

A test for multigroup comparison using partial least squares path modeling

Michael Kleesel

University of Siegen, Siegen, Germany and

Department of Design, Production and Management,

Faculty of Engineering Technology, University of Twente, Enschede, The Netherlands

Florian Schuberth

Department of Design, Production and Management,

Faculty of Engineering Technology, University of Twente, Enschede, The Netherlands

Jörg Henseler

Department of Design, Production and Management,

Faculty of Engineering Technology, University of Twente,

Enschede, The Netherlands and

Nova Information Management School, Universidade Nova de Lisboa,

Lisboa, Portugal, and

Bjoern Niehaves

University of Siegen, Siegen, Germany

Abstract

Purpose – People seem to function according to different models, which implies that in business and social sciences, heterogeneity is a rule rather than an exception. Researchers can investigate such heterogeneity through multigroup analysis (MGA). In the context of partial least squares path modeling (PLS-PM), MGA is currently applied to perform multiple comparisons of parameters across groups. However, this approach has significant drawbacks: first, the whole model is not considered when comparing groups, and second, the family-wise error rate is higher than the predefined significance level when the groups are indeed homogenous, leading to incorrect conclusions. Against this background, the purpose of this paper is to present and validate new MGA tests, which are applicable in the context of PLS-PM, and to compare their efficacy to existing approaches.

Design/methodology/approach – The authors propose two tests that adopt the squared Euclidean distance and the geodesic distance to compare the model-implied indicator correlation matrix across groups. The authors employ permutation to obtain the corresponding reference distribution to draw statistical inference about group differences. A Monte Carlo simulation provides insights into the sensitivity and specificity of both permutation tests and their performance, in comparison to existing approaches.

Findings – Both proposed tests provide a considerable degree of statistical power. However, the test based on the geodesic distance outperforms the test based on the squared Euclidean distance in this regard. Moreover, both proposed tests lead to rejection rates close to the predefined significance level in the case of no group differences. Hence, our proposed tests are more reliable than an uncontrolled repeated comparison approach.

Research limitations/implications – Current guidelines on MGA in the context of PLS-PM should be extended by applying the proposed tests in an early phase of the analysis. Beyond our initial insights, more research is required to assess the performance of the proposed tests in different situations.



Originality/value – This paper contributes to the existing PLS-PM literature by proposing two new tests to assess multigroup differences. For the first time, this allows researchers to statistically compare a whole model across groups by applying a single statistical test.

Keywords Monte Carlo simulation, Multigroup analysis, Partial least squares path modelling, Permutation test

Paper type Research paper

1. Introduction

The empirical testing of theories requires valid statistical methods to allow researchers to derive reliable implications. In the field of information systems (IS) and internet research, partial least squares path modeling (PLS-PM) is a widely used composite-based estimator for structural equation models with latent variables to investigate phenomena such as social networks (Cheung *et al.*, 2015), internet addiction (Lu and Wang, 2008) and mobile banking (Tam and Oliveira, 2017). It was originally developed by Herman A. O. Wold in the 1970s as an alternative estimator for structural equation models (Wold, 1975).

The existing literature on PLS-PM has provided substantial methodological contributions that have increased its application in various disciplines, such as strategic management (Hulland, 1999), IS research (Ringle *et al.*, 2012) and tourism research (Müller *et al.*, 2018). Notable milestones include the proposal of the confirmatory tetrad analysis (Gudergan *et al.*, 2008) and the heterotrait–monotrait ratio of correlations (Henseler *et al.*, 2015) and the development of consistent PLS (PLSc; Dijkstra and Henseler, 2015a), which enhances traditional PLS-PM to consistently estimate structural models containing common factors.

Researchers often assume that data sets in empirical research stem from a single homogeneous population. Contrary to this assumption, data sets used in social sciences are regularly affected by heterogeneity, which implies that the data were collected from different homogenous populations. Ignoring this fact, i.e., not taking heterogeneity into account, leads to questionable conclusions (Jedidi *et al.*, 1997). Hence, a multigroup analysis (MGA) can be conducted to investigate this issue caused by heterogeneity.

Heterogeneity has been recognized in the context of PLS-PM (e.g. Huma *et al.*, 2017), and several approaches have been adopted to define groups in the case of unobserved heterogeneity based on genetic algorithm segmentation (Ringle *et al.*, 2014) and iterative reweighted regression (Schlittgen *et al.*, 2016). Moreover, parametric and non-parametric approaches (Keil *et al.*, 2000; Chin and Dibbern, 2010; Henseler, 2012) have been proposed to assess parameter differences and, thus, heterogeneity across groups. However, the existing approaches have serious drawbacks. First, they do not compare the whole model but compare only specific parameters, e.g., path coefficients, to investigate heterogeneity. Second, the employed testing procedures rely on distributional assumptions, e.g., normal distributed data, which are often violated in empirical research. Finally, since the existing approaches rely on multiple comparisons, complex models with numerous relationships and more than two groups significantly affect the number of comparisons. Hence, researchers applying current approaches face the risk of a high family-wise error rate (FWER).

Against this background, this paper proposes two tests that allow for comparing a whole model across groups while maintaining the predefined significance level under the null hypothesis. For that purpose, we consider established distance measures, namely, the geodesic distance and the squared Euclidean distance, to measure the discrepancy of the model-implied correlation matrix of the indicators across groups. To obtain the reference distribution of the corresponding test statistic (distance measure) under the null hypothesis of no group differences, we employ a permutation procedure.

This paper is structured as follows: After the introduction, in Section 2, we review the existing literature on MGA in the context of PLS-PM and emphasize the importance of having a test that allows the comparison of the whole model across groups. In Section 3, we

propose two novel MGA tests and show how permutation can be used for significance testing. In Section 4, these new tests are evaluated by means of a Monte Carlo simulation. Finally, we extend current MGA guidelines in the context of PLS-PM by proposing a comprehensive test procedure that integrates our proposed test into existing approaches and discuss opportunities for future research.

2. MGA using PLS-PM

MGA can be used to explore differences across groups defined by group variables. Heterogeneity across groups in MGA occurs if there are significant differences across at least two groups. To address heterogeneity, researchers can estimate separate models per group or control for heterogeneity by means of a categorical moderator variable (Sarstedt, Henseler and Ringle, 2011). Regardless of how it is addressed, ignoring heterogeneity affects the complete underlying research model.

In terms of unobserved heterogeneity, cluster analysis such as k -means clustering has been widely used in the context of PLS-PM to identify partitions that are used for group-specific estimations (Hair *et al.*, 2018; Sarstedt and Mooi, 2014). A major shortcoming of this approach is that the structural model, which is a major aspect in structural equation modeling (SEM), is not taken into account. To overcome this drawback, the literature provides several approaches, such as finite mixture partial least squares (Hahn *et al.*, 2002; Sarstedt, Becker, Ringle and Schwaiger, 2011), the prediction-oriented segmentation in PLS-PM (Becker *et al.*, 2013) and the iterative reweighted regression segmentation method for PLS-PM (Schlittgen *et al.*, 2016). For a more complete overview of techniques, we refer to Hair *et al.* (2016).

In addition to uncovering unobserved heterogeneity, the previous literature has also suggested different approaches to test for observed heterogeneity across groups (Hair *et al.*, 2018). For two-group scenarios, a repeated application of unpaired sample t -tests has been proposed to identify differences between groups (Chin, 2000; Keil *et al.*, 2000). In doing so, the test statistic is assumed to follow a t -distribution where the standard errors of the parameter estimates are obtained by the bootstrap or jackknife procedure (Keil *et al.*, 2000). To overcome distributional assumptions, the previous literature has also provided a non-parametric test for MGA (Henseler, 2012). Although this test is similar to the former, it evaluates the bootstrap distribution of each group to analyze whether the estimates statistically differ between groups. Similarly, Chin (2003) and Chin and Dibbern (2010) propose a permutation test to evaluate group differences. Group-specific differences are compared with the corresponding reference distribution obtained by the permutation procedure. Apart from the analysis of two groups, approaches for multiple groups have been suggested, for example, the omnibus test of group differences, which is a combinatorial test comprising bootstrapping and permutation to mimic an overall F -test (Sarstedt, Henseler and Ringle, 2011).

A variety of tests allow testing heterogeneity across groups. However, these tests have some significant shortcomings: first, they do not consider the whole model; instead, they focus on only specific parameters, thus excluding information from the model. For instance, the procedure proposed by Chin (2003) and Chin and Dibbern (2010) suggests a simultaneous comparison of path coefficients. We argue that in the early stages of research or in complex models, a researcher might be interested not only in differences between path coefficients but also in comparing whole models across groups. Second, the use of repeated t -tests to investigate differences, e.g., in path coefficients, narrows the relaxation of distributional assumptions, which is considered a major advantage of PLS-PM. To maintain this advantage, a non-parametric test would be preferable. Finally, the simultaneous comparison of multiple parameters involves the risk of inducing an FWER, which is the probability that at least one single test makes one Type I error (falsely rejecting the null hypothesis) (Dudoit *et al.*, 2004). However, for a proper single statistical test, the type I error rate is usually determined by the significance level, and the simultaneous application of multiple tests increases the FWER if

not controlled. Hence, the aforementioned testing procedures in MGA face the risk of an FWER that is too high, i.e., the null hypothesis of no group differences is rejected too often. This issue is particularly relevant in scenarios with multiple groups and complex models, as the number of comparisons increases significantly. Let p be the number of parameters and G be the number of models; the number of overall comparisons c is calculated as follows:

$$c = p \binom{G}{2} = \frac{p \cdot G!}{2!(G-2)!} \tag{1}$$

For example, investigating whether four groups are heterogeneous with respect to ten parameters requires 60 statistical tests. Assuming a significance level of $\alpha=0.05$, the probability of falsely rejecting the null hypothesis of homogeneity across groups (the FWER) is $1-(1-\alpha)^c$. Without any further corrections, there is a 95.39 percent chance that at least one of the comparisons is statistically significant when the null hypothesis is indeed correct. This issue is also relevant when few parameters are compared (an overview is given in the Appendix, Table AI). Hence, MGA with repeated comparisons is required to take FWER into account.

To show how MGA is used in IS research and how the FWER is controlled, we conducted a literature review. We queried the Web of Science database, including publications from nine leading journals from the IS domain (*European Journal of Information Systems, Information Systems Journal, Information Systems Research, Internet Research, Journal of the Association for Information Systems, Journal of Information Technology, Journal of Management Information Systems, Journal of Strategic Information Systems* and *Management Information Systems Quarterly*). We used “multigroup” and “group differences” as keywords and included all articles applying PLS-PM. Since applying MGA is often a partial issue and therefore not mentioned explicitly in the papers’ abstract, we also included illustrative examples of references in the pertinent literature (Qureshi and Compeau, 2009). For each paper, we identified the grouping variable and its levels, how the significance level was adjusted and the number of path coefficients relevant for the MGA. Since the considered papers do not report any kind of correction of the significance level, we compute the FWER according to Equation (1). An overview is provided in Table I (sorted by year of publication).

Our review indicates that interest in MGA has increased. Fundamental papers that paved the way for MGA in the context of PLS-PM (Keil *et al.*, 2000; Qureshi and Compeau, 2009) have undoubtedly contributed to this development. At the same time, this review also shows that issues associated with multiple comparison tests, i.e., the FWER, have received little attention so far. Across all reviewed papers, there was no report of any kind of

Reference	Grouping variable	Paths	Comparisons	FWER (%)
Keil <i>et al.</i> (2000)	Culture (Finland, The Netherlands, Singapore)	5	15	53.67
Ahuja and Thatcher (2005)	Gender (male, female)	5	5	22.62
Srite and Karahanna (2006)	Cultural values (individualism, collectivism)	4	4	18.55
Zhu <i>et al.</i> (2006)	Users (EDI user, non-user)	16	16	55.99
Hsieh <i>et al.</i> (2008)	Economically (advantaged, disadvantaged)	9	9	36.98
Sia <i>et al.</i> (2009)	Cultural differences (Australia, Hong Kong)	6	6	26.49
Shen <i>et al.</i> (2010)	Gender (male, female)	6	6	26.49
Yeh <i>et al.</i> (2012)	Gender (male, female)	4	4	18.55
Dibbern <i>et al.</i> (2012)	Country (Germany, USA)	5	5	22.62
Zhou <i>et al.</i> (2015)	Indulgence (high indulgence, low indulgence)	4	4	18.55
Huma <i>et al.</i> (2017)	Organization (private, public)	6	6	26.46
Shi <i>et al.</i> (2018)	Gender (male, female)	3	3	14.26

Table I.
Multigroup analysis in IS research

correction (e.g. Bonferroni correction). Hence, we are inclined to assume that a correction was not applied. In conclusion, most papers might be affected by a considerable inflation of the FWER (14.26 percent < FWER < 55.99 percent). This issue particularly affects studies that investigate more than two groups and/or a high number of path coefficients.

3. A test for MGA

3.1 Measuring heterogeneity across groups

Here, we propose two new tests to compare whole models across groups to investigate heterogeneity. Similar to the test for overall model fit in PLS-PM (Dijkstra and Henseler, 2015b), which considers the discrepancy between the empirical indicator covariance matrix and the model-implied counterpart, we propose to investigate the distances between the indicator model-implied correlation matrices across groups[1].

To determine the differences between the model-implied indicator correlation matrices across groups, every measure that satisfies the properties of a distance (Deza and Deza, 2016, p. 3) can be used. Consequently, a distance greater than zero implies that two groups differ. If there is a statistically significant distance between the groups, further steps can be conducted to investigate the differences in more depth, e.g., investigation of specific path coefficients.

For the purpose of our research, i.e., assessing the distances between model-implied indicator correlation matrices, we consider two established distance measures: the geodesic distance and the squared Euclidean distance. While the squared Euclidean distance is well known, the geodesic distance is illustrated as follows: It belongs to Swain’s (1975) class of fitting functions and can be employed to estimate the model parameter in SEM. Properly scaled, it is asymptotically equal to the fitting function used in the maximum likelihood (ML) estimation for SEM.

In the case of two groups, the geodesic distance between the model-implied correlation matrix of group 1 ($\Sigma(\theta_1)$) and group 2 ($\Sigma(\theta_2)$) is calculated as follows:

$$d_g = \frac{1}{2} \sum_{i=1}^K \ln(\varphi_i)^2, \tag{2}$$

where φ_i is the i -th eigenvalue of the matrix $\Sigma(\theta_1)^{-1}\Sigma(\theta_2)$ and K is the number of rows of one of these two matrices. When the two matrices are equal, the geodesic distance is zero since all eigenvalues of a unit matrix are one.

The squared Euclidean distance between $\Sigma(\theta_1)$ and $\Sigma(\theta_2)$ is calculated as follows:

$$d_E = \frac{1}{2} \sum_{i=1}^K \sum_{j=1}^K (\sigma_{ij,1} - \sigma_{ij,2})^2, \tag{3}$$

where K is again the number of rows and $\sigma_{ij,1}$ and $\sigma_{ij,2}$ are elements of the respective matrix. If both matrices are identical, the squared Euclidean distance is zero; otherwise, this distance is greater than zero.

Since MGA is often conducted with more than two groups, we calculate the arithmetic mean of the distances of all possible pairs of groups. Note that the total number of group comparisons is $G(G-1)/2$, where G is the number of groups. Therefore, the average geodesic distance (D_g) for G groups is calculated as follows:

$$D_g = \frac{2}{G(G-1)} \sum_{g=2}^G \sum_{h=1}^{g-1} \sum_{i=1}^K \ln(\varphi_i)^2, \tag{4}$$

where φ_i is the i -th eigenvalue of the matrix $\Sigma(\theta_g)^{-1}\Sigma(\theta_h)$.

In a similar manner, we calculate the average squared Euclidean distance for more than two groups as follows:

$$D_e = \frac{2}{G(G-1)} \sum_{g=2}^G \sum_{h=1}^{g-1} \sum_{i=1}^K \sum_{j=1}^K (\sigma_{ij,g} - \sigma_{ij,h})^2, \quad (5)$$

where $\sigma_{ij,g}$ and $\sigma_{ij,h}$ are the elements of the corresponding model-implied correlation matrix. Since the squared Euclidean and the geodesic distance are either larger than or equal to zero, the two proposed average distances cannot be negative. Moreover, these distances are zero if all correlation matrices are equal; otherwise, they are larger than zero.

In terms of MGA, the considered null hypothesis is as follows: $H_0: \Sigma(\theta_1) = \dots = \Sigma(\theta_g) = \dots = \Sigma(\theta_G)$, where $\Sigma(\theta_g)$ is the model-implied population correlation matrix of the indicators for group g . To obtain the reference distribution of the distance measures including the average distances, we apply a permutation procedure, as described below.

3.2 Permutation tests

Permutation tests were introduced by Sir Ronald Fisher (1935) as a general approach for statistical inferences and have been considered the gold standard in medicine research (Edgington and Onghena, 2007, p. 9). There are three common types of permutation tests: exact permutation tests, moment–approximation permutation tests and resampling–approximation permutation tests (Berry *et al.*, 2014). All three types of tests share the characteristic that they use permutation to obtain the distribution of the test statistic under the null hypothesis. The exact permutation test obtains the reference distribution by calculating the test statistic for all possible permutations of the original data set. Thus, the number of calculations can grow considerably with an increasing number of observations. Consequently, the application of an exact permutation test is not always reasonable. The moment–approximation permutation test requires the computation of the exact moments of the test statistic, which are then used to fit a specific distribution. In turn, this distribution is used for the calculation of the p -value. The resampling–approximation permutation test is similar to the exact permutation test; however, the reference distribution of the test statistic is based on only a subset of all possible permutations of the original sample. Due to its feasibility, this test is widely used.

Multiple resampling–approximation permutation tests have been developed in the context of PLS-PM; for example, a permutation test for compositional invariance (Henseler *et al.*, 2016) and a permutation test to compare parameters across groups (Chin and Dibbern, 2010) have been developed. This type of permutation test has a distinct advantage compared to parametric tests. For example, it makes no assumptions about the distribution of the test statistic. Since PLS-PM also makes no distributional assumptions, this type of permutation test is perfectly in line with the PLS-PM's spirit. Moreover, such permutation tests have favorable properties for small sample sizes (Ludbrook and Dudley, 1994), and they are robust against extreme values. Therefore, we also choose this type of permutation test to compare the model-implied indicator correlation matrices across groups.

4. Monte Carlo simulation

4.1 Simulation design

We used a Monte Carlo simulation to evaluate the sensitivity (power) and specificity (Parikh *et al.*, 2008) of our two proposed permutation tests, where one is based on the average geodesic distance (D_g) and the other on the average squared Euclidean distance (D_e). While specificity is the test's ability to correctly reveal homogeneity across groups, sensitivity is the ability to correctly detect heterogeneity across groups. To compare the performance of our two proposed permutation tests to existing testing procedures, we also

included a test procedure based on repeated comparisons of path coefficients (RCPC), i.e., the path coefficients are compared across all groups (Chin, 2003; Chin and Dibbern, 2010).

Similar to the previous literature on MGA (e.g. Qureshi and Compeau, 2009), we used a structural population model with four constructs modeled as composites (Figure 1)[2]. All composites consist of three indicators. The population weights to form composites C_2 to C_4 are set to 0.3, 0.5 and 0.6. The weights from C_1 vary according to the following five scenarios, in which we compare three groups: the groups are homogenous; the groups have small differences among their structural models; the groups have moderate differences among their structural models; the population weights of the first composite vary across groups; and in addition to the previous scenario, the structural models also show small differences across groups. Table II presents the manipulated population parameters.

Furthermore, we varied the sample size per group from 100 to 500 observations. Finally, to investigate the robustness of our tests, we consider normally and non-normally distributed samples. To generate the non-normal data, we multiplied the samples drawn from the multivariate standard normal distribution by a scale factor, as proposed by Dijkstra and Henseler (2015b), leading to a kurtosis of approximately 1.74. We expect that the tests perform slightly worse when non-normally distributed, in contrast to the results of tests using normally distributed data. In total, we consider 50 experimental designs (5 scenarios \times 2 different distributions \times 5 sample size). For each design, we conduct 300 runs.

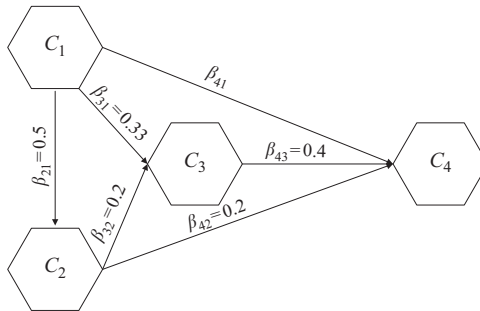


Figure 1. Structural population model

Scenario	D_g	D_e	g	β_{41}	w_{11}	w_{21}	w_{31}
(i) Homogeneity	0	0	1	0	0.30	0.50	0.60
			2	0	0.30	0.50	0.60
			3	0	0.30	0.50	0.60
(ii) Small structural difference	0.0471	0.0133	1	0	0.30	0.50	0.60
			2	0.1	0.30	0.50	0.60
			3	0.2	0.30	0.50	0.60
(iii) Moderate structural differences	0.3293	0.0266	1	0	0.60	0.50	0.30
			2	0.2	0.30	0.50	0.60
			3	0.4	0.30	0.50	0.60
(iv) Different weights ^a	0.2576	0.0337	1	0	0.60	0.50	0.30
			2	0	0.80	0.30	0.30
			3	0	0.38	0.38	0.66
(v) Structural differences and different weights ^a	0.3138	0.0409	1	0	0.60	0.50	0.30
			2	0.1	0.80	0.30	0.30
			3	0.2	0.38	0.38	0.66

Table II. Population parameters Notes: Group (g); average geodesic distance (D_g); average squared Euclidean distance (D_e); ^aWeights are rounded (two digits)

For consistent estimation of the weights, we employed Mode B (Dijkstra, 1981). To obtain the reference distribution of the two test statistics, we used 499 permutation runs. The simulation was implemented in the statistical programming environment *R* (version 3.4.0, R Core Team, 2017) using the *mvrnorm* function of the *MASS* package to draw data from the multivariate normal distribution (version 7.3-47, Ripley *et al.*, 2017) and the *matrixpls* package to estimate the specified model with the same structure as the population models (version 1.0.5, Rönkkö, 2017).

4.2 Simulation results

The produced rejection rates of the two permutation tests are displayed in Table III.

Homogeneity (scenario (i)). The degree of specificity is shown in the first rows of Table III (“Homogeneity”). For this scenario, our two tests maintain the predefined significance level of 5 percent quite well, while the repeated comparison testing procedure rejects the null hypothesis of no group differences far too often (> 48.0 percent).

Structural differences (scenarios (ii) and (iii)). Concerning small structural differences, both new tests are limited in terms of their rejection rates. In most of the conditions, the rejection rates are below the recommended threshold of 80 percent (Cohen, 1988). For moderate structural differences, both permutation tests reliably detect differences in most of the conditions. However, the repeated comparison test produces even higher rejection rates.

Different weights (scenarios (iv) and (v)). The results also confirm that our approach is able to detect heterogeneity in groups with different weights. In situations where only the weights differ, both new tests perform quite well, i.e., high rejection rates and outperform the RCPC approach in most conditions. It is notable that the new tests perform even better if both the structural model and the weights differ across groups.

Scenario	<i>n</i> per group	Normal data			Non-normal data		
		<i>D_g</i> (%)	<i>D_e</i> (%)	RCPC (%)	<i>D_g</i> (%)	<i>D_e</i> (%)	RCPC (%)
(i) Homogeneity	100	5.7	7.0	50.3	5.0	7.3	51.7
	200	5.0	4.3	51.0	3.0	2.7	53.3
	300	2.3	4.7	52.7	6.7	5.7	54.3
	400	4.7	4.7	55.0	4.3	2.3	47.0
	500	4.0	6.3	48.0	4.7	5.0	56.0
(ii) Small structural difference	100	12.3	9.0	68.7	8.7	4.7	64.7
	200	19.7	11.7	77.0	16.7	15.3	71.3
	300	32.0	22.7	84.7	22.7	18.3	78.7
	400	45.7	28.3	91.0	30.7	24.3	86.7
	500	56.7	36.7	96.3	41.0	33.7	92.3
(iii) Moderate structural differences	100	70.3	25.3	91.3	41.0	18.7	85.7
	200	99.0	59.3	99.7	84.3	46.7	97.0
	300	100.0	86.3	100.0	99.0	76.7	100.0
	400	100.0	96.3	100.0	99.7	86.3	100.0
	500	100.0	99.3	100.0	100.0	96.3	100.0
(iv) Different weights	100	54.3	59.0	52.7	37.3	42.7	56.3
	200	97.7	97.0	58.0	85.3	83.7	58.0
	300	100.0	100.0	62.3	99.0	99.3	60.3
	400	100.0	100.0	62.0	99.7	99.7	59.3
	500	100.0	100.0	63.0	100.0	100.0	62.0
(v) Structural differences and different weights	100	71.7	70.0	64.3	51.7	56.3	61.3
	200	99.7	99.0	80.7	93.0	94.7	72.3
	300	100.0	100.0	89.7	100.0	100.0	82.0
	400	100.0	100.0	93.3	100.0	100.0	89.3
	500	100.0	100.0	97.0	100.0	100.0	95.3

Table III.
Rejection rates

Sample size and data distribution. Across all conditions, all tests benefit from an increasing sample size, which results in a higher statistical power. Moreover, our results confirm that all tests perform slightly worse once the data are non-normally distributed. However, with a sufficiently large sample size, heterogeneity across groups can still be detected. With regard to moderate structural differences, 200 observations are necessary to obtain a sufficient degree of power (D_g : 99.0 percent) for normally distributed data. For non-normal data, 300 observations per group are necessary to achieve a similar level of statistical power (D_g : 99.0 percent).

Summary. Overall, the test based on the average geodesic distance produces higher rejection rates than the test based on the average squared Euclidean distance. The highest rejection rates for structural differences are produced for scenario (iii), i.e., moderate differences in the structural model across groups. Here, the test based on the average geodesic distance provides acceptable results, even if the sample size is small ($n = 200$; 99.0 percent). In contrast, the test based on the average squared Euclidean distance detects heterogeneity in only 59.3 percent of the cases.

5. Discussion

Despite the prevalence of heterogeneity in the social sciences, a test to compare a whole model across groups has not been available thus far. To allow for such a comparison, this study contributes to the existing literature by proposing two novel permutation tests based on the average geodesic distance (D_g) and the average squared Euclidean distance (D_e).

Our simulation study provides initial evidence that the two tests are viable options when the aim is to detect heterogeneity across multiple groups. Most importantly, the two proposed tests are capable of maintaining a predefined significance level. Hence, homogenous groups are not falsely classified as heterogeneous. This is a major advantage over the RCPC, which yields an inflation of FWER when not adjusted properly.

Based on the power results, the test based on the average geodesic distance is superior across the considered conditions. In particular, in situations with only small differences across groups, D_g outperforms D_e . As expected, an increasing sample size improves the power of all tests. Our simulation results also indicate that 100 observations are not sufficient for an acceptable power. Instead, 200 observations are required to exceed the threshold. This is in line with previous studies that highlighted the requirement of a sufficient sample size to detect heterogeneity (Qureshi and Compeau, 2009).

As indicated by the results of the variation in population composite weights, our approach is also able to detect differences within a composite model across groups. This highlights a major strength of this generic approach because it can be used for different purposes and is not limited to structural differences only. However, to apply MGA, it is important to establish measurement invariance in advance. Otherwise, an MGA may yield misleading or incorrect conclusions (Henseler *et al.*, 2016). Therefore, we recommend applying the presented tests after measurement invariance is established.

Against this background, current MGA guidelines in the context of PLS-PM need to be extended. If measurement invariances are established, we recommend initiating MGA by providing the results of one of the proposed tests, preferably the test based on the average geodesic distance. Only if there is a significant difference in the model-implied indicator correlation matrices across groups should existing techniques that allow for the investigation of single effects be applied. In fact, if a grouping variable does not lead to significant differences between groups, a researcher should either reject heterogeneity or respecify a grouping variable before conducting further analyses. Therefore, we propose new MGA guidelines that comprise our proposed tests and existing MGA procedures and consist of the following four steps displayed in Figure 2:

- (1) Establish measurement invariance: before conducting an MGA, a researcher should establish measurement invariance (Henseler *et al.*, 2016). Otherwise, an MGA is not

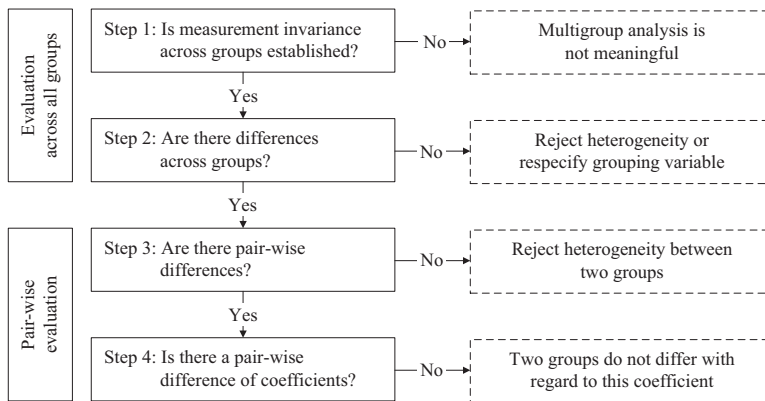


Figure 2.
A comprehensive test
procedure for MGA

meaningful. If measurement invariance is established, the subsequent steps can be applied to test for heterogeneity.

- (2) Overall evaluation: testing differences across all groups is considered the starting point for MGA. With this initial test, a researcher is able to determine whether groups differ significantly. This initial effort is particularly important when more than two groups are considered. If the test does not support heterogeneity, a researcher can either reject heterogeneity or respecify the grouping variable.
- (3) Pair-wise evaluation: If heterogeneity was found in the previous step, the purpose of this step is to investigate the heterogeneity in more detail. For that purpose, the proposed tests can be used for each pair of groups to examine which groups actually differ.
- (4) Effect-wise evaluation: finally, the differences are investigated with regard to specific coefficients such as path coefficients. For that purpose, researchers can draw from parametric approaches (Chin, 2000; Keil *et al.*, 2000) or non-parametric approaches (Henseler, 2012). As a result, we can determine whether there are group differences with respect to a specific effect.

6. Limitations and outlook

This paper presents initial insights into the efficacy of the tests for the comparison of the model-implied indicator correlation matrices across groups, while other questions remain unanswered and should be addressed by future research. The simulation study should be extended to further investigate the tests' performance. Important extensions include the consideration of unequal group sizes, a population model with a non-saturated structural model and path coefficients with positive and negative signs. Moreover, since PLS-PM in its current form, i.e., PLSc, is a composite-based estimator that can be used to estimate models containing both composites and factors, future research could further investigate the performance of our proposed tests for that type of model. We argue that the permutation tests should be based on the model-implied indicator correlation matrix to compare the whole model across groups. However, the permutation tests may also be based on other matrices, such as the model-implied construct correlation matrix, so that differences in only the structural model are investigated across groups. We chose to utilize two established distance measures, namely, the squared Euclidean distance and the geodesic distance. Since the literature provides several other distance measures that may also be useful (Deza and Deza, 2016), future research could compare their performance in the context of our proposed testing procedure. Because PLSc

(Dijkstra and Henseler, 2015a) encourages the use of PLS-PM for factor models, it also seems necessary to compare the test's performance to the performance of tests typically used in factor-based SEM. Finally, although not explicitly emphasized, our tests for multigroup comparison are of a confirmatory nature. Hence, it should be used with caution when applied to groups created by cluster analysis or similar techniques.

Notes

1. For a better comparison, we consider the model-implied correlation matrix of the indicators instead of their model-implied covariance matrix.
2. Since PLS-PM is often applied to models with more than four constructs (Ringle *et al.*, 2012), we also ran the simulation with a larger model that includes eight composites. Since the results were comparable, they are not reported here.

References

- Ahuja, M.K. and Thatcher, J.B. (2005), "Moving beyond intentions and toward the theory of trying: effects of work environment and gender on post-adoption information technology use", *MIS Quarterly*, Vol. 29 No. 3, pp. 427-459.
- Becker, J.-M., Rai, A., Ringle, C.M. and Völckner, F. (2013), "Discovering unobserved heterogeneity in structural equation models to avert validity threats", *MIS Quarterly*, Vol. 37 No. 3, pp. 665-694.
- Berry, K.J., Johnston, J.E. and Mielke, P.W. (2014), *A Chronicle of Permutation Statistical Methods*, Springer International Publishing, London.
- Cheung, C., Lee, Z.W.Y. and Chan, T.K.H. (2015), "Self-disclosure in social networking sites: the role of perceived cost, perceived benefits and social influence", *Internet Research*, Vol. 25 No. 2, pp. 279-299.
- Chin, W.W. (2000), "Frequently asked questions – partial least squares & PLS-graph", available at: <http://disc-nt.cba.uh.edu/chin/plsfaq.htm> (accessed September 22, 2017).
- Chin, W.W. (2003), "A permutation procedure for multi-group comparison of PLS models", in Vilares, M., Tenenhaus, M., Coelho, P., Esposito Vinzi, V. and Morineau, A. (Eds), *PLS and Related Methods*, Decisia, Lisbon, pp. 33-43.
- Chin, W.W. and Dibbern, J. (2010), "An introduction to a permutation based procedure for multi-group PLS analysis: results of tests of differences on simulated data and a cross cultural analysis of the sourcing of information system services between Germany and the USA", in Esposito Vinzi, V., Chin, W.W., Henseler, J. and Wang, H. (Eds), *Handbook of Partial Least Squares*, Springer, Berlin, pp. 171-193.
- Cohen, J. (1988), *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed., Routledge, New York, NY.
- Deza, M. and Deza, E. (2016), *Encyclopedia of Distances*, 4th ed., Springer, Heidelberg and New York, NY.
- Dibbern, J., Chin, W. and Heinzl, A. (2012), "Systemic determinants of the information systems outsourcing decision: a comparative study of German and United States firms", *Journal of the Association for Information Systems*, Vol. 13 No. 6, pp. 466-497.
- Dijkstra, T.K. (1981), *Latent Variables in Linear Stochastic Models: Reflections on "Maximum Likelihood" and "Partial Least Squares" Methods*, Groningen University, Groningen.
- Dijkstra, T.K. and Henseler, J. (2015a), "Consistent partial least squares path modeling", *MIS Quarterly*, Vol. 39 No. 2, pp. 297-316.
- Dijkstra, T.K. and Henseler, J. (2015b), "Consistent and asymptotically normal PLS estimators for linear structural equations", *Computational Statistics & Data Analysis*, Vol. 81, pp. 10-23.
- Dudoit, S., van der Laan, M.J. and Pollard, K.S. (2004), "Multiple testing. Part I. single-step procedures for control of general type I error rates", *Statistical Applications in Genetics and Molecular Biology*, Vol. 3 No. 1, pp. 1-69.
- Edgington, E. and Onghena, P. (2007), *Randomization Tests*, 4th ed., CRC Press, Hoboken.
- Fisher, R.A. (1935), *The Design of Experiments*, Oliver and Boyd, Edinburgh.

- Gudergan, S.P., Ringle, C.M., Wende, S. and Will, A. (2008), "Confirmatory tetrad analysis in PLS path modeling", *Journal of Business Research*, Vol. 61 No. 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*, Vol. 54 No. 3, pp. 243-269.
- Hair, J.F., Sarstedt, M., Matthews, L.M. and Ringle, C.M. (2016), "Identifying and treating unobserved heterogeneity with FIMIX-PLS: part I – method", *European Business Review*, Vol. 28 No. 1, pp. 63-76.
- Hair, J.F., Sarstedt, M., Ringle, C.M. and Gudergan, S.P. (2018), *Advanced Issues in Partial Least Squares Structural Equation Modeling*, Sage, Thousand Oaks, CA.
- Henseler, J. (2012), "PLS-MGA: a non-parametric approach to partial least squares-based multi-group analysis", in Gaul, W.A., Geyer-Schulz, A., Schmidt-Thieme, L. and Kunze, J. (Eds), *Challenges at the Interface of Data Analysis, Computer Science, and Optimization*, Springer, Berlin and Heidelberg, pp. 495-501.
- Henseler, J., Ringle, C.M. and Sarstedt, M. (2015), "A new criterion for assessing discriminant validity in variance-based structural equation modeling", *Journal of the Academy of Marketing Science*, Vol. 43 No. 1, pp. 115-135.
- Henseler, J., Ringle, C.M. and Sarstedt, M. (2016), "Testing measurement invariance of composites using partial least squares", *International Marketing Review*, Vol. 33 No. 3, pp. 405-431.
- 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*, Vol. 32 No. 1, pp. 97-119.
- Hulland, J. (1999), "Use of partial least squares (PLS) in strategic management research: a review of four recent studies", Vol. 20 No. 2, pp. 195-204.
- Huma, Z., Hussain, S., Thurasamy, R. and Malik, M.I. (2017), "Determinants of cyberloafing: a comparative study of a public and private sector organization", *Internet Research*, Vol. 27 No. 1, pp. 97-117.
- Jedidi, K., Jagpal, H.S. and DeSarbo, W.S. (1997), "Finite-mixture structural equation models for response-based segmentation and unobserved heterogeneity", *Marketing Science*, Vol. 16 No. 1, pp. 39-59.
- Keil, M., Tan, B.C.Y., Wei, K.-K., Saarinen, T., Tuunainen, V. and Wassenaar, A. (2000), "A cross-cultural study on escalation of commitment behavior in software projects", *MIS Quarterly*, Vol. 24 No. 2, pp. 299-325.
- Lu, H. and Wang, S. (2008), "The role of internet addiction in online game loyalty: an exploratory study", *Internet Research*, Vol. 18 No. 5, pp. 499-519.
- Ludbrook, J. and Dudley, H. (1994), "Issues in biomedical statistics: statistical inference", *ANZ Journal of Surgery*, Vol. 64 No. 9, pp. 630-636.
- Müller, T., Schubert, F. and Henseler, J. (2018), "PLS path modeling – a confirmatory approach to study tourism technology and tourist behavior", *Journal of Hospitality and Tourism Technology*, Vol. 9 No. 3, pp. 249-266.
- Parikh, R., Mathai, A., Parikh, S., Chandra Sekhar, G. and Thomas, R. (2008), "Understanding and using sensitivity, specificity and predictive values", *Indian Journal of Ophthalmology*, Vol. 56 No. 1, pp. 45-50.
- Qureshi, I. and Compeau, D. (2009), "Assessing between-group differences in information systems research: a comparison of covariance- and component-based SEM", *MIS Quarterly*, Vol. 33 No. 1, pp. 197-214.
- R Core Team (2017), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna.
- Ringle, C.M., Sarstedt, M. and Schlittgen, R. (2014), "Genetic algorithm segmentation in partial least squares structural equation modeling", *OR Spectrum*, Vol. 36 No. 1, pp. 251-276.
- Ringle, C.M., Sarstedt, M. and Straub, D.W. (2012), "A critical look at the use of PLS-SEM in MIS quarterly", *MIS Quarterly*, Vol. 36 No. 1, pp. iii-xiv.

- Ripley, B., Venables, B., Bates, D.M., Hornik, K., Gebhardt, A. and Firth, D. (2017), "MASS: support functions and datasets for venables and ripley's MASS", available at: <https://cran.r-project.org/package=MASS> (accessed September 13, 2017).
- Rönkkö, M. (2017), "Matrixpls: matrix-based partial least squares estimation", available at: <https://cran.r-project.org/package=matrixpls> (accessed September 13, 2017).
- Sarstedt, M. and Mooi, E. (2014), *A Concise Guide to Market Research: The Process, Data, and Methods Using IBM SPSS Statistics*, 2nd ed., Springer, Berlin.
- Sarstedt, M., Henseler, J. and Ringle, C.M. (2011), "Multigroup analysis in partial least squares (PLS) path modeling: alternative methods and empirical results", *Advances in International Marketing*, Vol. 22, pp. 195-218.
- Sarstedt, M., Becker, J.-M., Ringle, C.M. and Schwaiger, M. (2011), "Uncovering and treating unobserved heterogeneity with FIMIX-PLS: which model selection criterion provides an appropriate number of segments?", *Schmalenbach Business Review*, Vol. 63 No. 1, pp. 34-62.
- Schlittgen, R., Ringle, C.M., Sarstedt, M. and Becker, J.-M. (2016), "Segmentation of PLS path models by iterative reweighted regressions", *Journal of Business Research*, Vol. 69 No. 10, pp. 4583-4592.
- Shen, A.X., Lee, M.K., Cheung, C.M. and Chen, H. (2010), "Gender differences in intentional social action: we-intention to engage in social network-facilitated team collaboration", *Journal of Information Technology*, Vol. 25 No. 2, pp. 152-169.
- Shi, S., Mu, R., Lin, L., Chen, Y., Kou, G. and Chen, X.-J. (2018), "The impact of perceived online service quality on Swift Guanxi: implications for customer repurchase intention", *Internet Research*, Vol. 28 No. 2, pp. 432-455.
- Sia, C.L., Lim, K.H., Leung, K., Lee, M.K.O., Huang, W.W. and Benbasat, I. (2009), "Web strategies to promote internet shopping: is cultural-customization needed?", *MIS Quarterly*, Vol. 33 No. 3, pp. 491-512.
- Srite and Karahanna (2006), "The role of espoused national cultural values in technology acceptance", *MIS Quarterly*, Vol. 30 No. 3, pp. 679-704.
- Swain, A.J. (1975), "A class of factor analysis estimation procedures with common asymptotic sampling properties", *Psychometrika*, Vol. 40 No. 3, pp. 315-335.
- Tam, C. and Oliveira, T. (2017), "Understanding mobile banking individual performance: the DeLone & McLean model and the moderating effects of individual culture", *Internet Research*, Vol. 27 No. 3, pp. 538-562.
- Wold, H. (1975), "Path models with latent variables: the NIPALS approach", in Blalock, H.M., Aganbegian, A., Borodkin, F.M., Boudon, R. and Capocchi, V. (Eds), *Quantitative Sociology: International Perspectives on Mathematical and Statistical Modeling*, Academic Press, New York, NY, pp. 307-357.
- Yeh, J., Hsiao, K. and Yang, W. (2012), "A study of purchasing behavior in Taiwan's online auction websites: effects of uncertainty and gender differences", *Internet Research*, Vol. 22 No. 1, pp. 98-115.
- Zhou, Z., Jin, X.-L., Fang, Y. and Vogel, D. (2015), "Toward a theory of perceived benefits, affective commitment, and continuance intention in social virtual worlds: cultural values (indulgence and individualism) matter", *European Journal of Information Systems*, Vol. 24 No. 3, pp. 247-261.
- Zhu, K., Kraemer, K.L., Gurbaxani, V. and Xu, S.X. (2006), "Migration to open-standard interorganizational systems: network effects, switching costs, and path dependency", *MIS Quarterly*, Vol. 30, pp. 515-539.

Appendix. FWER in multiple comparison scenarios

Assuming that we have a fixed number of groups with a fixed number of parameters, the total number of comparisons can be determined according to Equation (1). Performing multiple comparisons without correction for Type I errors results in a $1-(1-\alpha)^c$ FWER, where c is the total number of comparisons and α is the predefined significance level for each comparison.

Table AI.
FWER in MGA

Number of groups	Number of parameters	Total number of comparisons (<i>c</i>)	FWER ($\alpha=5\%$) (%)
2	4	4	18.55
	8	8	33.66
	10	10	40.13
3	4	12	45.96
	8	24	70.80
	10	30	78.54
4	4	24	70.80
	8	48	91.47
	10	60	95.39
5	4	40	87.15
	8	80	98.35
	10	100	99.41

About the authors

Michael Klesel works as a Research Associate at the University of Siegen, Germany and is Visiting Scholar at the University of Twente, The Netherlands. His research interests include the individualization of information systems and structural equation modeling. He has published in the *Communications of the Association of Information Systems* and in leading conferences including the International Conference on Information Systems (ICIS), the European Conference on Information Systems (ECIS) and the American Conference on Information Systems (AMCIS).

Florian Schubert is Assistant Professor at the Chair of Product-Market Relations of the University of Twente, Enschede, The Netherlands. He obtained his PhD Degree in Econometrics at the Faculty of Business Management and Economics of the University of Würzburg, Germany. His main research interests are structural equation modeling, in particular on composite-based estimators and their enhancement.

Jörg Henseler holds the Chair of Product-Market Relations at the University of Twente, The Netherlands. His research interests include structural equation modeling and the interface of marketing and design research. He has published in *Computational Statistics & Data Analysis*, *European Journal of Information Systems*, *European Journal of Marketing*, *International Journal of Research in Marketing*, *Journal of the Academy of Marketing Science*, *Journal of Service Management*, *Journal of Supply Chain Management*, *Long Range Planning*, *Management Decision*, *MIS Quarterly*, *Organizational Research Methods* and *Structural Equation Modeling – A Multidisciplinary Journal*, among others. He is author of the ADANCO computer program, he lectures worldwide on theory and applications of structural equation models. Jörg Henseler is the corresponding author and can be contacted at: j.henseler@utwente.nl

Bjoern Niehaves is Full Professor and holds the Chair of Information Systems at the University of Siegen, Germany. He received a PhD Degree in Information Systems and a PhD Degree in Political Science from the University of Münster, Germany. Björn holds or held visiting positions at Harvard University (USA), the London School of Economics and Political Science (UK), Waseda University (Japan), Royal Institute of Technology (Sweden), Copenhagen Business School (Denmark), and Aalto University (Finland). He has published more than 200 research articles.

For instructions on how to order reprints of this article, please visit our website:

www.emeraldgrouppublishing.com/licensing/reprints.htm

Or contact us for further details: permissions@emeraldinsight.com