Multigroup Analysis using SmartPLS: Step-by-Step Guidelines for Business Research

Jun-Hwa Cheah
School of Business and Economics, Universiti Putra Malaysia, Selangor, Malaysia

Ramayah Thurasamy
School of Management, Universiti Sains Malaysia, Penang, Malaysia

Mumtaz Ali Memon
NUST Business School, National University of Sciences and Technology, Islamabad Pakistan

Francis Chuah
Othman Yeop Abdullah Graduate School of Business, Universiti Utara Malaysia, Malaysia

Hiram Ting
Faculty of Hospitality and Tourism Management, UCSI University, Malaysia
College of Business, Polytechnic University of the Philippines, Philippines

Abstract

Multigroup Analysis (MGA) using partial least squares path modelling (PLSPM) is an efficient approach to evaluate moderation across multiple relationships in a research model. This editorial explains the importance and the usage of MGA, especially when a study intends to understand heterogeneity effects (i.e., cross-cultural or gender differences) in a business research. We provide step-by-step guidelines to perform MGA using SmartPLS 3.3.2, including the assessment of measurement invariance using the measurement invariance of composite models (MICOM) procedure. We believe that this editorial will enable researchers and postgraduate students to conduct MGA with confidence to draw meaningful insights from the data and examine the differences across the groups under investigation.

Introduction

Partial least squares path modelling (PLSPM) has been widely used as a composite-based estimator to simultaneously investigate structural equation models with latent variables in business research. This technique was originally developed by Herman Wold in the 1970s as an alternative estimator to covariance-based structural equation modelling (Jöreskog and Wold, 1982; Wold, 1982). The main objective of PLSPM is to maximize the amount of variance explained in the endogenous constructs of the structural model (often viewed as prediction) to facilitate the explanation of the model's relationships (Chin et al., 2020; Hair, 2020; Hwang et al., 2020). The renewed interest in this technique since the early 2000s has resulted in its rapid development and substantial methodological contributions, such as the heterotrait–monotrait ratio of correlations (Henseler et al., 2015), the development of consistent PLS (PLSc) (Dijkstra and Henseler, 2015), discrete choice modelling in PLS (Hair et al., 2019b), weighted PLS (Cheah et al., 2020), necessary condition analysis (Richter et al., 2020), the endogeneity test in PLS (Hult et al., 2018; Sarstedt et al., 2020), PLSpredict (Shmueli et al., 2016, 2019), model selection criteria (Danks et al., 2020; Sharma, 2019a, 2019b), and the cross-validated predictive ability test (Liengaard et al., 2020).

Notwithstanding these recent developments, business researchers often assume that data in empirical research stems from a single homogenous population; on the contrary, in many real-world applications, the assumption of homogeneity is rather unrealistic. Accordingly, many researchers have begun to consider the notion of heterogeneity, where heterogeneous perceptions and evaluations of products and services form the basis of market segmentation. Chin and Dibbern (2010) stressed that ignoring heterogeneity often leads to questionable conclusions. In particular, studies that pool data as a single homogenous population are said to have failed to assess whether there are significant differences across two or more subgroups in the data. Multigroup analysis (MGA) is therefore the recommended approach to address this concern.

MGA or between-group analysis is a means to test predefined (also known as a priori) data groups to determine the existence of significant differences across group-specific parameter estimates (e.g., outer weights, outer loadings, and path coefficients) (Hair et al., 2017). MGA enables researchers to test for variations between different groups in two identical models when the groups are known. The ability to identify the presence or absence of multigroup differences is anchored in the PLSPM technique.

According to Hair et al. (2017), MGA in PLSPM is one of the most efficient ways to assess moderation across multiple relationships. Standard moderation examines a single structural relationship at the point of interaction between two exogenous variables' product and the endogenous variable (i.e., independent variable multiplied with moderator variable predicts dependent variable). In contrast, MGA offers a more complete picture of the moderator's influence on the analysis results as the focus

shifts from examining the moderator's impact on one specific relationship to examining its impact on all modelled relationships (Hair et al., 2017, 2018).

Though it is clearly important and easily accessed through standard PLS-SEM software, such as SmartPLS (Ringle et al., 2015; Sarstedt and Cheah, 2020) and R (see https://github.com/gastonstat/plspm), the utilisation of the MGA technique is relatively low. Additionally, the seminars and workshops on MGA, while frequently offered and made available on social platforms, may not always be practically useful for researchers and postgraduate students either due to the advanced content or incomprehensive presentation. The plethora of inquiries and comments on forums such as that of http://forum.smartpls.com/ further suggest that many novice users of PLSPM do not have the fundamental understanding and skills to conduct MGA and interpret the results. Guidelines and recommendations for conducting MGA remain scarce in methods and empirical papers. Therefore, a basic scholarly article on the application and interpretation of MGA is deemed both timely and necessary.

The present editorial addresses the aforementioned concerns by offering step-by-step guidelines to perform MGA in PLSPM in the business fields, including behavioural studies. It also discusses the key choices one must make when initiating the MGA procedure and interpreting its results. The illustration of the procedure is depicted with a tourism marketing model for a practical learning experience. Lastly, this editorial provides several recommendations on how MGA results should be interpreted. While our presentation is largely about the procedure of MGA, our overarching aim is to assist and encourage researchers and postgraduate students to perform MGA to deal with heterogeneity in a research model in a rigorous manner.

Why is MGA Important?

Given that the underlying assumption of MGA is heterogeneity across groups, it is useful in assessing differences across groups. According to Sarstedt and Ringle (2010), a researcher's failure to consider heterogeneity can threaten the validity of PLSPM results, which in turn leads to incorrect conclusions. In other words, assessing MGA in PLSPM substantially enhances the ability of researchers to identify meaningful differences in multiple relationships across group-specific results (Picon-Berjoyo et al., 2016; Schlagel & Sarstedt, 2016).

For instance, consider a study examining car purchase intention with two predictors, namely quality (X1) and price (X2), in an environment consisting of equal numbers of both male (Group 1) and female (Group 2) consumers. If we pool the data for analysis under the assumption that gender data stems from a homogeneous population, that is, that both male and female consumers share the same perceptions of price and quality, the path coefficient estimates would offer an incomplete picture of the model's relationships (see Figure 1). Using pooled data, the analysis yields a coefficient of 0.4 for both price and quality, leading the researcher to conclude that price and quality are equally important for males and females, even when they are not. This evidence

echoes Sarstedt et al.'s (2009) cautionary assertion that aggregating data and ignoring group-related differences in model estimates may lead to unexpected results.

Conversely, if the researcher were to account for heterogeneity and consider gender a categorical moderator variable, then the effect of quality (X1) on intention to purchase a car (Y) is much stronger among males (p1(1) = 0.60) than females (p1(2) = 0.20). In contrast, price (X2) has a stronger influence on intention to purchase a car (Y) among females (p2(2) = 0.50) than males (p2(1) = 0.30). This implies that one group of consumers (female) is more price sensitive but less quality sensitive, while the other group (male) is the opposite. From this illustration, it should be understood that ignoring heterogeneity may severely affect the underlying research model.

Figure 1: Heterogeneity Effect in PLSPM
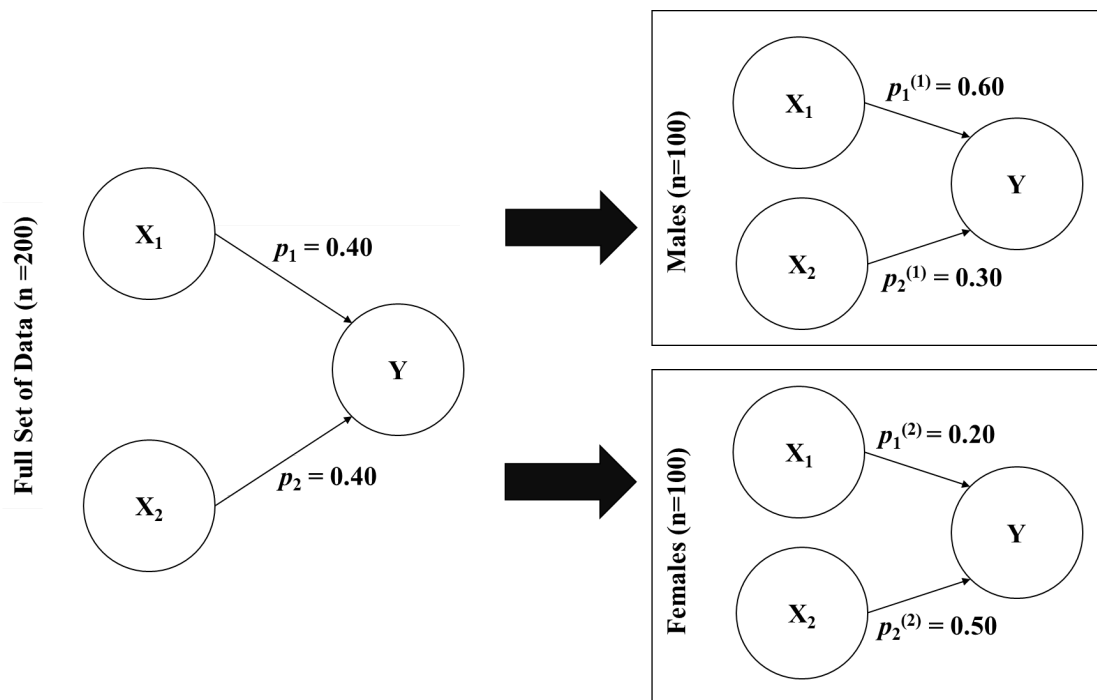
Overview of Steps to Run MGA in PLS-SEM

Figure 3 provides an overview of the steps to assess MGA in PLSPM. The overview is followed by an in-depth discussion of each step using an empirical example previously published in Tourism Management. SmartPLS 3.3.2 was used to illustrate the MGA assessment (Ringle et al., 2015; Sarstedt and Cheah, 2019).

Step 1: Data Preparation

One of the important aspects when dealing with MGA is sufficient statistical power once the data's sample is divided into subgroups. In other words, when performing MGA, it is imperative to ensure that the subgroups have sufficient power to detect the moderating effect (Becker et al., 2013; Hair et al., 2017).

The easiest way to ensure statistical power is to have large sample sizes for both groups. A statistical power of 80% should be computed and reported using G*Power to establish whether the absence of any moderating effect is due to low statistical power rather than the lack of a true moderating effect (Aguinis et al., 2017; Cohen, 1988; Hair et al., 2017; Memon et al., 2020). Alternatively, a researcher could use the gamma-exponential method or the inverse square root method to determine the minimum sample size required for PLSPM path models (Kock and Hadaya, 2016). Specifically, if researchers do not know the path coefficient with the minimum absolute magnitude, the minimum sample size required would be 160 based on the inverse square root method and 146 based on the gamma exponential method. Although both methods sometimes lead to a small overestimation, the slight imperfection allows us to safeguard against the issue of undetected moderation due to insufficient power.

Even if total sample size is relatively large, unequal sample sizes across the moderator-based subgroups would decrease statistical power and lead to the underestimation of moderating effects (Hair et al., 2017). To address this issue, researchers should try to select similar sample sizes for each group so that sample variance is maximized (see Aguinis et al., 2017). However, bear in mind that while oversampling a smaller group of respondents can improve the statistical power of the particular group, it also results in misrepresentation of the smaller group in the sampling frame relative to the actual study population.

Researchers also need to assess data quality before performing MGA in PLSPM. For instance, data should be checked for straight lining patterns, which occur when a respondent marks the same response for almost all survey items. For example, if a 7-point Likert scale was used to rate the items, response patterns with only '1' or '7' (end response) or only '4' (middle response) should be removed from the dataset (see Hair et al., 2019a). Straight lining responses minimise variability and lead to undetected (or detected but underestimated) moderator effects in MGA. Thus, researchers should ensure that the variable distributions for both sample groups are able to capture the full range of possible values. In addition, to detect moderator effects via MGA, a researcher should select well-established measures and assess both the reliability and validity of their measures (either reflective or formative) as recommended by Hair et al. (2019c). Low reliability and validity can cause the erroneous dismissal or underestimation of moderating effects in MGA.

Step 2: Generate Data Groups for MGA in PLSPM

After data preparation, the second step is to create groups by selecting the categorical variable of interest from the dataset. Theory and observation play an important role in generating the data groups. For instance, if empirical studies show that males and females produce different results in terms of the effects of a product's price and quality on actual purchase behaviour, then the researcher needs to position gender as a moderator to examine the overall relationships. This process would be similar if the researcher is interested to compare a path coefficient across more than two groups, as in cross-cultural studies (Malaysia vs Australia, Malaysia vs Singapore, and Australia vs Singapore). For more advanced MGA, a researcher can even compare more than one categorical variable using different combinations of categorical variables (e.g.

gender and generation), which creates multiple outcome groups (i.e., Gen-Y·females, Gen-Y·males, Gen-Z·females, and Gen-Z·males).

Step 3: Test for Measurement Invariance

Once the groups are generated, a researcher has to assess measurement invariance in PLSPM. Measurement invariance or measurement equivalence is a means of confirming that the measurement models specify measures of the same attribute under different conditions (Henseler et al., 2016). Differences in the paths (or β values) between latent variables could stem from different meanings attributed by a group's respondents to the phenomena being examined, rather than from true differences in the structural relationships. The reasons for these discrepancies include: (i) cross-national differences that emerge from culture-specific response styles (e.g., Johnson et al., 2005), such as acquiescence, i.e. the tendency to agree with questions regardless of their content (Sarstedt and Mooi, 2019); (ii) individual characteristics (e.g., gender, ethnicity, etc.) that entail responding to instruments in systematically different ways; and (iii) use of available scale options differently, i.e. the tendency to choose or not to choose extremes. Hult et al. (2008) stressed that failure to establish invariance can easily lead to the low power of statistical tests, the poor precision of estimators, and misleading results. Therefore, it is a fundamental step prior to conducting MGA because it gives researchers the confidence that group differences in model estimates do not result from the distinctive content and/or meanings of the latent variable across groups.

To safeguard the validity of outcomes and conclusions, Henseler et al. (2016) developed the measurement invariance of composite models (MICOM) procedure, which builds on the scores of the latent variable that fit with the trait of composite modelling in PLSPM. There are three steps in the MICOM procedure, namely, the assessments of (i) configural invariance (Step I), (ii) compositional invariance (Step II), and (iii) the equality of a composite's mean value and variance across groups (Step III) (see Hair et al., 2018 for further explanation of these invariance steps). If both configural invariance (Step I) and compositional invariance (Step II) are established, then *partial measurement invariance* is confirmed, and researchers can proceed to compare the path coefficients with the MGA. On the other hand, *full measurement invariance* is established if, in addition to fulfilling partial measurement invariance (Step I and Step II), composites exhibit equal means and variances across the groups (Step III). With full measurement invariance, pooling the data is a possible option (i.e. it will increase statistical power), rendering MGA unnecessary (see Henseler et al., 2016). However, if Step I, Step II, and either requirement of Step III (either equality of composite variance or equality of composite mean) are achieved, then the researcher can claim *partial measurement invariance* and proceed with MGA.

Figure 2: The Measurement Invariance of Composite Models (MICOM) Procedure
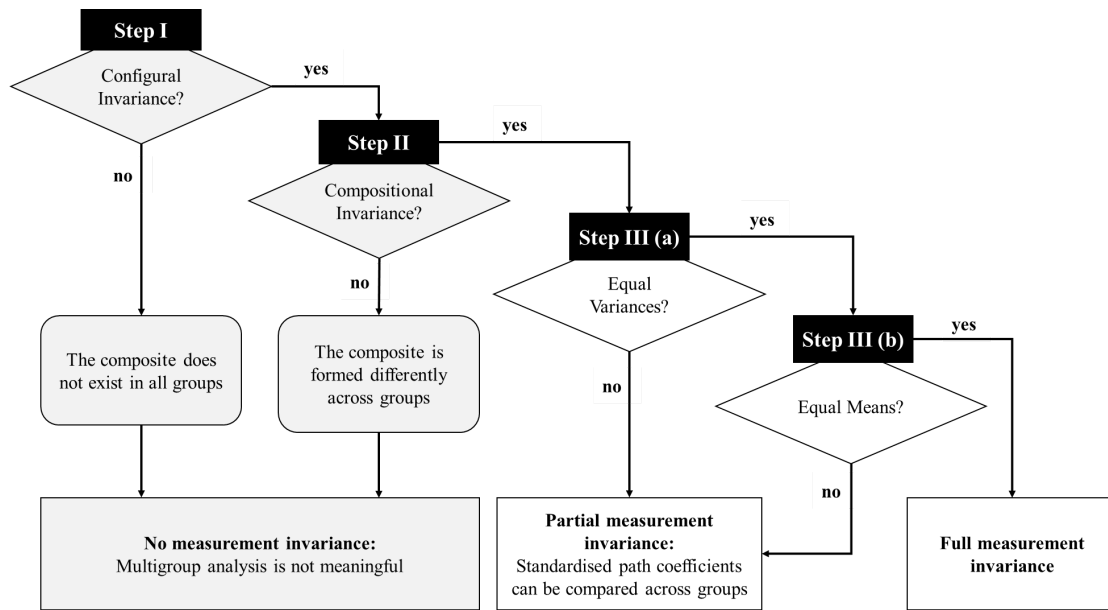
Note: The boxes shaded in grey indicate that achievement of both configural and compositional invariance are compulsory steps to proceed with the MGA technique.

Step 4: Tests for Multigroup Comparisons

Once measurement invariance, either partial or full, is established using MICOM, the researcher can begin assessing group differences using MGA in PLSPM. MGA is often used to compare parameters (e.g., path coefficients, outer weights, outer loadings, etc.) between two or more groups when they are known a priori (Hair et al., 2018).

To make group comparisons, SmartPLS offers five different assessment approaches based on bootstrapping (Hair et al., 2018). The SmartPLS software has three approaches, i.e. Henseler's bootstrap-based MGA (Henseler et al., 2009), the Parametric Test (Keil et al., 2000), and the Welch-Satterthwait Test (Welch, 1947). The fourth approach to assess group comparisons is the permutation test (Chin and Dibbern, 2010), which can be estimated using the MICOM path coefficient option in SmartPLS. However, if researchers would like to compare more than two groups (e.g., Malaysia vs Australia, Malaysia vs Singapore, and Australia vs Singapore), they can opt for the fifth approach, which is the Omnibus Test of Group Differences (OTG) (Sarstedt et al., 2011).

Each of these approaches has its advantages and disadvantages (see Hair et al., 2018). For instance, the parametric test approach is rather lenient and subject to type I errors. This approach also relies on the assumption of normal distribution, so it is not consistent with the nature of PLS that uses non-parametric assumptions (Hair et al., 2018; Sarstedt et al., 2011). The Welch-Satterthwaite test, meanwhile, is a variant of the parametric test that does not assume equal variances when comparing the means of two groups. In contrast, both Henseler's PLS-MGA test (Henseler et al., 2009) and the permutation test use nonparametric assumptions.

Henseler's PLS-MGA procedure (Henseler et al., 2009) is a result of the probability value of a one-tailed test by comparing each bootstrap estimate of one group to all the bootstrap estimates of the same parameter in the other group (Hair et al., 2011). Since the bootstrap distributions for this procedure are not necessarily symmetrical, the use of a two-sided hypothesis is not possible (Henseler et al., 2009) This approach appears to be appropriate; however, the interpretation of the results may be somewhat challenging to due to the condition of the one-tailed test. To keep the explanation simple, researchers must be aware that Henseler's PLS-MGA result is significant at the 5% probability level, whereby the p-value for the difference in group-specific path coefficients should be smaller than 0.05 or larger than 0.95.

Next, the permutation test is a separate option that is run concurrently with Step II of the test for measurement invariance. That is, the output of path coefficients from the measurement invariance test is a means of comparing the path coefficients of subgroups. Notably, Chin and Dibbern (2010) and Hair et al. (2018) recommend the permutation test due to its ability to control for type I errors and its relatively conservative nature compared to the parametric test. Nevertheless, researchers must ensure that there are no large differences in group-specific sample sizes to prevent adverse consequences on the permutation test's performance. Hair et al. (2018) suggested that when one group's sample is more than double the size of the other group, researchers are recommended to choose between two options, which are (i) to select Henseler's PLS-MGA approach (in the case of testing a one-sided hypothesis) or (ii) to randomly draw another sample for the large group that is comparable in size to the smaller group, and subsequently compare the two samples using the permutation test.

The fifth approach, the OTG, combines the bootstrapping procedure with the permutation test to mimic an overall F-test. Although it sounds like an ANOVA F-test to analyse more than two groups, this procedure maintains the Type I error level (familywise error rate) and delivers an acceptable level of statistical power while not relying on distribution assumptions. The key disadvantage of this procedure is that the OTG has not yet been included in SmartPLS. Researchers interested in the OTG approach can attempt it using the R code to generate results (available under the download section at https://www.pls-sem.net/downloads/advanced-issues-in-pls-sem-1/). Moreover, a researcher cannot assume that use of the OTG is sufficient, given that this procedure always demonstrates significant parameter differences, especially when the number of bootstraps runs increases. On top of that, OTG results do not provide a clear answer on the presence of specific differences between path coefficients across groups. Consequently, researchers must conduct all possible pairwise group comparisons while controlling the familywise error rate (Type I error) to avoid alpha inflation. This is easily controlled using either the *Bonferroni* correction or the *Šidák* correction (Hair et al., 2018). The Bonferroni correction can be applied with the formula alpha/m. For example, if researchers compare five groups, there would be *m*=10 comparisons, yielding a significance level of 0.05/10 (i.e., 0.005), instead of 0.05. On the other hand, the Šidák correction uses a different formula, that is, $1-(1-alpha)^{1/m}$. For example, for five groups with 10 comparisons, one would use a significance level of $1-(1-0.05)^{1/10}$ (i.e., 0.005116) instead of 0.05. Both corrections are plausible, but researchers should note that the Bonferroni correction has weaker statistical power than the Šidák correction in detecting pairwise group comparisons when there is an

exponential increase in the number of comparisons (see Hair et al., 2018). Readers can refer to some examples of using the OTG with the Šidák correction in papers published in tourism (see Ting et al. 2019) and retailing (Osakwe et al., 2020).

Figure 3: Guidelines for Running MGA in PLSPM
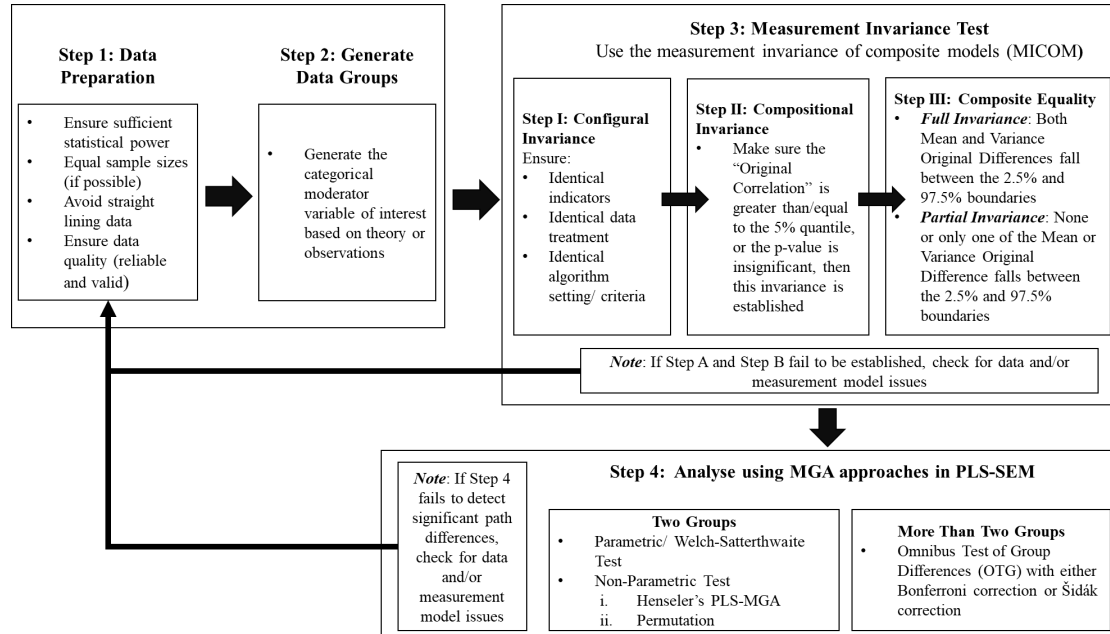
Case Study Illustration: MICOM and MGA in PLSPM

To illustrate the use of MGA in conjunction with the MICOM procedure in PLSPM, we used the model of ethnic food consumption intention at the touring destination by Ting et al. (2019). The model was built on the Theory of Reasoned Action (TRA) to investigate tourists' intention to consume Dayak food when they visit Sarawak, Malaysia. Following the work of Ting et al. (2019), the TRA model had a total of nine reflective indicators rated on a seven-point Likert scale. These measurements comprised the items for *attitude* towards consuming Dayak food, *subjective norms* towards consuming Dayak food, *perceived behavioural control* towards consuming Dayak food, and *intention* to consume Dayak food, which have been validated and published in Tourism Management. Instead of analysing the same research model from the Tourism Management manuscript, this case study examined a more parsimonious model using attitude, subjective norms, and intention to consume Dayak food with an aggregate dataset of 248 observations from two regions (i.e., Peninsular Malaysia and Singapore). This theoretical model is presented in Figure 4, where the research questions to be investigated were "*Do the tourists of Peninsular Malaysia and Singapore differ with regard to the effects of their (i) subjective norms on attitude, (ii) subjective norms on intention, and (iii) attitude on intention?*"

Figure 4: Ethnic (Dayak) Food Consumption Model in SmartPLS 3.3.2
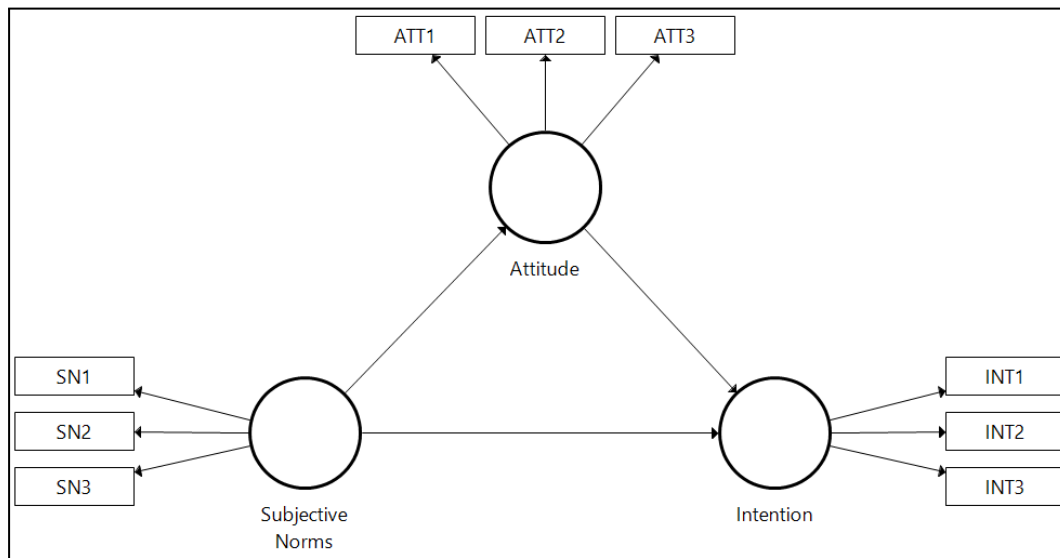
Based on the proposed research questions, the following three research hypotheses were formulated:

i.   Region moderates the relationship between Subjective Norms and Attitude.
ii.   Region moderates the relationship between Subjective Norms and Intention.
iii.   Region moderates the relationship between Attitude and Intention.

Step 1: Data Preparation

When engaging in MGA, researchers must first ascertain that the number of observations in each group meets the rule of thumb for minimum sample size. According to the power analysis results using G*Power, 64 observations per group were needed to detect R^2 values of about 0.25 at a significance level of 5% and a power level of 80%. Therefore, the group-specific sample sizes for both Peninsular Malaysia (n = 139) and Singapore (n = 109) were considered sufficiently large with approximately equal sample sizes. Notably, this dataset did not have issues with missing values or reliability and validity.

Step 2: Generate Data Groups

We started generating data groups in SmartPLS by first double-clicking on the inserted data (**Food Cross Border Study**), upon which the software opened a new tab (also known as data view). This data view provides basic overall information on the dataset (see Figure 5, Panel A). At this point, three icons or options were seen, namely **Add Data Group**, **Generate Data Groups**, and **Clear Data Groups**. To define groups, we clicked on the **Generate Data Groups** icon, which opened a new dialogue box (see Figure 5, Panel B). Next to **Name Prefix,** we specified a prefix, or initial name, for SmartPLS to use in naming the groups in the results. In this example, we renamed the default prefix to Group_. Next, under **Group Column,** we could specify

one or more grouping variables. Upon left-clicking on the box next to Group Column 0, a list of all variables in the dataset appeared, with each variable showing its number of unique values. For this example, we scrolled down and located the variable called "Region", which had two unique values (Peninsular Malaysia and Singapore). Then, we left-clicked on "Region" and closed the dialogue box by left-clicking on **OK** to return to data view. The data view now exhibited an additional tab labelled Data Groups, which typically offers information on group-specific sample sizes (see Figure 5, Panel C). Then, when the cursor was hovered to the right of the four groups, the **Delete** and **Edit** buttons appeared. We clicked on "**Edit**" and were subsequently led to a new dialogue box, "**Configure Data Group**" (see Figure 5, Panel D). We then renamed the groups. For instance, GROUP_Region(1.0) was renamed to "Peninsular Malaysia", and in a similar manner, GROUP_Region(2.0) was renamed to "Singapore". Researchers should note that we used only two groups for illustration to keep the discussion parsimonious. Finally, we closed the data view and returned to the modelling window that illustrated the model in Figure 4.

Figure 5: Panel A shows the data view in SmartPLS, Panel B shows the selection of data groups to be generated, Panel C shows the data view with data groups generated based on regions, and Panel D shows the renaming of groups.
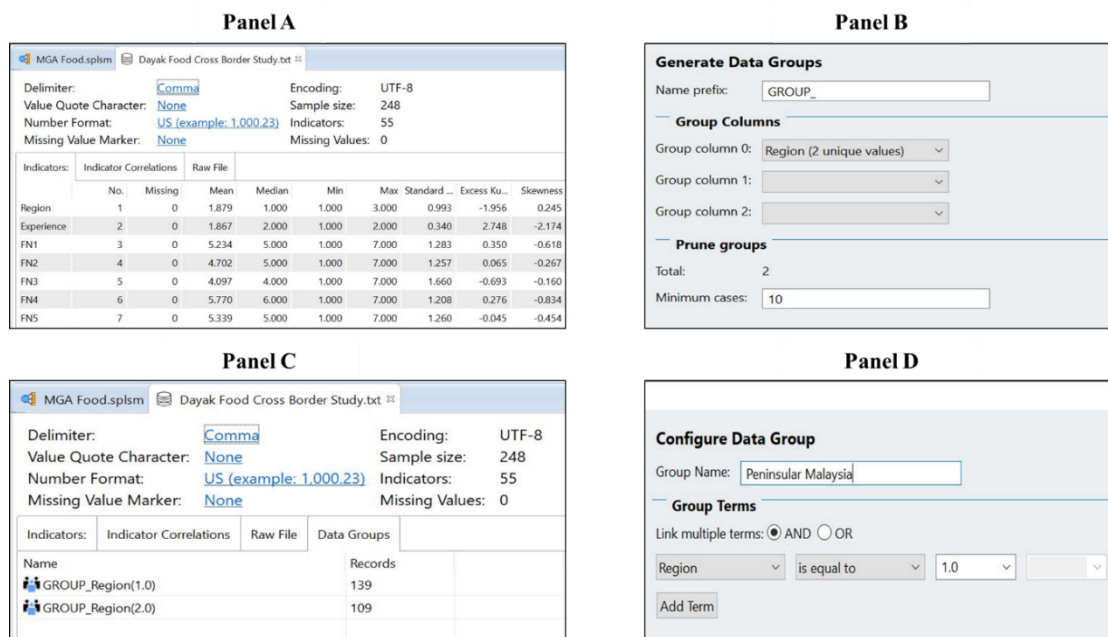
Step 3: Measurement Invariance Test using MICOM

To run the MICOM procedure, configural invariance (Step I) must be established. In doing so, we ensured: (i) the use of equal indicators in both groups when checking reliability and validity; (ii) similar data treatment (e.g., dealing with missing values using mean value replacement or case-wise deletion); and (iii) similar PLSPM algorithm settings (e.g., path weighting with a maximum of 300 iterations and a stop criterion of 10^{-7}). If all these criteria are taken care of, configural invariance is established. Notably, when running MICOM in SmartPLS, configural invariance (Step I) is automatically confirmed.

In Step II, we clicked on **Calculate → Permutation**, following which we saw a permutation dialogue box that allowed us to specify the groups to be compared. To do so, we selected Peninsular Malaysia from the menu next to Group A and selected Singapore for Group B. We then set a minimum of 1,000 (5,000 is also acceptable) permutations and two-tailed testing at a significance level of 0.05 (see Figure 6, Panel A). **If directional hypotheses** are involved, **the one-tailed test can be applied**. Next, we clicked on **Start Calculation,** allowing SmartPLS to estimate the model for each group separately.

In the results report, we hovered the cursor over **Quality Criteria → MICOM** procedure. The first tab, labelled Step 2 (also known as Step II in the MICOM assessment) (see Figure 6, Panel B), portrayed the results of the compositional invariance assessment. This result is typically slightly different from those presented because the permutation technique in MICOM uses a random procedure (like the bootstrapping resampling procedure). To assess the result, we compared the results of correlation c between the composite scores of the first and second group (column Original Correlations) with the 5% quantile, which revealed that the quantile was smaller than (or equal to) correlation c for all the constructs. This result was corroborated by the permutation's p-values that were larger than 0.05, indicating the correlation was not significantly lower than one. As shown in Figure 6 (Panel B), having fulfilled these criteria, the results suggested that compositional invariance was established. Thus, the results of Step 2 (or Step II) supported partial measurement invariance. We could then confidently compare standardised path coefficients across the groups through MGA in PLSPM.

Additionally, to check whether full measurement invariance was achieved, we clicked on the tab labelled Step 3 (or Step III) (see Figure 6, Panel C) to assess the composites' (constructs') equality of mean values and variances across groups. We checked the first column (mean original difference) and ensured that the value of each construct fell within the 95% confidence interval. This was done by comparing the mean original difference to the lower (2.5%) and upper (97.5%) boundaries shown in Columns Three and Four. If the mean original difference falls within the range of the lower and upper boundaries, then the first part of Step III has been met, thus providing initial evidence of invariance. Based on Figure 6 (Panel C), both the mean differences of Intention and Subjective Norms fell within the 95% confidence interval of the lower (2.5%) and upper (97.5%) boundaries while Attitude did not, indicating that there were no significant differences between the two groups' (Peninsular Malaysia vs Singapore) mean values for two of the three latent variables. For instance, for Intention, the original difference in the mean values of the latent variable scores was 0.136, which was within the lower boundary of -0.239 and the upper boundary of 0.242. The results in the fifth column "Permutation p-value", being higher than 0.05 for Intention and Subjective Norms but not for Attitude, further corroborated the findings. The next columns exhibited the results of composite variances, for which the interpretation is similar to that of mean differences. Again, not all the confidence intervals straddled the original value, and the p-values for Intention and Subjective Norms were clearly smaller than 0.05. Again, Attitude showed no significant difference in composite variances. Since the Step III results

concluded that not all the composite mean values and variances were equal, only partial measurement invariance was supported.

Figure 6: Panel A shows the permutation dialog box, Panel B shows the MICOM Step II results, and Panel C shows the MICOM Step III results
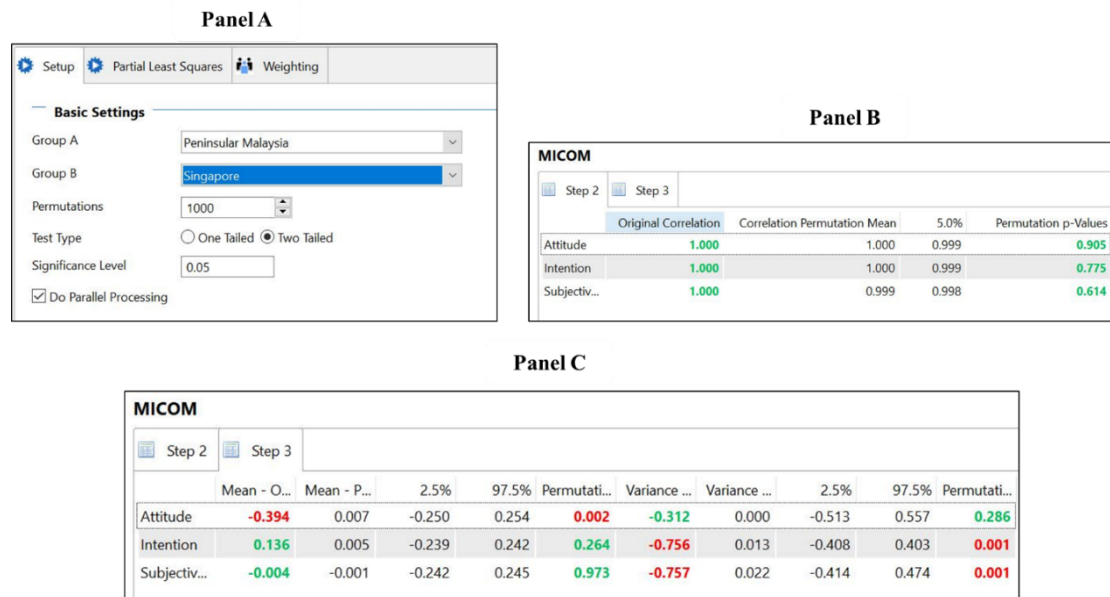
Step 4: Test of MGA Comparisons

Once measurement invariance was established (regardless of partial or full), we continued by examining group comparisons using MGA. We chose to first focus on the permutation test results generated earlier in Step III. Specifically, from the permutation test, we accessed the results by going to **Final Results → Path Coefficients** in the SmartPLS permutation output. The first two columns in Figure 7 exhibit the original path coefficients in Group 1 (Peninsular Malaysia) and Group 2 (Singapore), followed by their differences in the original data set and the permutation test. The results revealed that only one relationship was statistically different between Group 1 (Peninsular Malaysia) and Group 2 (Singapore), that is the relationship between subjective norms and attitude ($p < 0.05$).

Figure 7: Permutation Test Results in SmartPLS

To further analyse group-specific differences, we ran another MGA using SmartPLS. To do so, we clicked on **Calculate → Multi-Group Analysis (MGA)**. In the dialogue box (see Figure 8), we selected Peninsular Malaysia under Group A by ticking the box next to it. Next, we ticked on Singapore under Group B. While researchers can

manipulate the settings related to the PLS algorithm (e.g., bootstrapping procedure, missing value treatment, and variable weighting) at this point, we recommend the use of standard settings for all options, as suggested by Hair et al. (2017). Hence, we clicked on **Start Calculation**.

Figure 8: Multigroup Analysis Dialogue Box in SmartPLS

Once the result was generated, the report opened. We went to **Final Results → Path Coefficients** and selected the **PLS-MGA** tab to access the MGA results (see Figure 9, Panel A). Since PLS-MGA uses a one-tailed test, the p-values in the report showed whether the path coefficient was significantly larger in the first group (i.e. Peninsular Malaysia) than in the second group (i.e. Singapore). Alternatively, researchers can also use the '1-p' value to assess whether there is a significant difference in the other direction. The results indicated that only the relationship between subjective norms and attitude had a significant p-value ($p < 0.05$). Similar results were reported by the Parametric Test (see Figure 9, Panel B) and the Welch-Satterthwaite test (see Figure 9, Panel C). This implies that there was a significant difference across regions (Peninsular Malaysia vs Singapore) in the relationship between subjective norms and attitude. We also compared the bootstrapping result to assess the difference between the regions' path coefficients. Figure 9 (Panel D) shows that Peninsular Malaysia ($\beta = 0.730$) had a stronger path coefficient than Singapore ($\beta = 0.197$).

Figure 9: Panel A reports PLS-MGA results, Panel B reports the Parametric PLS Multigroup Test results, Panel C reports the Welch-Satterthwaite Test results, and Panel D reports bootstrapping results for path coefficient comparison

Conclusion and Future Studies

The assumption that all study samples are drawn from homogeneous populations may lead to incorrect conclusions. Therefore, researchers must routinely assess heterogeneity effects (i.e., size of firms, country, age, gender, income, etc.) through MGA. This editorial provides step-by-step guidelines on MGA procedures, which will boost beginners' confidence in using MGA and subsequently, improve the rigour of their research projects. Additionally, MGA assessments will also improve our knowledge of previously misunderstood theoretical relationships. In addition to the above guidelines, we recommend five practical considerations for researchers interested in MGA, as follows:

i. Those who want to mimic or work on covariance-based (or common factor) modelling when performing MGA in PLSPM can click on **Consistent PLS Algorithm** and look for **Consistent Permutation** and/or **Consistent Multi-Group Analysis**.

ii. If a researcher faces difficulties establishing partial invariance, especially using a reflective construct, he/she can remedy it by examining the outer loadings under Permutation Test or Multi-Group Test. The researcher needs to click on **Final Results → Outer Loadings**, then check for any significant difference in the outer loadings of constructs that fail to achieve compositional invariance. Subsequently, the researcher should remove the item with significant outer loading differences, because both sample groups could have perceived the item's meaning differently than intended. It is important to bear in mind that this remedy can only be used for reflective constructs because the nature of reflective indicators is interchangeable (see Hair et al., 2017).

iii. When performing the Permutation Test (either MICOM or group comparison of path coefficients), one should not rely completely on the p-value and neglect the confidence intervals. Usually, the p-value and confidence intervals are consistent. However, in certain conditions where the confidence interval is not symmetric and the mean value is on the edge of the confidence interval, p-values and confidence intervals may contradict one another. As p-values more easily obtain significant results due to large sample sizes (see Lin et al., 2013), it is often considered a "rough yardstick" to reject a model or hypothesis (Hahn, 2016). The confidence interval, however, offers better representation as it accounts for the skewness of the sampling distribution, whereas the p-value assumes a symmetric distribution. Therefore, we recommend that researchers interpret their results according to confidence intervals.

iv. The results of the Permutation and MGA tests (Henseler's MGA, Parametric, and Welch-Satterthwaite tests) will vary because these approaches rely on the bootstrapping procedure; however, their differences should not be substantial.

v. Based on a comparison of all the MGA techniques, we find that MGA results do not entail different outcomes across methods, even though we recommend using the permutation test in comparing two groups. This includes the use of a multimethod approach to provide additional confidence in the final results (see Hair et al., 2018).

This editorial focused on the comparison between two groups; however, there are times when a researcher encounters more than two groups. Hence, future research should provide step-by-step instructions on comparing more than two groups using both the OTG method and the Bonferonni/ Šidák correction. In addition, while there are calls for the use of consistent PLS (PLSc) (Dijkstra and Henseler, 2015), there are no known studies that have looked into the combination of this algorithm with MGA and the MICOM procedure. Subsequently, more research is needed to compare this algorithm with the various methods of MGA to better understand the results' stability or differences in each method. Lastly, future studies should attempt to combine both observed heterogeneity and unobserved heterogeneity when examining potential differences among groups using the segmentation approach. Using larger datasets or big data, both observed heterogeneity (MGA test) and unobserved heterogeneity (FIMIX, PLS-POS, and PLS-IRRS) can be complementary techniques because they

help to first uncover other areas that split data into several groups to identify and/or understand the differences among subpopulations (see Sarstedt et al., 2017).

References

Aguinis, H., Edwards, J. R., & Bradley, K. J., (2017), "Improving our understanding of moderation and mediation in strategic management research", *Organizational Research Methods*, vol. 20, no. 4, pp. 665-685.

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

Cheah, J. H., Roldán, J. L., Ciavolino, E., Ting, H., & Ramayah, T., (2020), "Sampling weight adjustments in partial least squares structural equation modelling: Guidelines and illustrations", *Total Quality Management & Business Excellence*, pp. 1-20.

Chin, W., Cheah, J. H., Liu, Y., Ting, H., Lim, X. J., & Cham, T. H., (2020), "Demystifying the role of causal-predictive modelling using partial least squares structural equation modelling in information systems research", *Industrial Management & Data Systems*, vol. 120, no. 12, pp. 2161-2209.

Chin, W. W., & 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 *Handbook of partial least squares* (pp. 171-193). Springer, Berlin, Heidelberg.

Cohen, J., (1988), *Statistical power analysis for the behavioural sciences* (2nd ed.), Lawrence Erlbaum: New Jersey.

Danks, N. P., Sharma, P. N., & Sarstedt, M., (2020), "Model selection uncertainty and multimodel inference in partial least squares structural equation modelling (PLS-SEM)", *Journal of Business Research*, vol. 113, pp. 13-24.

Dijkstra, T. K., & Henseler, J., (2015), "Consistent partial least squares path modeling", *MIS Quarterly*, vol. 39, no. 2, pp. 297-316.

Hahn, E. D., & Ang, S. H., (2017), "From the editors: New directions in the reporting of statistical results in the Journal of World Business", *Journal of World Business*, vol. 52, no. 2, pp. 125-126.

Hair Jr, J. F., (2020), "Next-generation prediction metrics for composite-based PLS-SEM", *Industrial Management & Data Systems*, Forthcoming.

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E., (2019a), *Multivariate data analysis*, (8th ed.), Cengage Learning: London.

Hair, J. F., Hult, G. T. M., Ringle, C. M., & Sarstedt, M., (2017b), A primer on Partial Least Squares Structural Equation Modeling (PLS-SEM), (2nd ed.), Sage Publication: Thousand Oaks.

Hair, J. F., Ringle, C. M., Gudergan, S. P., Fischer, A., Nitzl, C., & Menictas, C., (2019b), "Partial least squares structural equation modeling-based discrete choice modeling: An illustration in modeling retailer choice", *Business Research*, vol. 12, no. 1, pp. 115-142.

Hair, J. F., Risher, J. J., Sarstedt, M., & Ringle, C. M., (2019c), "When to use and how to report the results of PLS-SEM", *European Business Review*, vol. 31, no. 1, pp. 2-24.

Hair, J. F., Sarstedt, M., Ringle, C. M., & Gudergan, S. P., (2018), *Advanced issues in Partial Least Squares Structural Equation Modeling (PLS-SEM)*, Sage Publication: Thousand Oaks.

Henseler, J., Ringle, C. M., & 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., & Sarstedt, M., (2016), "Testing measurement invariance of composites using partial least squares", *International Marketing Review*, vol. 33, no. 3, pp. 405-431.

Henseler, J., Ringle, C. M., & Sinkovics, R. R., (2009), "The use of partial least squares path modeling in international marketing", In *New challenges to international marketing* (pp. 277-319): Emerald Group Publishing Limited.

Hult, G. T. M., Hair Jr, J. F., Proksch, D., Sarstedt, M., Pinkwart, A., & Ringle, C. M., (2018), "Addressing endogeneity in international marketing applications of partial least squares structural equation modeling", *Journal of International Marketing*, vol. 26, no. 3, pp. 1-21.

Hult, G. T. M., Ketchen, D. J., Griffith, D. A., Finnegan, C. A., Gonzalez-Padron, T., Harmancioglu, N., & Cavusgil, S. T., (2008), "Data equivalence in cross-cultural international business research: Assessment and guidelines", *Journal of International Business Studies*, vol. 39, no. 6, pp. 1027-1044.

Hwang, H., Sarstedt, M., Cheah, J. H., & Ringle, C. M., (2020), "A concept analysis of methodological research on composite-based structural equation modeling: Bridging PLSPM and GSCA", *Behaviormetrika*, vol. 47, no. 1, pp. 219-241.

Johnson, T., Kulesa, P., Cho, Y. I., & Shavitt, S., (2005), "The relation between culture and response styles: Evidence from 19 countries", *Journal of Cross-Cultural Psychology*, vol. 36, no. 2, pp. 264-277.

Jöreskog, K. G., & Wold, H. O. A., (1982), "The ML and PLS techniques for modeling with latent variables: Historical and comparative aspects", in Wold, H.O.A. and J€oreskog, K.G. (Eds), *Systems under indirect observation: Part I*, North-Holland, Amsterdam, pp. 263-270.

Keil, M., Tan, B. C., & Wei, K. K., (2000), "A cross-cultural study on escalation of commitment behaviour in software projects", *MIS Quarterly*, vol. 24, no. 2, pp. 299-325.

Kock, N., & Hadaya, P., (2018), "Minimum sample size estimation in PLS‑SEM: The inverse square root and gamma‑exponential methods", *Information Systems Journal*, vol. 28, no. 1, pp. 227-261.

Liengaard, B. D., Sharma, P. N., Hult, G. T. M., Jensen, M. B., Sarstedt, M., Hair, J. F., & Ringle, C. M., (2020), "Prediction: Coveted, yet forsaken? Introducing a cross-validated predictive ability test in partial least squares path modeling", *Decision Sciences*, Forthcoming.

Lin, M., Lucas Jr, H. C., & Shmueli, G., (2013), "Research commentary - too big to fail: Large samples and the p-value problem", *Information Systems Research*, vol. 24, no. 4, pp. 906-917.

Memon, M. A., Ting, H., Cheah, J. H., Ramayah, T., Chuah, F., & Cham, T. H., (2020), "Sample size for survey research: Review and recommendations", *Journal of Applied Structural Equation Modelling*, vol. 4, no. 2, pp. 1-20.

Osakwe, C. N., Ruiz, B., Amegbe, H., Chinje, N. B., Cheah, J. H., & Ramayah, T., (2020), "A multi-country study of bank reputation among customers in Africa: Key antecedents and consequences", *Journal of Retailing and Consumer Services*, vol. 56, pp. 102-182.

Picón-Berjoyo, A., Ruiz-Moreno, C., & Castro, I., (2016), "A mediating and multigroup analysis of customer loyalty", *European Management Journal*, vol. 34, no. 6, pp. 701-713.

Richter, N. F., Schubring, S., Hauff, S., Ringle, C. M., & Sarstedt, M., (2020), "When predictors of outcomes are necessary: Guidelines for the combined use of PLS-SEM and NCA", *Industrial Management & Data Systems*, Forthcoming.

Ringle, C. M., Wende, S., & Becker, J. M., (2015), *SmartPLS 3*, SmartPLS, Beonningstedt.

Sarstedt, M., & Cheah, J. H., (2019), "Partial least squares structural equation modeling using SmartPLS: A software review", *Journal of Marketing Analytics*, vol. 7, no. 3, pp. 196-202.

Sarstedt, M., & Mooi, E. A., (2019), *A concise guide to market research: The process, data, and methods using IBM SPSS statistics*, Springer: London.

Sarstedt, M., Henseler, J., & Ringle, C. M., (2011), "Multigroup analysis in Partial Least Squares (PLS) Path Modeling: Alternative methods and empirical results", Sarstedt, M., Schwaiger, M., & Taylor, C. R., (Ed.) Measurement and research methods in

international marketing (*Advances in International Marketing*, vol. 22), Emerald Group Publishing Limited, Bingley, pp. 195-218.

Sarstedt, M., & Ringle, C. M., (2010)., "Treating unobserved heterogeneity in PLS path modeling: A comparison of FIMIX-PLS with different data analysis strategies", *Journal of Applied Statistics*, vol. 37, no. 8, pp. 1299-1318.

Sarstedt, M., Ringle, C. M., Cheah, J. H., Ting, H., Moisescu, O. I., & Radomir, L., (2020), "Structural model robustness checks in PLS-SEM", *Tourism Economics*, vol. 26, no. 4, pp. 531-554.

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

Schlägel, C., & Sarstedt, M., (2016), "Assessing the measurement invariance of the four-dimensional cultural intelligence scale across countries: A composite model approach", *European Management Journal*, vol. 34, no. 6, pp. 633-649.

Sharma, P., Sarstedt, M., Shmueli, G., Kim, K. H., & Thiele, K. O., (2019a), "PLS-based model selection: The role of alternative explanations in information systems research", *Journal of the Association for Information Systems*, vol. 20, no. 4, p. 4.

Sharma, P. N., Shmueli, G., Sarstedt, M., Danks, N., & Ray, S., (2019b), "Prediction‑oriented model selection in partial least squares path modeling", *Decision Sciences*, Forthcoming.

Shmueli, G., Ray, S., Estrada, J. M. V., & Chatla, S. B., (2016), "The elephant in the room: Predictive performance of PLS models", *Journal of Business Research*, vol. 69, no. 10, pp. 4552-4564.

Shmueli, G., Sarstedt, M., Hair, J. F., Cheah, J. H., Ting, H., Vaithilingam, S., & Ringle, C. M., (2019), "Predictive model assessment in PLS-SEM: Guidelines for using PLSpredict", *European Journal of Marketing*, vol. 53, no. 11, pp. 2322-2347.

Ting, H., Fam, K. S., Hwa, J. C. J., Richard, J. E., & Xing, N., (2019), "Ethnic food consumption intention at the touring destination: The national and regional perspectives using multi-group analysis", *Tourism Management*, vol. 71, pp. 518-529

Welch, B. L., (1947), "The generalization of student's problem when several different population variances are involved", *Biometrika*, vol. 34, no. 1/2, pp. 28-35.

Wold, H. O. A., (1982), "Soft modeling: The basic design and some extensions", in J€oreskog, K.G. and Wold, H.O.A. (Eds), *Systems under Indirect Observations: Part II*, North-Holland, Amsterdam, pp. 1-54.