L., Kannan, P. K., and Ratchford, B. T. 2008. "Incorporating Subjective Characteristics in Product Design and Evaluations," *Journal of Marketing Research* (45:2), pp. 182-194.

Mason, C. H., and Perreault, W. D. 1991. "Collinearity, Power, and Interpretation of Multiple Regression Analysis," *Journal of Marketing Research* (28:3), pp. 268-280.

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

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* (45:5-6), pp. 395-423.

Muthén, B. O. 1989. "Latent Variable Modeling in Heterogeneous Populations," *Psychometrika* (54:4), pp. 557-585.

Muthén, B. O. 1994. "Multilevel Covariance Structure Analysis," *Sociological Methods & Research* (22:3), pp. 376-398.

Palumbo, F., Romano, R., and Esposito Vinzi, V. 2008. "Fuzzy PLS Path Modeling: A New Tool For Handling Sensory Data," in *Data Analysis, Machine Learning and Applications: Proceedings of the 31st Annual Conference of the Gesellschaft für Klassifikation*, C. Preisach, H. Burkhardt, L. Schmidt-Thieme, and R. Decker (eds.), Berlin: Springer, pp. 689-696.

Papies, D., and Clement, M. 2008. "Adoption of New Movie Distribution Services on the Internet," *Journal of Media Economics* (21:3), pp. 131-157.

Parasuraman, A., Zeithaml, V. A., and Berry, L. L. 1988. "SERVQUAL: A Multiple-Item Scale for Measuring Consumer Perceptions of Service Quality," *Journal of Retailing* (64:1), pp. 12-40.

Petter, S., Rai, A., and Straub, D. 2012. "The Critical Importance of Construct Measurement Specification: A Response to Aguirre-Urreta and Marakas," *MIS Quarterly* (36:1), pp. 147-156.

Petter, S., Straub, D., and Rai, A. 2007. "Specifying Formative Constructs in Information Systems Research," *MIS Quarterly* (31:4), pp. 623-656.

Popkowski Leszczyc, P. T., and Bass, F. M. 1998. "Determining the Effects of Observed and Unobserved Heterogeneity on Consumer Brand Choice," *Applied Stochastic Models and Data Analysis* (14:2), pp. 95-115.

Qureshi, I., and Compeau, D. 2009. "Assessing Between-Group Differences in Information Systems Research: A Comparison of Covariance- and Component-Based SEM," *MIS Quarterly* (33:1), pp. 197-214.

Rabe-Hesketh, S., Skrondal, A., and Pickles, A. 2004. "Generalized Multilevel Structural Equation Modeling," *Psychometrika* (69:2), pp. 167-190.

Rai, A., Patnayakuni, R., and Seth, N. 2006. "Firm Performance Impacts of Digitally Enabled Supply Chain Integration Capabilities," *MIS Quarterly* (30:2), pp. 225-246.

R Core Team. 2013. "R: A Language and Environment for Statistical Computing," R Foundation for Statistical Computing, Vienna.

Reinartz, W. J., Echambadi, R., and Chin, W. W. 2002. "Generating Non-Normal Data for Simulation of Structural Equation Models Using Mattson's Method," *Multivariate Behavioral Research* (37:2), pp. 227-244.

Reinartz, W. J., Haenlein, M., and Henseler, J. 2009. "An Empirical Comparison of the Efficacy of Covariance-Based and Variance-Based SEM," *International Journal of Research in Marketing* (26:4), pp. 332-344.

Reinecke, J. 2006. "Special Issue: Mixture Structural Equation Modeling," *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences* (2:3), pp. 83-85.

Rigdon, E. E., Ringle, C. M., and Sarstedt, M. 2010. "Structural Modeling of Heterogeneous Data with Partial Least Squares," in *Review of Marketing Research,* N. K. Malhotra (ed.), Armonk, NY: M. E. Sharpe, pp. 255-296.

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* (8), pp. 19-49.

Ringle, C. M., Sarstedt, M., and Schlittgen, R. 2010b. "Finite Mixture and Genetic Algorithm Segmentation in Partial Least Squares Path Modeling: Identification of Multiple Segments in a Complex Path Model," in *Advances in Data Analysis, Data Handling and Business Intelligence*, A. Fink, B. Lausen, W. Seidel, and A. Ultsch (eds.), Berlin: Springer, pp. 167-176.

Ringle, C. M., Sarstedt, M., Schlittgen, R., and Taylor, C. R. 2013. "PLS Path Modeling and Evolutionary Segmentation," *Journal of Business Research*, forthcoming.

Ringle, C. M., Sarstedt, M., and Straub, D. 2012. "A Critical Look at the Use of PLS-SEM in *MIS Quarterly*," *MIS Quarterly* (36:1), pp. iii-viii.

Ringle, C. M., Wende, S., and Will, A. 2005. "SmartPLS 2.0," www.smartpls.de.

Rust, R. T., and Verhoef, P. C. 2005. "Optimizing the Marketing Interventions Mix in Intermediate-Term CRM," *Marketing Science* (24:3), pp. 477-489.

Sánchez, G. 2009. "PATHMOX Approach: Segmentation Trees in Partial Least Squares Path Modeling," unpublished doctoral dissertation, Universitat Politècnica de Catalunya.

Sánchez, G., and Aluja, T. 2006. "PATHMOX: A PLS-PM Segmentation Algorithm," in *Proceedings of the IASC Symposium on Knowledge Extraction by Modelling*, International Association for Statistical Computing Island of Capri, Italy.

Sánchez, G., and Aluja, T. 2012. "R Package pathmox: Segmentation Trees in Partial Least Squares Path Modeling (Version 0.1-1)," http://cran.r-project.org/web/packages/pathmox/.

Sánchez, G., and Trinchera, L. 2013. "R Package PLSPM (version 0.3.5)," http://cran.r-project.org/web/packages/plspm/.

Sarstedt, M. 2008. "A Review of Recent Approaches for Capturing Heterogeneity in Partial Least Squares Path Modelling," *Journal of Modelling in Management* (3:2), 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* (63:1), 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 *Advances in International Marketing, Volume 22,* M. Sarstedt, M. Schwaiger and C. R. Taylor (eds.), Bingley, UK: Emerald Group Publishing Limited, pp. 195-218.

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* (37:8), pp. 1299-1318.

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* (3:3), pp. 185-206.

Sörbom, D. 1974. "A General Method for Studying Differences in Factor Means and Factor Structure between Groups," *British Journal of Mathematical and Statistical Psychology* (27:2), pp. 229-239.

Späth, H. 1979. "Algorithm 39 Clusterwise Linear Regression," *Computing* (22:4), pp. 367-373.

Squillacciotti, S. 2005. "Prediction Oriented Classification in PLS Path Modeling," in *PLS & Marketing: Proceedings of the 4th International Symposium on PLS and Related Methods*, T. Aluja,

J. Casanovas, V. Esposito Vinzi and M. Tenenhaus (eds.), Paris: DECISIA, pp. 499-506.

Squillacciotti, S. 2010. "Prediction-Oriented Classification in PLS Path Modeling," in *Handbook of Partial Least Squares: Concepts, Methods and Applications*, V. Esposito Vinzi, W. W. Chin, J. Henseler, and H. Wang (eds.), Berlin: Springer, pp. 219-233.

Srite, M., and Karahanna, E. 2006. "The Role of Espoused National Cultural Values in Technology Acceptance," *MIS Quarterly* (30:3), pp. 679-704.

Steenkamp, J.-B. E. M., and Baumgartner, H. 1998. "Assessing Measurement Invariance in Cross-National Consumer Research," *Journal of Consumer Research* (25:1), pp. 78-90.

Straub, D. W. 1989. "Validating Instruments in MIS Research," *MIS Quarterly* (13:2), pp. 147-169.

Tenenhaus, M., Esposito Vinzi, V., Chatelin, Y.-M., and Lauro, C. 2005. "PLS Path Modeling," *Computational Statistics & Data Analysis* (48:1), pp. 159-205.

Tueller, S., and Lubke, G. 2010. "Evaluation of Structural Equation Mixture Models: Parameter Estimates and Correct Class Assignment," *Structural Equation Modeling: A Multidisciplinary Journal* (17:2), pp. 165-192.

Van de Ven, A. H. 2007. *Engaged Scholarship: A Guide for Organizational and Social Research*, New York: Oxford University Press

Vandenberg, R. J., and Lance, C. E. 2000. "A Review and Synthesis of the Measurement Invariance Literature: Suggestions, Practices, and Recommendations for Organizational Research," *Organizational Research Methods* (3:1), pp. 4-70.

Venkatesh, V. 2000. "Determinants of Perceived Ease of Use: Integrating Control, Intrinsic Motivation, and Emotion into the Technology Acceptance Model," *Information Systems Research* (11:4), pp. 342-365.

Venkatesh, V., and Bala, H. 2008. "Technology Acceptance Model 3 and a Research Agenda on Interventions," *Decision Sciences* (39:2), pp. 273-315.

Venkatesh, V., and Davis, F. D. 2000. "A Theoretical Extension of the Technology Acceptance Model: Four Longitudinal Field Studies," *Management Science* (46:2), pp. 186-204.

Venkatesh, V., and Morris, M. G. 2000. "Why Don't Men Ever Stop to Ask for Directions? Gender, Social Influence, and Their Role in Technology Acceptance and Usage Behavior," *MIS Quarterly* (24:1), pp. 115-139.

Venkatesh, V., Morris, M. G., Davis, G. B., and Davis, F. D. 2003. "User Acceptance of Information Technology: Toward a Unified View," *MIS Quarterly* (27:3), pp. 425-478.

Wang, J., and Keil, M. 2007. "A Meta-Analysis Comparing the Sunk Cost Effect for IT and Non-IT Projects," *Information Resources Management Journal* (20:3), pp. 1-18.

Wedel, M., and DeSarbo, W. S. 1994. "A Review of Latent Class Regression Models and their Applications," in *Advanced Methods for Marketing Research,* R. P. Bagozzi (ed.), Cambridge, UK: Blackwell Business, pp. 353-388.

Wedel, M., and Kamakura, W. 2000. *Market Segmentation: Conceptual and Methodological Foundations* (2nd ed.), New York: Kluwer Academic Publishers.

Wetzels, M., Odekerken-Schröder, G., and van Oppen, C. 2009. "Using PLS Path Modeling for Assessing Hierarchical Construct Models: Guidelines and Empirical Illustration," *MIS Quarterly* (33:1), pp. 177-195.

Wold, H. 1982. "Soft Modeling: The Basic Design and Some Extensions," in *Systems Under Indirect Observations: Part I*, K. G. Jöreskog and H. Wold (eds.), Amsterdam: North-Holland, pp. 1-54..

Wu, J., and Lederer, A. 2009. "A Meta-Analysis of the Role of Environment-Based Voluntariness in Information Technology Acceptance," *MIS Quarterly* (33:2), pp. 419-432.

About the Authors

Jan-Michael Becker is a postdoctoral researcher at the University of Cologne. He received his doctoral degree in Marketing from the University of Cologne, Germany and his diploma in Information Systems from the University of Hamburg, Germany. He has been a visiting scholar at Georgia State University several times. His research interests focus on structural equation modeling, PLS path modeling, unobserved heterogeneity, and mixture models, as well as brand management and bridging marketing and IS problems.

Arun Rai is Regents' Professor and the Harkins Chair in the Center for Process Innovation and the Department of Computer Information Systems at the Robinson College of Business, Georgia State University. His research has examined how firms can leverage information technologies in their strategies, inter-firm relationships, and processes, and how systems can be successfully developed and implemented. His articles have appeared in *Management Science, MIS Quarterly, Information Systems Research, Journal of Management Information Systems, Journal of Operations Management,* and other journals. He serves, or has served, as a senior editor at *Information Systems Research, MIS Quarterly*, and *Journal of Strategic Information Systems* and as an associate editor at *Information Systems Research*, *Management Science, Journal of MIS,* and *MIS Quarterly*. He was named Fellow of the Association for Information Systems in 2010 in recognition for outstanding contributions to the Information Systems discipline.

Christian M. Ringle is Professor of Management at the Hamburg University of Technology (TUHH), Germany, and Visiting Professor at the University of Newcastle, Australia. His research concerns improvements of quantitative methods for business research applied to study management and marketing issues. His work has been published in outlets that include *MIS Quarterly*, *International Journal of Research in Marketing*, *Journal of the Academy of Marketing Science*, *Journal of Service Research*, *Journal of Business Research*, and *Long Range Planning*.

Franziska Völckner is Professor of Marketing at the University of Cologne, Germany. Her research interest focuses on building and managing market-based assets. This interest bridges the areas of branding, consumer behavior, marketing metrics, and marketing strategy. Her work has been published in several academic journals, including *Journal of Marketing, Journal of Marketing Research, International Journal of Research in Marketing, Journal of the Academy of Marketing Science, Journal of Service Research, Journal of Business Research,* and *Marketing Letters*, among others.

MIS Quarterly

RESEARCH ESSAY

DISCOVERING UNOBSERVED HETEROGENEITY IN STRUCTURAL EQUATION MODELS TO AVERT VALIDITY THREATS

Jan-Michael Becker
Department of Marketing and Brand Management, University of Cologne,
Cologne, 50923, GERMANY {j.becker@wiso.uni-koeln.de}

Arun Rai
Center for Process Innovation and Department of Computer Information Systems, Robinson College of Business,
Georgia State University, Atlanta, GA 30303 U.S.A. {arunrai@gsu.edu}

Christian M. Ringle
Institute for Human Resource Management and Organizations, Hamburg University of Technology (TUHH),
Hamburg, 21073, GERMANY {ringle@tuhh.de} and
Faculty of Business and Law, University of Newcastle, Callaghan, NSW 2308 AUSTRALIA {christian.ringle@newcastle.edu.au}

Franziska Völckner
Department of Marketing and Brand Management, University of Cologne,
Cologne, 50923, GERMANY {voelckner@wiso.uni-koeln.de}

Appendix A

Meta-Analyses of Information Systems Studies

Table A1. Meta-Analyses of IS Studies: Inconsistent Results Across a Range of Phenomena

IS Phenomenon	Reference, Journal	Scope	Meta-Analysis Purpose	Moderators/Contingency Variables Examined	Nature of Inconsistent Findings *(emphasis added)*
Decision Support System (DSS) Implementation Success	Alavi and Joachimsthaler 1992, MISQ	144 findings from 33 studies	Investigating the relationship between user-related factors and DSS implementation success	*Authors suggest that moderators could explain the large variance in effect sizes across studies.*	"*Reviews of information systems implementation research…have revealed that collectively, implementation studies have yielded* ***conflicting and somewhat confusing findings***."
Group Support Systems (GSS)	Dennis et al. 2001, MISQ	61 articles	Developing a new model for interpreting GSS effects on firm performance.	• Fit between the Task and the GSS Structures • Appropriation Support Received	"*Many previous papers have lamented the fact that the findings of past GSS* ***research have been inconsistent****. This paper develops a new model for interpreting GSS effects on performance…*"

Table A1. Meta-Analyses of IS Studies: Inconsistent Results Across a Range of Phenomena (Continued)

IS Phenomenon	Reference, Journal	Scope	Meta-Analysis Purpose	Moderators/Contingency Variables Examined	Nature of Inconsistent Findings *(emphasis added)*
IT Investment Payoff	Kohli and Deveraj 2003, ISR	66 studies	Examining structural variables that explain why some IT payoff studies observe a positive effect and some do not.	• Dependent Classification • Sample Size • Data Source • Type of IT Impact • Type of IT Assets • Industry	"*…some studies have shown* ***mixed results*** *in establishing a relationship between IT investment and firm performance.*"
IT Innovation Adoption	Lee and Xia 2006, I&M	54 correla-tions from 21 studies	Investigating the effects of organizational size on IT innovation adoption.	• Type of Innovation • Type of Organization • Stage of Adoption • Scope of Size • Industry Sector	"*…empirical results on the relationship between them have been* ***disturbingly mixed and inconsisten****t.…explain and resolve these mixed results by… examining the effects of six moderators on the relationship.*"
IT Project Escalation	Wang and Keil 2007, IRMJ	12 articles with 20 separate experiment s	Investigating the effect size of sunk cost on project escalation and deter-mining whether there is a difference in effect sizes between IT and non-IT projects.	• IT vs. Non-IT Projects	"*…because of the strong magnitude and* ***heterogeneity of effect sizes*** *for the sunk cost effect, we need more primary studies that investigate potential moderators of sunk cost.*"
Turnover of IT Professionals	Joseph et al. 2007, MISQ	33 studies	Integrating the 43 antecedents of turnover intentions of IT professionals in a unified framework using meta-analytic structural equation modeling.	• Age • Gender Ratio of Sample • Operationalization of Turnover Intention • Operationalization of Antecedents	"*…our narrative review finds* ***several inconsistent*** *(e.g., organization tenure and role conflict)* ***and inconclusive*** *(e.g., age and gender) findings.*"
IS Implementation Success	Sharma and Yetton 2003, MISQ	22 studies	Proposing a contingent model in which task interdependence moderates the effect of management support on implementation success.	• Task Interdependence	"*A meta-analysis of the empirical literature provides strong support for the model and begins to explain the* ***wide variance in empirical findings****.*" "*The theory developed and findings reported above help to explain the* ***inconsistent findings*** *in the literature.*"
	Sabherwal et al. 2006, Mgmt.Scien ce	612 findings from 121 studies	Explaining the interrelationships among four constructs representing the success of a specific information system and the relationships of these IS success constructs with four user-related constructs and two constructs representing the context.	*Authors suggest that possible moderators include voluntari-ness of IS adoption and user characteristics such as age and gender.*	"*Despite considerable empirical research, results on the relationships among constructs related to information system (IS) success, as well as the determinants of IS success,* ***are often inconsistent****.*"
	Sharma and Yetton 2007, MISQ	27 studies	Proposing a contingent model in which the effect of training on IS implementation success is a function of technical complexity and task interdependence.	• Technical Complexity • Task Interdependence	"*Research has investigated the main effect of training on information systems implementation success. However,* ***empirical support for this model is inconsistent****.*"

Table A1. Meta-Analyses of IS Studies: Inconsistent Results Across a Range of Phenomena (Continued)

IS Phenomenon	Reference, Journal	Scope	Meta-Analysis Purpose	Moderators/Contingency Variables Examined	Nature of Inconsistent Findings *(emphasis added)*
Technology Acceptance	King and He 2006, I&M	88 studies	Summarizing TAM research and investigating conditions under which TAM may have different effects.	• Type of Users • Type of Usage	"*...all TAM relationships are not borne out in all studies; there is* ***wide variation in the predicted effects*** *in various studies…*" "*Since there* ***are inconsistencies in TAM results****, a meta-analysis is more likely to appropriately integrate the positive and the negative.*"
	Schepers and Wetzels 2007, I&M	51 articles containing 63 studies	Analyzing the role of subjective norms and three inter-study moderating factors.	• Type of Respondents • Type of Technology • Culture	"*First, the subjective norm has had a* ***mixed and inconclusive role****….Some studies found considerable impacts of it on the dependent variables. However, others did not find significant effects.*"
	Wu and Lederer 2009, MISQ	71 studies	Investigating the impact of environment-based voluntariness on the relationships among the four primary TAM constructs (i.e., ease of use, perceived usefulness, behavioral intention, and usage).	• Environment-Based Voluntariness	"*The Q statistic for each of the five correlations exceeded its cutoff, and thus the analyses confirmed heterogeneity for each (p < 0.01).* ***That is, of all the correlations vary across studies more than would be produced by sampling error.***"

Appendix B

Prediction-Oriented Segmentation for PLS Path Modeling (PLS-POS)

Overview

As a distance-based segmentation method, the PLS prediction-oriented segmentation (PLS-POS) method builds on earlier work on distance-measure-based segmentation—that is, the PLS typological path modeling (PLS-TPM) approach (Squillacciotti 2005) and its enhancement, the response-based detection of respondent segments in PLS (REBUS-PLS) (Esposito Vinzi et al. 2008). To extend the distance-measure-based PLS segmentation methods (including overcoming the methodological limitation of PLS-TPM and REBUS-PLS being applicable only to PLS path models with reflective measures (Esposito Vinzi et al. 2008; Sarstedt 2008)), the PLS-POS algorithm introduces three novel features: (1) it uses an explicit PLS-specific *objective criterion* to form homogeneous groups, (2) it includes a new *distance measure* that is appropriate for PLS path model with both reflective and formative measures and is able to uncover unobserved heterogeneity in formative measures, and (3) it ensures continuous *improvement of the objective criterion* throughout the iterations of the algorithm (hill-climbing approach). Table B1 shows the key technical differences of the new PLS-POS method in comparison with the main distance-based methods (i.e., PLS-TPM and REBUS-PLS) and the popular finite-mixture method for PLS (i.e., FIMIX-PLS).

The following sections explain in greater detail PLS-POS' distinctive features. To begin with, we focus on the description of PLS-POS' *objective criterion*. An explanation of the *distance measure* employed and its extension to use it for formative measurement models follows. Finally, we provide details on the *algorithm* with its specific steps and procedures and how it ensures the continuous improvement of the objective criterion.

Objective Criterion of PLS-POS

The main segmentation objective in PLS is to form homogenous groups of observations that show increased endogenous variables' explained variance (R^2) and, thus, provide an improved prediction (compared to the overall sample), which is in accordance with Anderberg's (1973, p.

Table B1. Comparison of the Technical Differences of FIMIX-PLS, PLS-TPM, REBUS-PLS, and PLS-POS

	Finite-Mixture Segmentation Approach	**Distance-Based Clustering Approaches**		
Algorithm Feature	**FIMIX-PLS** (Hahn et al. 2002)	**PLS-TPM** (Squillacciotti 2005; Squillacciotti 2010)	**REBUS-PLS** (Esposito Vinzi et al. 2010; Esposito Vinzi et al. 2008)	**PLS-POS**
Distributional Assumptions	Yes	No	No	No
Pre-clustering	No pre-clustering; random split of observations	Hierarchical classification based on redundancy residuals of the overall model	Hierarchical classification based on communality and structural residuals of the overall model	No pre-clustering; random split of observations and assignment to closest segment according to the distance measure
Distance measure	Has no distance measure†	Based on redundancy residuals of a single reflective endogenous latent variable	Based on communality residuals of all latent variables and structural residuals of all endogenous latent variables	Based on structural residuals of all endogenous latent variables with an extension that also accounts for heterogeneity in formative measures
Accounts for sources of heterogeneity in reflective measures?	No	No	Yes	No
Accounts for sources of heterogeneity in formative measures?	No	No‡	No ‡	Yes
Accounts for sources of heterogeneity in the structural model?	Yes	Yes	Yes	Yes
Assignment of observations to segments in each iteration	Proportional assignment of **all** observations to all segments based on the conditional multivariate normal densities to optimize the likelihood function	Assigns **all** observations to the closest segment	Assigns **all** observations to the closest segment	Assigns **only one** observation to the closest segment and assures improvement of an objective criterion (*R^2* of all endogenous latent variables) before accepting the change
Stop criterion	Extremely small improvement in log likelihood below critical value (or maximum number of iterations)	Stability of the classes' composition (no reassignment of observations); or maximum number of iterations	Stability of the classes' composition (number of reassignments below a critical percentage value of observations); or maximum number of iterations	Infinitesimal improvement in objective criterion (or maximum number of iterations)

†FIMIX-PLS assumes that each endogenous latent variable is distributed as a finite mixture of conditional multivariate normal densities. It uses these densities to estimate probabilities of segment memberships for each observation (proportional assignment) to optimize the likelihood function (which implicitly maximizes the segment-specific explained variance as part of the likelihood function).
‡"As in PLS-TPM, … [REBUS-PLS] 'distance' has, so far, only been implemented on models with reflective blocks. Although this is not to be considered a strict limitation for many applications, it must be pointed out that REBUS-PLS requires all blocks to be reflective" (Esposito Vinzi et al. 2008, p. 444). This requirement for models with only reflective measures also holds for the REBUS-PLS implementation in the PLSPM package (Sánchez and Trinchera 2013) for the statistical software R (R Core Team 2013).

195) notion of "clustering for maximum prediction." Consequently, possible PLS-specific and, thus, prediction-oriented objective criteria include the following: (1) the sum of the manifest variables' redundancy residuals in the reflective measures, (2) the sum of endogenous latent variables' R^2 values in the structural model, and (3) the goodness-of-fit criterion (GoF; Tenenhaus et al. 2005)[1] for assessing both the structural model and the reflective measures.

Including the residual terms of the manifest variables would only be appropriate to assess the explained variance and, thus, the predictive performance in reflective measures. Because PLS path modeling allows for the use of reflective and formative measures, objective criteria that draw on the manifest variables' residual terms do not support the general applicability of PLS-POS in both measurement models (i.e., reflective and formative). Consequently, the redundancy and community residual in the reflective measures, which are also included in the PLS-GoF measure, are not a useful criterion for the purpose of the PLS segmentation method.

An appropriate PLS-specific objective criterion maximizes the sum of the endogenous latent variables' R^2 values. In accordance with the PLS algorithm's objective (Lohmöller 1989; Wold 1982), PLS-POS focuses on maximizing the predictivity of each group by minimizing the sum of the endogenous latent variables' squared residuals in the PLS path model. Thus, the sum of each group's sum of R^2 values represents the objective criterion, which is explicitly defined and calculated in the PLS-POS algorithm. Every reassignment of observations in PLS-POS ensures improvement of the objective criterion (hill climbing approach; see description of the algorithm below). This objective criterion is suitable for any PLS path model regardless of whether such models include reflective or formative measures.

Distance Measure

To reassign observations, PLS-POS builds on the idea of Squillacciotti (2005) and Esposito Vinzi et al. (2008) to use a distance measure. We propose a new distance measure that is applicable to both reflective and formative measures and accounts for heterogeneity in the structural and the formative measurement model. This observation-to-group distance measure identifies appropriate observations to form homogenous groups and thereby depicts suitable candidates to improve the objective criterion. Within a group, each observation's capability to predict the endogenous latent variables in the PLS path model determines its distance to that group: the shorter the distance of observation i to group g, the higher the predictivity of observation i in group g.

It is important to understand the conceptual difference between observation i's membership in its current group k ($k = g$; $k, g \varepsilon G$) and its distance to an alternative group g ($k \neq g$; $k, g \varepsilon G$). For every endogenous latent variable b ($b \varepsilon B$), the latent variable scores of its direct predecessors $Y_{a_b ik}^{exogenous}$ and the corresponding structural model path coefficients $p_{a_b g}$ allow for the group-specific prediction of the endogenous latent variable scores $\left(\hat{Y}_{big}\right)$ via linear combinations $\left(\hat{Y}_{big} = \sum_{a_b=1}^{A_b} Y_{a_b ik}^{exogenous} \times p_{a_b g}\right)$. To calculate $\hat{Y}_{big}$, we use the latent variable scores of an observation's current group k and draw on the alternative group g's PLS path coefficients $p_{a_b g}$. The difference between the predicted value $\hat{Y}_{big}$ and the current group's latent variable scores Y_{bik} from the PLS path model estimation is the residual of observation i in group g for the endogenous latent variable b (Equation 1):

$$e_{big}^2 = \left(\hat{Y}_{big} - Y_{bik}\right)^2 = \left(\sum_{a_b=1}^{A_b} Y_{a_b ok}^{exogenous} \times p_{a_b g} - Y_{bik}^{endogenous}\right)^2 \tag{1}$$

The result of e_{big}^2 is an observation's predictivity in its current group when $k = g$ ($k, g \varepsilon G$). Furthermore, using the path coefficients $p_{a_b g}$ of alternative group-specific PLS estimations for $k \neq g$ ($k, g \varepsilon G$) provides a heuristic outcome for observation i's predictivity in each of the G-1 other possible group assignments. This establishes the new prediction-oriented PLS-POS distance measure, as presented by Equation (2):

$$D_{kig} = \sum_{b=1}^{B} \sqrt{\frac{e_{big}^2}{\sum_{i=1}^{I_k} e_{big}^2}} \tag{2}$$

The residuals of each observation i are divided by the sum of the residuals of all observations in i's current group k (I_k; sample size in group k). This ratio's square root is the distance of an observation i to group g for an endogenous latent variable b ($b \varepsilon B$). The sum over all

[1]Against its naming, PLS-GoF does not represent a measure of fit for PLS path modeling; see Henseler and Sarstedt (2012) for a discussion.

endogenous variables B in the PLS path model provides the total distance measure D_{kig}. The smaller the sum of the endogenous latent variables' squared residual values, the higher the predictivity of observation i in group g of the underlying PLS path model.

The distinction between formative and reflective measures requires that one pays particular attention in PLS path modeling (e.g., Diamantopoulos et al. 2001; Gudergan et al. 2008; Jarvis et al. 2003). Formative measures require (1) taking into account the indicators' heterogeneity for each measurement model within each group and/or (2) uncovering the significant differences in weights between the groups. Therefore, calculating the group-specific residual term in models with formative measures requires an extension of the group-specific residual e^2_{big} in the distance measure. The latent variable scores $Y_{a_b jik}$ are replaced by linear combinations of the manifest variable scores $x_{a_b jik}$ and the corresponding measurement model's formative weights $\pi_{a_b jg}$. Equation (3) shows the calculation of the residual term for formative measures in the PLS path model.

$$e^2_{big} = \left(\sum_{a_b=1}^{A_b} \sum_{j}^{J} x_{a_b jik} \times \pi_{a_b jg} \times p_{a_b g} - Y_{bik}^{endogenous} \right)^2 \tag{3}$$

The formative latent variable scores become a group-wise reestimated prediction of the associated manifest variables j when the squared residual is determined.

Algorithm

The segmentation process starts by randomly partitioning the overall sample into the prespecified number of G equal groups (Figure B1, Step 1). Calculating all group-specific PLS path model estimates reveals each observation's distance to its own and all other G-1 groups. A partitioning approach that assigns each observation to the group to which it has the shortest distance improves the initial segmentation.

Subsequently, the PLS-POS algorithm computes the group-specific PLS path modeling results (Figure B1, Step 2), updates the objective function (Figure B1, Step 3), and computes the observations' distances to all groups (Figure B1, Step 4.1). PLS-POS uses the distance measure to reassign observations based on the maximum value of the difference between an observation's distance to its current group (i.e., the group to which the observation has been assigned) and its distance to an alternative group (Equation 4).

$$\text{difference } \Delta_{kig} = \text{distance to current group } k\ (D_{kik}) - \text{distance to alternative group } g\ (D_{kig}) \tag{4}$$

Positive differences indicate that an observation has a shorter distance to the alternative group and, thus, potentially fits better in that group in terms of predictivity. This computation is conducted for all observations (Figure B1, Step 4.1). Each observation's maximum positive difference becomes part of the list of candidates (Figure B1, Step 4.2). Negative values are not considered because reassigning these observations possibly decreases the objective criterion. Subsequently, the candidates are sorted in descending order in terms of their positive distance differences (Figure B1, Step 4.3).

After the STOP statement, PLS-POS provides the group-specific PLS path model estimates for the final segmentation solution (Figure B1, Step 7). The maximum number of iterations should be sufficiently high (e.g., twice the number of observations in the overall sample) to obtain a solution that is close to the global optimum. The maximum search depth equals the number of observations in the sorted list of candidate observations for reassignment and, thus, may not exceed the number of observations in the overall sample. In early explorative research stages, one may use a reduced search depth for performance reasons. However, to determine the final segmentation result, the search depth should equal the maximum number of observations to ensure that the segmentation solution that minimizes the PLS-POS objective criterion (i.e., the endogenous latent variables' R^2 values in the PLS path model) has been identified.

Finally, three important issues are worth noting. First, PLS-POS only reassigns observations that improve the objective criterion. As such, the algorithm ensures the continuous improvement of the objective criterion and potentially provides a solution that is at least close to the global optimum. Second, in each iteration step, the algorithm changes the assignment of only one observation and calculates the group-specific PLS estimates of all observations and their new distance measures. Thus, in contrast to the alternative distance-based PLS segmentation approaches suggested in the literature to date (e.g., Esposito Vinzi et al. 2008; Squillacciotti 2005), PLS-POS avoids moving a sizeable set (more or less) of similar candidates from one group to another without improving the objective criterion. Third, owing to the implementation of a hill-climbing approach, PLS-POS could face the problem of ending in local optima. Wedel and Kamakura (2000) recommend running hill-climbing algorithms several times to attain alternative starting partitions and, finally, to select the best segmentation solution. The same procedure should be applied in the application of PLS-POS.

Step 1: Create an initial segmentation to start the algorithm
    Step 1.1: Randomly split the overall sample into K equally sized groups
    Step 1.2: Compute the group-specific PLS estimates for the path model
    Step 1.3: Establish each observation's distance to each group
    Step 1.4: Assign each observation to the closest group
DO LOOP
Step 2: Compute the group-specific PLS estimates for the path model
Step 3: Determine the result of the objective criterion
Step 4: Create a list of candidate observations for reassignment
    Step 4.1: Establish the K-1 differences between each observation's distance to its current group and an alternative group
    Step 4.2: IF an observation has one or more positive differences of distances, then
        Add the maximum difference and the observation's corresponding alternative group assignment to a list of candidates
        ELSE: Do nothing
    Step 4.3: IF the list is empty, then
        GO TO STOP
        ELSE: Sort the list of candidate observations in descending order in terms of their positive distance differences
Step 5: Improve the segmentation result
    Step 5.1: Select the first observation in the list of candidate observations for reassignment
    DO LOOP
    Step 5.2: Reassign the observation
    Step 5.2: Compute the group-specific PLS estimates for the path model
    Step 5.3: Determine the result of the objective criterion
    Step 5.4: IF the observation's reassignment improves the objective criterion, then
        Save the current assignment and GO TO Step 6
        ELSE: Undo changes and continue with Step 5.5
    Step 5.5: IF the list contains a subsequent observation following the currently selected observation on the list of candidates AND the maximum search depth has not been reached, then
        Select the next observation
        ELSE: GO TO Step 6
    UNTIL the objective criterion is improved
Step 6: IF the maximum number of iterations OR the maximum search depth has been reached, then
    GO TO STOP
    ELSE: GO TO Step 2
UNTIL STOP
Step 7: Compute the group-specific PLS path model estimates and provide the final segmentation results

Figure B1. The PLS-POS Algorithm

Appendix C

Design of the Multicollinearity Factor for the Simulation Study

The design of the simulation study for the formative measurement model includes three levels of multicollinearity between the formative indicators in the model. To simulate different levels of multicollinearity, we revert to Mason and Perreault's (1991) seminal study on multicollinearity (see also Grewal et al. 2004). We vary two levels of correlation patterns among the predictor variables reflecting conditions typically encountered by researchers and practitioners. In addition, a situation in which the indicators are uncorrelated (orthogonal) serves as a baseline for comparison (i.e., a perfect formative measure) because this model is unaffected by multicollinearity.

Table C1 shows the two multicollinearity levels based on Mason and Perreault, including the trace of $(X'X)^{-1}$, det(X'X), and condition number, as well as each variable's variance inflation factor (VIF) associated with a given level of multicollinearity.

Table C1. Levels of Multicollinearity

		Level 1				Level 2			
		X_1	X_2	X_3	X_4	X_1	X_2	X_3	X_4
	X_1	1.00				1.00			
	X_2	.65	1.00			.80	1.00		
	X_3	.40	.40	1.00		.60	.60	1.00	
	X_4	.00	.00	.00	1.00	.00	.00	.00	1.00
VIF		1.80	1.80	1.24	1.00	2.96	2.96	1.67	1.00
Trace $(X'X)^{-1}$		5.85				8.59			
Det(X'X)		.47				.22			
Condition no.		2.38				3.42			

Note: VIF = variance inflation factor

Appendix D

Simulation on the Effects of Unobserved Heterogeneity

The objective of this simulation study is to evaluate the implications of unobserved heterogeneity for structural model parameter estimates in PLS path models. The results show that unobserved heterogeneity has a strong adverse effect on PLS estimation outcomes: (1) parameter estimates are biased, (2) nonsignificant path coefficients at the group level become significant at the overall sample level that combines groups, (3) sign differences in the parameter estimates between groups are manifested as nonsignificant results at the overall sample level, and (4) explained variance of the model (R^2 of the endogenous variables) decreases. These erroneous estimates can lead to both Type I and Type II errors and to invalid inferences.

The simulation study uses a path model with two exogenous variables having a direct effect on one endogenous variable (all variables measured with five reflective indicators). We generate data for the true path coefficients of two groups by considering three situations of unobserved heterogeneity:

- Situation 1, where the path coefficients between group 1 and group 2 differ but show the same sign. We consider scenarios where all parameter estimates are positive (situation 1a) and negative (situation 1b) and where the magnitude in parameter differences between groups is low (.1) and high (.5).

- Situation 2, where unobserved heterogeneity causes sign reversal in parameter estimates across the two groups (i.e., group 1 has a positive path coefficient, while group 2 has a negative one).

- Situation 3, where one group has a nonsignificant parameter estimate and the other group has a significant parameter estimate. We distinguish between two different levels of parameter differences represented by the effect size of the significant parameter, namely .2 and .7.

We generated 100 sets of data for each condition and estimated the group-specific path coefficients, the overall sample path coefficients, and the t-values of these coefficients by employing the bootstrapping procedure on 1,000 subsamples (Henseler et al. 2009).

Table D1 presents the results. The left side shows the group-specific mean estimates of the path coefficients and their average t-values.[2] The columns on the right side show the mean path coefficients of the overall sample and the interpretation of the results in terms of bias, Type I and II errors, and variance explained (R^2).

[2]For a significance level of $\alpha = 0.05$ the t-value has to exceed the threshold of 1.98 in these conditions.

The results show that in all situations, biases in the parameter estimates distort effect sizes and cause misinterpretation of the path coefficients, which is especially problematic for comparative hypotheses (e.g., path coefficient 1 > path coefficient 2). Type I and Type II errors are exacerbated in situations where the group-specific parameters show inconsistent signs (i.e., situation 2 where signs are reversed across groups) and when at least one of the groups involves nonsignificant parameters while the other group does not (i.e., situation 3). In contrast, when all parameters are significant and show the same sign (situation 1), our results suggest that it is not very likely that Type II errors occur. In this situation, the existence of Type II errors depends on the effect size and the degree to which the increased power of the combined sample size compensates for the increase in standard errors due to unobserved heterogeneity. For all parameter constellations in our simulation study, the increased sample size compensates for the increase in standard errors.

The R^2 decreases in almost all situations, implying an inferior model fit at the overall sample level. We find particularly strong decreases in R^2 in situations in which the group-specific effect sizes are high; in contrast, R^2 is almost unaffected in situations showing low group-specific effect sizes.

Table D1. Results of the Simulation Study

	Group-Specific Parameter Estimates		Pooled Parameter Estimate				
	Group 1 (n = 200)	**Group 2 (n = 200)**	**Parameter (n = 400)**	**Biased?**	**Type I Error**	**Type II Error**	**Lower R^2**
1a.	**.7** *(t = 18.57)* **.2** *(t = 3.94)* *R^2 = .53*	**.2** *(t = 3.84)* **.7** *(t = 19.64)* *R^2 = .53*	**.45** *(t = 11.36)* **.45** *(t = 11.54)* *R^2 = .41*	Yes	–	No	Yes
	.3 *(t = 4.95)* **.2** *(t = 3.31)* *R^2 = .13*	**.2** *(t = 3.36)* **.3** *(t = 4.79)* *R^2 = .13*	**.25** *(t = 5.70)* **.25** *(t = 5.73)* *R^2 = .12*	Yes	–	No	(Yes)
1b.	**-.7** *(t = 18.95)* **-.2** *(t = 3.70)* *R^2 = .53*	**-.2** *(t = 4.01)* **-.7** *(t = 19.27)* *R^2 = .53*	**-.45** *(t = -11.19)* **-.45** *(t = -11.44)* *R^2 = .24*	Yes	–	No	Yes
	-.3 *(t = 5.03)* **-.2** *(t = 3.14)* *R^2 = .13*	**-.2** *(t = 3.25)* **-.3** *(t = 5.09)* *R^2 = .13*	**-.25** *(t = -5.61)* **-.25** *(t = -5.80)* *R^2 = .12*	Yes	–	No	(Yes)
2.	**.7** *(t = 19.43)* **.2** *(t = 3.99)* *R^2 = .53*	**-.7** *(t = 19.09)* **-.2** *(t = 3.78)* *R^2 = .53*	**.00** *(t = .01)* **.00** *(t = .00)* *R^2 = .00*	Yes	–	100% 100%	Yes
3.	**.7** *(t = 19.94)* **.0** *(t = .01)* *R^2 = .49*	**.0** *(t = .01)* **.7** *(t = 19.89)* *R^2 = .49*	**.35** *(t = 7.61)* **.35** *(t = 7.38)* *R^2 = .24*	Yes	100% 100%	No	Yes
	.2 *(t = 3.38)* **.0** *(t = .00)* *R^2 = .04*	**.0** *(t = .01)* **.2** *(t = 3.17)* *R^2 = .04*	**.10** *(t = 1.88)* **.10** *(t = 1.90)* *R^2 = .02*	Yes	20% 40%	80% 60%	(Yes)
4.	**.0** *(t = .00)* **.0** *(t = .01)* *R^2 = .00*	**.0** *(t = .01)* **.0** *(t = .00)* *R^2 = .00*	**.00** *(t = .00)* **.00** *(t = .00)* *R^2 = .00*	–	No	–	–

Appendix E

ANOVA Results—Model 1 (Reflective Measures)

Tables E1 to E4 present the ANOVA results for model 1 (reflective measures) explaining MAB by method (PLS-POS/FIMIX-PLS) and the six design factors. All significant and substantial effects (i.e., all effects that explain more than 2 percent of the total variance in MAB implying a partial η^2 of more than .02) are highlighted in grey.

We find that the R^2, structural model heterogeneity, data distribution, and the interaction of structural model heterogeneity and R^2 have a substantial and significant effect on the MAB of both methods. Furthermore, there is a significant and substantial difference in the parameter recovery (MAB) of the two methods (PLS-POS and FIMIX-PLS) and for the interaction effects between the method and structural model heterogeneity and between the method and R^2.

Table E1. Between-Subjects Effects (Part I)

Source of Variance in MAB	df	F	Sig.	Partial η^2
Intercept	1	14,658.62	.000	.568
SMH	3	1,121.71	.000	.232
R^2	3	1,948.85	.000	.344
Sample Size	2	70.77	.000	.013
Reliability	1	1.88	.170	.000
Data Distribution	1	497.52	.000	.043
RSS	1	22.62	.000	.002
SMH × R^2	9	178.96	.000	.126
SMH × Sample Size	6	9.64	.000	.005
SMH × Reliability	3	1.33	.262	.000
SMH × Data Distribution	3	21.15	.000	.006
SMH × RSS	3	25.17	.000	.007
R^2 × Sample Size	6	11.44	.000	.006
R^2 × Reliability	3	.75	.524	.000
R^2 × Data Distribution	3	14.72	.000	.004
R^2 × RSS	3	29.76	.000	.008
Sample Size × Reliability	2	.48	.620	.000
Sample Size × Data Distribution	2	14.17	.000	.003
Sample Size × RSS	2	63.92	.000	.011
Reliability × Data Distribution	1	4.04	.044	.000
Reliability × RSS	1	.11	.735	.000
Data Distribution × RSS	1	267.72	.000	.023
SMH × R^2 × Sample Size	18	1.75	.026	.003
SMH × R^2 × Reliability	9	1.27	.249	.001
SMH × R^2 × Data Distribution	9	6.00	.000	.005
SMH × R^2 × RSS	9	2.32	.013	.002
SMH × Sample Size × Reliability	6	1.39	.216	.001

Note: df = degrees of freedom; MAB = mean absolute bias; RSS = relative segment size; SMH = structural model heterogeneity; all significant and substantial effects (i.e., all effects that explain more than 2% of the total variance in MAB implying a partial η^2 of more than .02) are highlighted in grey.

Table E2. Between-Subjects Effects (Part II)				
Source of Variance in MAB	**df**	**F**	**Sig.**	**Partial η^2**
SMH × Sample Size × Data Distribution	6	5.22	.000	.003
SMH × Sample Size × RSS	6	9.23	.000	.005
SMH × Reliability × Data Distribution	3	2.19	.087	.001
SMH × Reliability × RSS	3	3.50	.015	.001
SMH × Data Distribution × RSS	3	2.30	.075	.001
R^2 × Sample Size × Reliability	6	1.88	.080	.001
R^2 × Sample Size × Data Distribution	6	1.83	.089	.001
R^2 × Sample Size × RSS	6	13.00	.000	.007
R^2 × Reliability × Data Distribution	3	1.85	.135	.000
R^2 × Reliability × RSS	3	.42	.740	.000
R^2 × Data Distribution × RSS	3	7.83	.000	.002
Sample Size × Reliability × Data Distribution	2	1.65	.191	.000
Sample Size × Reliability × RSS	2	2.19	.112	.000
Sample Size × Data Distribution × RSS	2	17.14	.000	.003
Reliability × Data Distribution × RSS	1	1.08	.299	.000
SMH × R^2 × Sample Size × Reliability	18	.53	.948	.001
SMH × R^2 × Sample Size × Data Distribution	18	1.68	.036	.003
SMH × R^2 × Sample Size × RSS	18	2.11	.004	.003
SMH × R^2 × Reliability × Data Distribution	9	.68	.725	.001
SMH × R^2 × Reliability × RSS	9	.80	.614	.001
SMH × R^2 × Data Distribution × RSS	9	1.52	.135	.001
SMH × Sample Size × Reliability × Data Distribution	6	.60	.730	.000
SMH × Sample Size × Reliability × RSS	6	.79	.577	.000
SMH × Sample Size × Data Distribution × RSS	6	2.41	.025	.001
SMH × Reliability × Data Distribution × RSS	3	2.06	.104	.001
R^2 × Sample Size × Reliability × Data Distribution	6	1.52	.168	.001
R^2 × Sample Size × Reliability × RSS	6	1.04	.399	.001
R^2 × Sample Size × Data Distribution × RSS	6	4.75	.000	.003
R^2 × Reliability × Data Distribution × RSS	3	.26	.851	.000
Sample Size × Reliability × Data Distribution × RSS	2	.53	.588	.000
SMH × R^2 × Sample Size × Reliability × Data Distribution	18	.70	.817	.001
SMH × R^2 × Sample Size × Reliability × RSS	18	.70	.811	.001
SMH × R^2 × Sample Size × Data Distribution × RSS	18	.99	.473	.002
SMH × R^2 × Reliability × Data Distribution × RSS	9	.50	.874	.000
SMH × Sample Size × Reliability × Data Distribution × RSS	6	1.71	.115	.001
R^2 × Sample Size × Reliability × Data Distribution × RSS	6	1.41	.206	.001
SMH × R^2 × Sample Size × Reliability × Data Distribution × RSS	18	.96	.502	.002
Error	11,136			

Note: df = degrees of freedom; MAB = mean absolute bias; RSS = relative segment size; SMH = structural model heterogeneity.

Table E3. Within-Subjects Effects (Part I)				
Source of Variance in MAB	df	F	Sig.	Partial η^2
Method	1	952.31	.000	.079
Method × SMH	3	217.47	.000	.055
Method × R^2	3	137.14	.000	.036
Method × Sample Size	2	4.66	.009	.001
Method × Reliability	1	.00	.974	.000
Method × Data Distribution	1	87.97	.000	.008
Method × RSS	1	104.01	.000	.009
Method × SMH × R^2	9	12.84	.000	.010
Method × SMH × Sample Size	6	2.79	.010	.002
Method × SMH × Reliability	3	.26	.854	.000
Method × SMH × Data Distribution	3	37.26	.000	.010
Method × SMH × RSS	3	.88	.450	.000
Method × R^2 × Sample Size	6	1.84	.087	.001
Method × R^2 × Reliability	3	.02	.995	.000
Method × R^2 × Data Distribution	3	19.48	.000	.005
Method × R^2 × RSS	3	3.98	.008	.001
Method × Sample Size × Reliability	2	.27	.765	.000
Method × Sample Size × Data Distribution	2	17.60	.000	.003
Method × Sample Size × RSS	2	16.60	.000	.003
Method × Reliability × Data Distribution	1	.02	.876	.000
Method × Reliability × RSS	1	.149	.700	.000
Method × Data Distribution × RSS	1	14.37	.000	.001
Method × SMH × R^2 × Sample Size	18	.89	.589	.001
Method × SMH × R^2 × Reliability	9	1.33	.215	.001
Method × SMH × R^2 × Data Distribution	9	2.07	.029	.002
Method × SMH × R^2 × RSS	9	4.56	.000	.004
Method × SMH × Sample Size × Reliability	6	.73	.626	.000
Method × SMH × Sample Size × Data Distribution	6	3.94	.001	.002
Method × SMH × Sample Size × RSS	6	1.72	.112	.001
Method × SMH × Reliability × Data Distribution	3	.74	.527	.000
Method × SMH × Reliability × RSS	3	1.02	.381	.000
Method × SMH × Data Distribution × RSS	3	18.88	.000	.005
Method × R^2 × Sample Size × Reliability	6	.28	.945	.000
Method × R^2 × Sample Size × Data Distribution	6	2.09	.051	.001
Method × R^2 × Sample Size × RSS	6	3.57	.002	.002
Method × R^2 × Reliability × Data Distribution	3	.29	.835	.000
Method × R^2 × Reliability × RSS	3	1.28	.278	.000
Method × R^2 × Data Distribution × RSS	3	8.97	.000	.002
Method × Sample Size × Reliability × Data Distribution	2	.69	.501	.000
Method × Sample Size × Reliability × RSS	2	.13	.876	.000
Method × Sample Size × Data Distribution × RSS	2	8.98	.000	.002
Method × Reliability × Data Distribution × RSS	1	.00	.993	.000

Note: df = degrees of freedom; MAB = mean absolute bias; RSS = relative segment size; SMH = structural model heterogeneity; all significant and substantial effects (i.e., all effects that explain more than 2% of the total variance in MAB implying a partial η^2 of more than .02) are highlighted in grey.

Table E4. Within-Subjects Effect (Part II)				
Source of Variance in MAB	**df**	**F**	**Sig.**	**Partial η^2**
Method × SMH × R^2 × Sample Size × Reliability	18	.56	.930	.001
Method × SMH × R^2 × Sample Size × Data Distribution	18	1.95	.009	.003
Method × SMH × R^2 × Sample Size × RSS	18	1.47	.092	.002
Method × SMH × R^2 × Reliability × Data Distribution	9	.95	.484	.001
Method × SMH × R^2 × Reliability × RSS	9	1.07	.380	.001
Method × SMH × R^2 × Data Distribution × RSS	9	1.96	.040	.002
Method × SMH × Sample Size × Reliability × Data Distribution	6	.54	.775	.000
Method × SMH × Sample Size × Reliability × RSS	6	1.23	.286	.001
Method × SMH × Sample Size × Data Distribution × RSS	6	2.62	.015	.001
Method × SMH × Reliability × Data Distribution × RSS	3	.30	.828	.000
Method × R^2 × Sample Size × Reliability × Data Distribution	6	1.20	.305	.001
Method × R^2 × Sample Size × Reliability × RSS	6	.56	.766	.000
Method × R^2 × Sample Size × Data Distribution × RSS	6	2.59	.016	.001
Method × R^2 × Reliability × Data Distribution × RSS	3	.34	.798	.000
Method × Sample Size × Reliability × Data Distribution × RSS	2	.34	.711	.000
Method × SMH × R^2 × Sample Size × Reliability × Data Distribution	18	.49	.965	.001
Method × SMH × R^2 × Sample Size × Reliability × RSS	18	.44	.980	.001
Method × SMH × R^2 × Sample Size × Data Distribution × RSS	18	1.76	.024	.003
Method × SMH × R^2 × Reliability × Data Distribution × RSS	9	.47	.897	.000
Method × SMH × Sample Size × Reliability × Data Distribution × RSS	6	1.62	.138	.001
Method × R^2 × Sample Size × Reliability × Data Distribution × RSS	6	.32	.928	.000
Method × SMH × R^2 × Sample Size × Reliability × Data Distribution × RSS	18	.83	.667	.001
Error(Method)	11,136			

Note: df = degrees of freedom; MAB = mean absolute bias; RSS = relative segment size; SMH = structural model heterogeneity.

Appendix F

ANOVA Results—Model 2 (Formative Measures)

Tables F1 to F7 present the ANOVA results for model 2 (formative measures) explaining MAB by method (PLS-POS/FIMIX-PLS) and the seven design factors. All significant and substantial effects (i.e., all effects that explain more than 2 percent of the total variance in MAB implying a partial η^2 of more than .02) are highlighted in grey.

We find that the R^2, structural and measurement model heterogeneity, sample size, multicollinearity and data distribution, the interaction of structural and measurement model heterogeneity, and the interaction of sample size and relative segment size have a substantial and significant effect on the MAB of both methods. Furthermore, there is a significant and substantial difference in the parameter recovery (MAB) of the two methods (PLS-POS and FIMIX-PLS) and for the two-way interaction effects between method and R^2, multicollinearity, and structural and measurement model heterogeneity. Method even has a significant and substantial interaction effect with both structural and measurement model heterogeneity (three-way interaction).

Table F1. Between-Subjects Effects (Part I)

Source of Variance in MAB	df	F	Sig.	Partial η^2
Intercept	1	142,696.80	.00	.740
SMH	3	7,605.33	.00	.313
MMH	2	2,912.99	.00	.104
R^2	3	4,286.31	.00	.204
Sample Size	2	864.77	.00	.033
RSS	1	629.83	.00	.012
Data Distribution	1	1,465.75	.00	.028
Multicollinearity	2	848.18	.00	.033
SMH × MMH	6	298.09	.00	.034
SMH × R^2	9	44.28	.00	.008
MMH × R^2	6	5.82	.00	.006

Note: df = degrees of freedom; MAB = mean absolute bias; MMH = measurement model heterogeneity; RSS = relative segment size; SMH = structural model heterogeneity; all significant and substantial effects (i.e., all effects that explain more than 2% of the total variance in MAB implying a partial η^2 of more than .02) are highlighted in grey.

Table F2. Between-Subjects Effects (Part II)

Source of Variance in MAB	df	F	Sig.	Partial η^2
SMH × Sample Size	6	31.10	.00	.004
MMH × Sample Size	4	15.06	.00	.001
R^2 × Sample Size	6	46.43	.00	.006
SMH × RSS	3	78.68	.00	.005
MMH × RSS	2	.69	.50	.000
R^2 × RSS	3	87.86	.00	.005
Sample Size × RSS	2	1,426.86	.00	.054
SMH × Data Distribution	3	12.04	.00	.001
MMH × Data Distribution	2	7.61	.00	.000
R^2 × Data Distribution	3	3.21	.02	.000
Sample Size × Data Distribution	2	28.39	.00	.001
RSS × Data Distribution	1	2.26	.13	.000
SMH × Multicollinearity	6	109.17	.00	.013
MMH × Multicollinearity	4	287.84	.00	.022
R^2 × Multicollinearity	6	5.39	.00	.001
Sample Size × Multicollinearity	4	28.36	.00	.002
RSS × Multicollinearity	2	15.71	.00	.001
Data Distribution × Multicollinearity	2	16.50	.00	.001
SMH × MMH × R^2	18	25.86	.00	.009
SMH × MMH × Sample Size	12	5.18	.00	.001
SMH × R^2 × Sample Size	18	.78	.73	.000
MMH × R^2 × Sample Size	12	.48	.93	.000
SMH × MMH × RSS	6	5.48	.00	.001
SMH × R^2 × RSS	9	.60	.80	.000
MMH × R^2 × RSS	6	2.66	.01	.000
SMH × Sample Size × RSS	6	42.87	.00	.005
MMH × Sample Size × RSS	4	6.23	.00	.000
R^2 × Sample Size × RSS	6	59.73	.00	.007
SMH × MMH × Data Distribution	6	3.35	.00	.000
SMH × R^2 × Data Distribution	9	12.58	.00	.002
MMH × R^2 × Data Distribution	6	1.79	.10	.000
SMH × Sample Size × Data Distribution	6	9.02	.00	.001
MMH × Sample Size × Data Distribution	4	2.33	.05	.000
R^2 × Sample Size × Data Distribution	6	2.76	.01	.000
SMH × RSS × Data Distribution	3	13.81	.00	.001
MMH × RSS × Data Distribution	2	1.50	.22	.000
R^2 × RSS × Data Distribution	3	2.64	.05	.000
Sample Size × RSS × Data Distribution	2	21.48	.00	.001
SMH × MMH × Multicollinearity	12	18.31	.00	.004
SMH × R^2 × Multicollinearity	18	7.30	.00	.003
MMH × R^2 × Multicollinearity	12	1.16	.31	.000
SMH × Sample Size × Multicollinearity	12	11.15	.00	.003
MMH × Sample Size × Multicollinearity	8	3.17	.00	.001
R^2 × Sample Size × Multicollinearity	12	.88	.57	.000
SMH × RSS × Multicollinearity	6	12.44	.00	.001
MMH × RSS × Multicollinearity	4	8.08	.00	.001

Note: df = degrees of freedom; MAB = mean absolute bias; MMH = measurement model heterogeneity; RSS = relative segment size; SMH = structural model heterogeneity; all significant and substantial effects (i.e., all effects that explain more than 2% of the total variance in MAB implying a partial η^2 of more than .02) are highlighted in grey.

Table F3. Between-Subjects Effects (Part III)

Source of Variance in MAB	df	F	Sig.	Partial η^2
R^2 × RSS × Multicollinearity	6	1.29	.26	.000
Sample Size × RSS × Multicollinearity	4	18.22	.00	.001
SMH × Data Distribution × Multicollinearity	6	.94	.46	.000
MMH × Data Distribution × Multicollinearity	4	3.81	.00	.000
R^2 × Data Distribution × Multicollinearity	6	.88	.51	.000
Sample Size × Data Distribution × Multicollinearity	4	11.09	.00	.001
RSS × Data Distribution × Multicollinearity	2	12.97	.00	.001
SMH × MMH × R^2 × Sample Size	36	.75	.86	.001
SMH × MMH × R^2 × RSS	18	.86	.63	.000
SMH × MMH × Sample Size × RSS	12	5.31	.00	.001
SMH × R^2 × Sample Size × RSS	18	1.92	.01	.001
MMH × R^2 × Sample Size × RSS	12	.36	.98	.000
SMH × MMH × R^2 × Data Distribution	18	1.65	.04	.001
SMH × MMH × Sample Size × Data Distribution	12	3.87	.00	.001
SMH × R^2 × Sample Size × Data Distribution	18	1.36	.14	.000
MMH × R^2 × Sample Size × Data Distribution	12	.68	.78	.000
SMH × MMH × RSS × Data Distribution	6	1.80	.09	.000
SMH × R^2 × RSS × Data Distribution	9	1.57	.12	.000
MMH × R^2 × RSS × Data Distribution	6	.54	.78	.000
SMH × Sample Size × RSS × Data Distribution	6	8.98	.00	.001
MMH × Sample Size × RSS × Data Distribution	4	3.19	.01	.000
R^2 × Sample Size × RSS × Data Distribution	6	1.04	.40	.000
SMH × MMH × R^2 × Multicollinearity	36	2.16	.00	.002
SMH × MMH × Sample Size × Multicollinearity	24	.79	.75	.000
SMH × R^2 × Sample Size × Multicollinearity	36	1.62	.01	.001
MMH × R^2 × Sample Size × Multicollinearity	24	1.04	.41	.000
SMH × MMH × RSS × Multicollinearity	12	2.41	.00	.001
SMH × R^2 × RSS × Multicollinearity	18	1.19	.26	.000
MMH × R^2 × RSS × Multicollinearity	12	1.38	.17	.000
SMH × Sample Size × RSS × Multicollinearity	12	9.08	.00	.002
MMH × Sample Size × RSS × Multicollinearity	8	1.95	.05	.000
R^2 × Sample Size × RSS × Multicollinearity	12	1.38	.17	.000
SMH × MMH × Data Distribution × Multicollinearity	12	6.34	.00	.002

Note: df = degrees of freedom; MAB = mean absolute bias; MMH = measurement model heterogeneity; RSS = relative segment size; SMH = structural model heterogeneity.

Table F4. Between-Subjects Effects (Part IV)

Source of Variance in MAB	df	F	Sig.	Partial η^2
SMH × R^2 × Data Distribution × Multicollinearity	18	1.72	.03	.001
MMH × R^2 × Data Distribution × Multicollinearity	12	1.12	.34	.000
SMH × Sample Size × Data Distribution × Multicollinearity	12	10.19	.00	.002
MMH × Sample Size × Data Distribution × Multicollinearity	8	.87	.54	.000
R^2 × Sample Size × Data Distribution × Multicollinearity	12	2.23	.01	.001
SMH × RSS × Data Distribution × Multicollinearity	6	9.02	.00	.001
MMH × RSS × Data Distribution × Multicollinearity	4	.49	.74	.000
R^2 × RSS × Data Distribution × Multicollinearity	6	1.10	.36	.000
Sample Size × RSS × Data Distribution × Multicollinearity	4	24.61	.00	.002
SMH × MMH × R^2 × Sample Size × RSS	36	.75	.86	.001
SMH × MMH × R^2 × Sample Size × Data Distribution	36	.74	.88	.001
SMH × MMH × R^2 × RSS × Data Distribution	18	1.20	.25	.000
SMH × MMH × Sample Size × RSS × Data Distribution	12	1.62	.08	.000
SMH × R^2 × Sample Size × RSS × Data Distribution	18	.69	.83	.000
MMH × R^2 × Sample Size × RSS × Data Distribution	12	1.20	.27	.000
SMH × MMH × R^2 × Sample Size × Multicollinearity	72	1.13	.21	.002
SMH × MMH × R^2 × RSS × Multicollinearity	36	1.66	.01	.001
SMH × MMH × Sample Size × RSS × Multicollinearity	24	1.66	.02	.001
SMH × R^2 × Sample Size × RSS × Multicollinearity	36	.52	.99	.000
MMH × R^2 × Sample Size × RSS × Multicollinearity	24	.75	.81	.000
SMH × MMH × R^2 × Data Distribution × Multicollinearity	36	.95	.55	.001
SMH × MMH × Sample Size × Data Distribution × Multicollinearity	24	1.52	.05	.001
SMH × R^2 × Sample Size × Data Distribution × Multicollinearity	36	1.33	.09	.001
MMH × R^2 × Sample Size × Data Distribution × Multicollinearity	24	.90	.60	.000
SMH × MMH × RSS × Data Distribution × Multicollinearity	12	1.52	.11	.000
SMH × R^2 × RSS × Data Distribution × Multicollinearity	18	1.90	.01	.001
MMH × R^2 × RSS × Data Distribution × Multicollinearity	12	1.45	.14	.000
SMH × Sample Size × RSS × Data Distribution × Multicollinearity	12	8.65	.00	.002
MMH × Sample Size × RSS × Data Distribution × Multicollinearity	8	1.13	.34	.000
R^2 × Sample Size × RSS × Data Distribution × Multicollinearity	12	.85	.60	.000
SMH × MMH × R^2 × Sample Size × RSS × Data Distribution	36	.98	.51	.001
SMH × MMH × R^2 × Sample Size × RSS × Multicollinearity	72	.84	.84	.001
SMH × MMH × R^2 × Sample Size × Data Distribution × Multicollinearity	72	1.07	.33	.002
SMH × MMH × R^2 × RSS × Data Distribution × Multicollinearity	36	1.24	.15	.001
SMH × MMH × Sample Size × RSS × Data Distribution × Multicollinearity	24	1.12	.32	.001
SMH × R^2 × Sample Size × RSS × Data Distribution × Multicollinearity	36	1.09	.32	.001
MMH × R^2 × Sample Size × RSS × Data Distribution × Multicollinearity	24	.87	.65	.000
SMH × MMH × R^2 × Sample Size × RSS × Data Distribution × Multicollinearity	72	1.05	.36	.002
Error	50,112			

Note: df = degrees of freedom; MAB = mean absolute bias; MMH = measurement model heterogeneity; RSS = relative segment size; SMH = structural model heterogeneity.

Table F5. Within-Subjects Effects (Part I)

Source of Variance in MAB	df	F	Sig.	Partial η^2
Method	1	3,938.52	.00	.073
Method × SMH	3	3,987.98	.00	.193
Method × MMH	2	6,771.05	.00	.213
Method × R^2	3	826.32	.00	.047
Method × Sample Size	2	227.55	.00	.009
Method × RSS	1	171.66	.00	.003
Method × Data Distribution	1	2.97	.08	.000
Method × Multicollinearity	2	1,739.12	.00	.065
Method × SMH × MMH	6	976.49	.00	.105
Method × SMH × R^2	9	83.50	.00	.015
Method × MMH × R^2	6	6.13	.00	.001
Method × SMH × Sample Size	6	22.80	.00	.003
Method × MMH × Sample Size	4	3.13	.01	.000
Method × R^2 × Sample Size	6	3.95	.00	.000
Method × SMH × RSS	3	60.96	.00	.004
Method × MMH × RSS	2	12.78	.00	.001
Method × R^2 × RSS	3	15.69	.00	.001
Method × Sample Size × RSS	2	163.40	.00	.006
Method × SMH × Data Distribution	3	54.31	.00	.003
Method × MMH × Data Distribution	2	3.39	.03	.000
Method × R^2 × Data Distribution	3	5.19	.00	.000
Method × Sample Size × Data Distribution	2	12.45	.00	.000
Method × RSS × Data Distribution	1	56.16	.00	.001
Method × SMH × Multicollinearity	6	372.96	.00	.043
Method × MMH × Multicollinearity	4	257.24	.00	.020
Method × R^2 × Multicollinearity	6	9.69	.00	.001
Method × Sample Size × Multicollinearity	4	22.84	.00	.002
Method × RSS × Multicollinearity	2	5.85	.00	.000
Method × Data Distribution × Multicollinearity	2	11.81	.00	.000
Method × SMH × MMH × R^2	18	11.49	.00	.004
Method × SMH × MMH × Sample Size	12	2.44	.00	.001
Method × SMH × R^2 × Sample Size	18	3.68	.00	.001
Method × MMH × R^2 × Sample Size	12	1.39	.16	.000
Method × SMH × MMH × RSS	6	14.80	.00	.002
Method × SMH × R^2 × RSS	9	12.50	.00	.002
Method × MMH × R^2 × RSS	6	2.61	.02	.000
Method × SMH × Sample Size × RSS	6	47.94	.00	.006
Method × MMH × Sample Size × RSS	4	13.37	.00	.001
Method × R^2 × Sample Size × RSS	6	19.62	.00	.002
Method × SMH × MMH × Data Distribution	6	1.74	.11	.000
Method × SMH × R^2 × Data Distribution	9	5.01	.00	.001
Method × MMH × R^2 × Data Distribution	6	3.04	.01	.000
Method × SMH × Sample Size × Data Distribution	6	7.68	.00	.001
Method × MMH × Sample Size × Data Distribution	4	.30	.88	.000
Method × R^2 × Sample Size × Data Distribution	6	3.34	.00	.000
Method × SMH × RSS × Data Distribution	3	3.68	.01	.000
Method × MMH × RSS × Data Distribution	2	.76	.47	.000
Method × R^2 × RSS × Data Distribution	3	.43	.73	.000
Method × Sample Size × RSS × Data Distribution	2	19.04	.00	.001
Method × SMH × MMH × Multicollinearity	12	28.62	.00	.007
Method × SMH × R^2 × Multicollinearity	18	5.04	.00	.002
Method × MMH × R^2 × Multicollinearity	12	.46	.94	.000

Note: df = degrees of freedom; MAB = mean absolute bias; MMH = measurement model heterogeneity; RSS = relative segment size; SMH = structural model heterogeneity; all significant and substantial effects (i.e., all effects that explain more than 2% of the total variance in MAB implying a partial η^2 of more than .02) are highlighted in grey.

Table F6. Within-Subjects Effects (Part II)				
Source of Variance in MAB	**df**	**F**	**Sig.**	**Partial η^2**
Method × SMH × Sample Size × Multicollinearity	12	11.91	.00	.003
Method × MMH × Sample Size × Multicollinearity	8	1.40	.19	.000
Method × R^2 × Sample Size × Multicollinearity	12	.91	.53	.000
Method × SMH × RSS × Multicollinearity	6	16.91	.00	.002
Method × MMH × RSS × Multicollinearity	4	3.91	.00	.000
Method × R^2 × RSS × Multicollinearity	6	1.19	.31	.000
Method × Sample Size × RSS × Multicollinearity	4	20.68	.00	.002
Method × SMH × Data Distribution × Multicollinearity	6	6.57	.00	.001
Method × MMH × Data Distribution × Multicollinearity	4	3.63	.01	.000
Method × R^2 × Data Distribution × Multicollinearity	6	.99	.43	.000
Method × Sample Size × Data Distribution × Multicollinearity	4	24.39	.00	.002
Method × RSS × Data Distribution × Multicollinearity	2	28.84	.00	.001
Method × SMH × MMH × R^2 × Sample Size	36	1.35	.08	.001
Method × SMH × MMH × R^2 × RSS	18	1.48	.08	.001
Method × SMH × MMH × Sample Size × RSS	12	1.99	.02	.000
Method × SMH × R^2 × Sample Size × RSS	18	2.48	.00	.001
Method × MMH × R^2 × Sample Size × RSS	12	2.34	.01	.001
Method × SMH × MMH × R^2 × Data Distribution	18	.86	.63	.000
Method × SMH × MMH × Sample Size × Data Distribution	12	2.68	.00	.001
Method × SMH × R^2 × Sample Size × Data Distribution	18	1.28	.19	.000
Method × MMH × R^2 × Sample Size × Data Distribution	12	.37	.97	.000
Method × SMH × MMH × RSS × Data Distribution	6	1.18	.32	.000
Method × SMH × R^2 × RSS × Data Distribution	9	3.45	.00	.001
Method × MMH × R^2 × RSS × Data Distribution	6	.51	.80	.000
Method × SMH × Sample Size × RSS × Data Distribution	6	8.37	.00	.001
Method × MMH × Sample Size × RSS × Data Distribution	4	1.21	.31	.000
Method × R^2 × Sample Size × RSS × Data Distribution	6	1.13	.34	.000
Method × SMH × MMH × R^2 × Multicollinearity	36	1.29	.11	.001
Method × SMH × MMH × Sample Size × Multicollinearity	24	1.28	.16	.001
Method × SMH × R^2 × Sample Size × Multicollinearity	36	1.36	.08	.001
Method × MMH × R^2 × Sample Size × Multicollinearity	24	1.05	.40	.001
Method × SMH × MMH × RSS × Multicollinearity	12	3.27	.00	.001
Method × SMH × R^2 × RSS × Multicollinearity	18	1.02	.43	.000
Method × MMH × R^2 × RSS × Multicollinearity	12	1.40	.16	.000
Method × SMH × Sample Size × RSS × Multicollinearity	12	8.14	.00	.002
Method × MMH × Sample Size × RSS × Multicollinearity	8	2.47	.01	.000
Method × R^2 × Sample Size × RSS × Multicollinearity	12	1.36	.18	.000
Method × SMH × MMH × Data Distribution × Multicollinearity	12	2.63	.00	.001
Method × SMH × R^2 × Data Distribution × Multicollinearity	18	1.65	.04	.001
Method × MMH × R^2 × Data Distribution × Multicollinearity	12	.82	.63	.000
Method × SMH × Sample Size × Data Distribution × Multicollinearity	12	7.24	.00	.002
Method × MMH × Sample Size × Data Distribution × Multicollinearity	8	1.01	.42	.000
Method × R^2 × Sample Size × Data Distribution × Multicollinearity	12	1.42	.15	.000
Method × SMH × RSS × Data Distribution × Multicollinearity	6	6.94	.00	.001
Method × MMH × RSS × Data Distribution × Multicollinearity	4	1.40	.23	.000
Method × R^2 × RSS × Data Distribution × Multicollinearity	6	1.59	.15	.000
Method × Sample Size × RSS × Data Distribution × Multicollinearity	4	15.65	.00	.001
Method × SMH × MMH × R^2 × Sample Size × RSS	36	1.88	.00	.001
Method × SMH × MMH × R^2 × Sample Size × Data Distribution	36	.80	.80	.001
Method × SMH × MMH × R^2 × RSS × Data Distribution	18	1.00	.45	.000

Note: df = degrees of freedom; MAB = mean absolute bias; MMH = measurement model heterogeneity; RSS = relative segment size; SMH = structural model heterogeneity.

Table F7. Within-Subjects Effects (Part III)				
Source of Variance in MAB	**df**	**F**	**Sig.**	**Partial η^2**
Method × SMH × MMH × Sample Size × RSS × Data Distribution	12	2.14	.01	.001
Method × SMH × R^2 × Sample Size × RSS × Data Distribution	18	1.53	.07	.001
Method × MMH × R^2 × Sample Size × RSS × Data Distribution	12	.77	.68	.000
Method × SMH × MMH × R^2 × Sample Size × Multicollinearity	72	.91	.70	.001
Method × SMH × MMH × R^2 × RSS × Multicollinearity	36	1.28	.12	.001
Method × SMH × MMH × Sample Size × RSS × Multicollinearity	24	1.95	.00	.001
Method × SMH × R^2 × Sample Size × RSS × Multicollinearity	36	1.37	.07	.001
Method × MMH × R^2 × Sample Size × RSS × Multicollinearity	24	.90	.60	.000
Method × SMH × MMH × R^2 × Data Distribution × Multicollinearity	36	.98	.50	.001
Method × SMH × MMH × Sample Size × Data Distribution × Multicollinearity	24	2.46	.00	.001
Method × SMH × R^2 × Sample Size × Data Distribution × Multicollinearity	36	1.49	.03	.001
Method × MMH × R^2 × Sample Size × Data Distribution × Multicollinearity	24	.70	.85	.000
Method × SMH × MMH × RSS × Data Distribution × Multicollinearity	12	1.75	.05	.000
Method × SMH × R^2 × RSS × Data Distribution × Multicollinearity	18	1.71	.03	.001
Method × MMH × R^2 × RSS × Data Distribution × Multicollinearity	12	1.37	.17	.000
Method × SMH × Sample Size × RSS × Data Distribution × Multicollinearity	12	8.67	.00	.002
Method × MMH × Sample Size × RSS × Data Distribution × Multicollinearity	8	1.29	.24	.000
Method × R^2 × Sample Size × RSS × Data Distribution × Multicollinearity	12	.78	.68	.000
Method × SMH × MMH × R^2 × Sample Size × RSS × Data Distribution	36	.85	.73	.001
Method × SMH × MMH × R^2 × Sample Size × RSS × Multicollinearity	72	1.05	.36	.002
Method × SMH × MMH × R^2 × Sample Size × Data Distribution × Multicollinearity	72	1.20	.11	.002
Method × SMH × MMH × R^2 × RSS × Data Distribution × Multicollinearity	36	1.53	.02	.001
Method × SMH × MMH × Sample Size × RSS × Data Distribution × Multicollinearity	24	2.53	.00	.001
Method × SMH × R^2 × Sample Size × RSS × Data Distribution × Multicollinearity	36	1.33	.09	.001
Method × MMH × R^2 × Sample Size × RSS × Data Distribution × Multicollinearity	24	1.25	.18	.001
Method × SMH × MMH × R^2 × Sample Size × RSS × Data Distribution × Multicollinearity	72	.96	.58	.001
Error(Method)	50,112			

Note: df = degrees of freedom; MAB = mean absolute bias; MMH = measurement model heterogeneity; RSS = relative segment size; SMH = structural model heterogeneity.

References

Alavi, M., and Joachimsthaler, E. A. 1992. "Revisiting DSS Implementation Research: A Meta- Analysis of the Literature and Suggestions for Researchers," *MIS Quarterly* (16:1), pp. 95-116.

Anderberg, M. R. 1973. *Cluster Analysis for Applications*, New York: Academic Press.

Dennis, A. R., Wixom, B. H., and Vandenberg, R. J. 2001. "Understanding Fit and Appropriation Effects in Group Support Systems via Meta-Analysis," *MIS Quarterly* (25:2), pp. 167-193.

Diamantopoulos, A., and Winklhofer, H. M. 2001. "Index Construction with Formative Indicators: An Alternative to Scale Development," *Journal of Marketing Research* (38:2), pp. 269-277.

Esposito Vinzi, V., Trinchera, L., and Amato, S. 2010. "PLS Path Modeling: From Foundations to Recent Developments and Open Issues for Model Assessment and Improvement," in *Handbook of Partial Least Squares: Concepts, Methods and Applications*, V. Esposito Vinzi, W. W. Chin, J. Henseler, and H. Wang (eds.), Berlin: Springer, pp. 47-82.

Esposito Vinzi, V., Trinchera, L., Squillacciotti, S., and Tenenhaus, M. 2008. "REBUS-PLS: A Response-Based Procedure for Detecting Unit Segments in PLS Path Modelling," *Applied Stochastic Models in Business & Industry* (24:5), pp. 439-458.

Grewal, R., Cote, J. A., and Baumgartner, H. 2004. "Multicollinearity and Measurement Error in Structural Equation Models: Implications for Theory Testing," *Marketing Science* (23:4), pp. 519-529.

Gudergan, S. P., Ringle, C. M., Wende, S., and Will, A. 2008. "Confirmatory Tetrad Analysis in PLS Path Modeling," *Journal of Business Research* (61:12), pp. 1238-1249.

Hahn, C., Johnson, M. D., Herrmann, A., and Huber, F. 2002. "Capturing Customer Heterogeneity Using a Finite Mixture PLS Approach," *Schmalenbach Business Review (SBR)* (54:3), pp. 243-269.

Henseler, J., Ringle, C. M., and Sinkovics, R. R. 2009. "The Use of Partial Least Squares Path Modeling in International Marketing," in *Advances in International Marketing,* R. R. Sinkovics and P. N. Ghauri (eds.), Bingley, United Kingdom: Emerald Group Publishing Limited, pp. 277-320.

Henseler, J., and Sarstedt, M. 2012. "Goodness-of-Fit Indices for Partial Least Squares Path Modeling," *Computational Statistics*, (http://link.springer.com/article/10.1007%2Fs00180-012-0317-1).

Jarvis, C. B., MacKenzie, S. B., and Podsakoff, P. M. 2003. "A Critical Review of Construct Indicators and Measurement Model Misspecification in Marketing and Consumer Research," *Journal of Consumer Research* (30:2), pp. 199-218

Joseph, D., Kok-Yee, N., Koh, C., and Soon, A. 2007. "Turnover of Information Technology Professionals: A Narrative Review, Meta-Analytic Structural Equation Modeling, and Model Development," *MIS Quarterly* (31:3), pp. 547-577.

King, W. R., and He, J. 2006. "A Meta-Analysis of the Technology Acceptance Model," *Information & Management* (43:6), pp. 740-755.

Kohli, R., and Devaraj, S. 2003. "Measuring Information Technology Payoff: A Meta-Analysis of Structural Variables in Firm-Level Empirical Research," *Information Systems Research* (14:2), pp. 127-145.

Lee, G., and Xia, W. 2006. "Organizational Size and IT Innovation Adoption: A Meta-Analysis," *Information & Management* (43:8), pp. 975-985.

Lohmöller, J.-B. 1989. *Latent Variable Path Modeling with Partial Least Squares*, Heidelberg: Physica.

Mason, C. H., and Perreault, W. D. 1991. "Collinearity, Power, and Interpretation of Multiple Regression Analysis," *Journal of Marketing Research* (28:3), pp. 268-280.

R Core Team. 2013. "R: A Language and Environment for Statistical Computing," R Foundation for Statistical Computing, Vienna.

Sabherwal, R., Jeyaraj, A., and Chowa, C. 2006. "Information System Success: Individual and Organizational Determinants," *Management Science* (52:12), pp. 1849-1864.

Sánchez, G., and Trinchera, L. 2013. "R Package PLSPM (version 0.3.5)," http://cran.r-project.org/web/packages/plspm/.

Sarstedt, M. 2008. "A Review of Recent Approaches for Capturing Heterogeneity in Partial Least Squares Path Modelling," *Journal of Modelling in Management* (3:2), pp. 140-161.

Schepers, J., and Wetzels, M. 2007. "A Meta-Analysis of the Technology Acceptance Model: Investigating Subjective Norm and Moderation Effects," *Information & Management* (44:1), pp. 90-103.

Sharma, R., and Yetton, P. 2003. "The Contingent Effects of Management Support and Task Interdependence on Successful Information Systems Implementation," *MIS Quarterly* (27:4), pp. 533-555.

Sharma, R., and Yetton, P. 2007. "The Contingent Effects of Training, Technical Complexity, and Task Interdependence on Successful Information Systems Implementation," *MIS Quarterly* (31:2), pp. 219-238.

Squillacciotti, S. 2005. "Prediction Oriented Classification in PLS Path Modeling," in *PLS & Marketing: Proceedings of the 4th International Symposium on PLS and Related Methods*, T. Aluja, J. Casanovas, V. Esposito Vinzi and M. Tenenhaus (eds.), Paris: DECISIA, pp. 499-506.

Squillacciotti, S. 2010. "Prediction-Oriented Classification in PLS Path Modeling," in *Handbook of Partial Least Squares: Concepts, Methods and Applications,* V. Esposito Vinzi, W. W. Chin, J. Henseler, and H. Wang (eds.), Berlin: Springer, pp. 219-233.

Tenenhaus, M., Esposito Vinzi, V., Chatelin, Y.-M., and Lauro, C. 2005. "PLS Path Modeling," *Computational Statistics & Data Analysis* (48:1), pp. 159-205

Wang, J., and Keil, M. 2007. "A Meta-Analysis Comparing the Sunk Cost Effect for IT and Non-IT Projects," *Information Resources Management Journal* (20:3), pp. 1-18.

Wedel, M., and Kamakura, W. 2000. *Market Segmentation: Conceptual and Methodological Foundations* (2nd ed.), New York: Kluwer Academic Publishers.

Wold, H. 1982. "Soft Modeling: The Basic Design and Some Extensions," in: *Systems Under Indirect Observations: Part I,* K. G. Jöreskog and H. Wold (eds.), Amsterdam: North-Holland, pp. 1-54.

Wu, J., and Lederer, A. 2009. "A Meta-Analysis of the Role of Environment-Based Voluntariness in Information Technology Acceptance," *MIS Quarterly* (33:2), pp. 419-432.

Copyright of MIS Quarterly is the property of MIS Quarterly & The Society for Information Management and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.