Between-Group Equivalence in Comparisons Using PLS: Results from Three Simulation Studies

Completed Research Paper

Miguel I. Aguirre-Urreta
Texas Tech University
Lubbock, TX
USA
miguel.aguirre-urreta@ttu.edu

Mikko Rönkkö
Aalto University
Espoo
Finland
mikko.ronkko@aalto.fi

Abstract

The examination of between-group differences in theoretical relationships of interest is central to the conduct of research in the organizational and behavioral sciences, including Information Systems research. One such approach for examining these differences relies on the conduct of multi-group PLS analyses, where each group of interest is modeled separately, and the structural results from each analysis are then compared. Though this approach has been employed in the empirical literature, the ability of separate analyses to obtain models that are equivalent from a measurement perspective – which is a pre-requisite to making any comparisons between them – has not been carefully studied. In this research we perform three simulation studies under varying invariant conditions that highlight the performance of PLS in this regard. Our results indicate that sampling variability plays a major role in whether equivalent results can be obtained, and showcase the conditions in which the technique performs best.

Keywords: Partial least squares, measurement invariance, simulation, between-group comparisons

Introduction

The process of theorizing, modeling, and testing between-group differences is central to Information Systems (IS) research, as it is in many other organizational and social sciences. Many of our most important research streams incorporate the examination of differences between subject groups; for example, technology acceptance research (e.g., the TAM and UTAUT models) has examined how the central relationships of the theory are different for men and women (e.g., Venkatesh & Morris, 2000; Venkatesh, Morris, Davis, & Davis, 2003). Traditionally, these between-group differences would have been estimated and assessed using techniques such as moderated OLS regression or ANOVA. More recently, however, researchers have turned to path modeling approaches to incorporate those in more complex research models with multiple independent, mediating, and dependent variables, which would be more difficult to model using traditional techniques. For this purpose, two distinct group of techniques have emerged, one based on simultaneous multi-group analysis using latent variables (e.g., covariance-based SEM) and another based on composites of observed variables, such as Partial Least Squares (PLS) (Qureshi & Compeau, 2009).

In the case of covariance-based SEM there is a rich methodological literature that has focused on how to specify multiple groups in a single analysis, impose constrains across them, and test between-group differences by effecting comparisons between models imposing different constraints in their specification.

Several different fit indexes, both absolute and relative, have been employed for this purpose. Central within this literature, and for our research as well, is the strong recognition that comparisons should only be made once it has been ascertained that the measures employed are equivalent across the groups under examination, as the measures should have the same meaning to respondents in all groups before any meaningful comparisons can be made (Vandenberg, 2002; Vandenberg & Lance, 2000). For the particular case of between-group comparisons of paths relating latent variables of interest in a research model, the literature recommends that configural (pattern of indicators relating to the latent variables of interest), metric (factor loadings), scalar (indicator intercepts) and factorial (factor variances) invariance be established before any structural parameters are compared across groups (Vandenberg & Lance, 2000, p. 56).

The literature on between-group comparisons using PLS, on the other hand, has largely focused on approaches to testing the statistical significance of path differences across groups. The most common of these approaches, and the one examined by Qureshi and Compeau (2009), involves a *t*-test with pooled standard errors. In this approach, model parameters are estimated for each group independently, and those are then compared and their significance assessed. This approach is the one most commonly employed in the IS literature (e.g., Ahuja & Thatcher, 2005; Chai, Das, & Rao, 2011; Keil et al., 2000; Sia et al., 2009, to name a few examples). Whether the composites obtained as proxies for the latent variables of interest in each separate group are comparable across those, however, has not been considered. For example, the work of Qureshi and Compeau (2009), on which this research is based, focused solely on the comparison of structural paths, with the data generated for such purpose being invariant by design in the data generation process (Qureshi & Compeau, 2009, p. 200). As we see in our first study, however, that models are invariant in the population does not necessarily lead to composites that are themselves invariant in any given sample, when research models are estimated separately for each group. Moreover, even if not explicitly considered in the multi-group PLS literature, the underlying rationale that measures and composites should be deemed comparable before any comparison are performed remains nonetheless valid (Millsap, 2011). In the particular case of PLS, the latent variables of interest are measured with indicators directly related to them, each of which is partly a function of the latent variable and some random measurement error. Those indicators are then combined into weighted composites that then stand in for the latent variables in the subsequent analyses. As such, ensuring that these composites are comparable across groups before conducting any comparisons of the structural parameters is a necessary condition, albeit one that has not been extensively considered in the literature.

This research seeks to examine the performance of PLS from a measurement equivalence standpoint by extending the research of Qureshi and Compeau (2009) in three different ways. In the first study we do so under the condition of full equivalence, which is the same in which the work of Qureshi and Compeau (2009) was conducted, even if the actual equivalence of the composites was not examined there. In the second study we examine violations of measurement invariance in the form of unequal loadings across the two groups. Finally, in our last study, we examine violations of structural invariance (which are typically of theoretical interest to researchers). Our results highlight a number of interesting findings about the performance of PLS under these conditions, which have not been previously examined from the perspective of measurement equivalence prior to conducting between-group comparisons.

Study 1: Full Equivalence

This research employed the models previously developed by Qureshi and Compeau (2009) in order to evaluate the performance of PLS in various scenarios. In particular, our research models have the same number of latent variables, pattern of relationships, and values for those parameters as in the original study. In that research, a base model and five alternative ones were used to compare the performance of PLS and covariance-based SEM on between-group comparisons of structural paths. Adoption of the same structural models thus allows for a direct extension of previous research. The structural portion of the models is depicted in Figure 1. Whereas the correlation between the exogenous latent variables was kept constant at 0.40 throughout, all other paths in the models were varied, as shown in Table 1, where the last column is the difference in paths across models (as compared to the Base model). All values are shown in standardized metric.

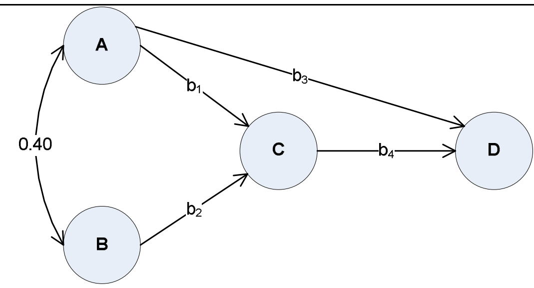

Figure 1. Structural Portion of Simulation Models

Table 1. Base Model and Comparison Models

Model	b_1	b_2	b_3	b_4	| diff |
Base (B)	.05	.20	.35	.60	-
C1	.10	.25	.40	.55	.05
C2	.20	.35	.50	.45	.15
C3	.30	.45	.60	.35	.25
C4	.40	.55	.70	.25	.35
C5	.50	.65	.80	.15	.45

In this first simulation study, full equivalence of both loadings and structural paths was examined. Each of the six models was compared against results from a different sample from the exact same model, and the correspondence between the two was calculated based on the results obtained from the PLS analyses. The following experimental conditions were varied, in a fully-crossed factorial design. First, the combined sample size over both groups was 100, 200, 300, or 400. Second, the number of indicators per latent variable was the same across all latent variables in the model, each being reflectively-measured with 3, 6, 9, or 12 indicators. Third, the loadings relating each latent variable to its indicators were the same in each experimental condition across all latent variables, at 0.70, 0.80, or 0.90, to represent varying degrees of indicator quality. Fourth, the two groups were of the same size (e.g., a 50/50 split of the total sample size), or the first group was three times larger than the second (e.g., a 75/25 split of the total sample size). Within each of these 96 conditions (4 x total sample size, 4 x number of indicators, 3 x loading strength, 2 x group sizes) we estimated 1,000 replications, with all data drawn from a multivariate normal distribution. All variables in the models, latent and manifest, had unit variance and zero means. No cross-loadings or correlated errors were included. The process was repeated for each of the models shown in Table 1.

The simulation was implemented in the R statistical environment by first specifying each model as described above. Data generation and analysis were conducted with the *matrixpls* package (Rönkkö, 2014), which employs the *simsem* package (Pornprasertmanit, Miller, & Schoemann, 2013) for data generation. All simulation code is available from the first author upon request. Throughout this research we used the Root Mean Square Deviation (RMSD), calculated between vectors of comparable estimates, such as vectors of weights or loadings across groups, to analyze our results. The formula for the RMSD is as follows:

$$RMSD = \sqrt{\frac{\sum_{i}^{N}(x_i - y_i)^2}{N}}$$

In this formulation, x_i and y_i represent individual elements of two comparable vectors, such as weights across two groups. Each element of the two vectors is compared to its corresponding counterpart from the other group, and their difference is squared, thus preserving the magnitude of the difference without concern for directionality, nor allowing differences of one sign for some pair of elements to cancel out with similar differences but of an opposite sign for another pair of elements, thus masking their existence. The sum of the squared differences is then averaged over the number of elements (*N*) in the compared vectors, which are assumed to be of equal length, as would be the case when the same items are used to measure latent variables across two groups of interest. The averaging does away with effects due to the length of the compared vectors. Finally, the average squared differences are transformed back, by means of taking their squared root, to the same metric as the original measurements, thus making their interpretation more straightforward. The RMSD is the square root of the Mean Squared Error (MSE), thus expressing the latter in the same metric as the original quantities being compared. When two comparable vectors are identical, which is the ideal scenario from the point of view of the researcher, the RMSD would be zero. RMSD values greater than zero indicate a deviation from the base case of full invariance. Though how much of a deviation would be deemed acceptable to still perform comparisons is a matter of judgement, the RMSD statistic helps inform researchers in this regard.

Results

Table 2 presents average RMSD for indicator weights within each condition (only the case of three indicators is included due to space limitations); each value shown is the average RMSD over one thousand replications. There are several effects of interest that can be observed from the results in Table 2. Recall that, in this simulation, the models from which the data were drawn for each group and then compared were fully invariant; that is, the population value of the RMSD statistic is zero in all these simulation runs.

TABLE 2. Average RMSD of Weights across Groups – Study 1 (3 indicators only)

Indicators	Loadings	Total Size	Groups	Model B	Model C1	Model C2	Model C3	Model C4	Model C5
3	0.70	100	Equal	0.264	0.224	0.151	0.111	0.088	0.074
			Unequal	0.288	0.250	0.183	0.144	0.112	0.092
		200	Equal	0.197	0.148	0.096	0.073	0.060	0.051
			Unequal	0.216	0.174	0.117	0.089	0.071	0.060
		300	Equal	0.155	0.114	0.075	0.058	0.048	0.041
			Unequal	0.179	0.138	0.090	0.070	0.057	0.048
		400	Equal	0.127	0.094	0.064	0.050	0.041	0.035
			Unequal	0.153	0.114	0.076	0.059	0.049	0.041
	0.80	100	Equal	0.204	0.155	0.097	0.069	0.053	0.043
			Unequal	0.218	0.178	0.123	0.088	0.066	0.052
		200	Equal	0.135	0.095	0.060	0.046	0.036	0.029
			Unequal	0.158	0.116	0.075	0.055	0.043	0.034
		300	Equal	0.100	0.069	0.047	0.036	0.029	0.023
			Unequal	0.119	0.091	0.057	0.044	0.034	0.028
		400	Equal	0.081	0.058	0.040	0.031	0.025	0.020
			Unequal	0.103	0.075	0.048	0.037	0.029	0.024
	0.90	100	Equal	0.143	0.096	0.052	0.037	0.028	0.021
			Unequal	0.151	0.117	0.072	0.047	0.034	0.025
		200	Equal	0.082	0.056	0.032	0.024	0.019	0.014
			Unequal	0.099	0.070	0.040	0.029	0.022	0.017
		300	Equal	0.058	0.039	0.026	0.020	0.015	0.011
			Unequal	0.072	0.054	0.031	0.023	0.018	0.013
		400	Equal	0.045	0.033	0.022	0.017	0.013	0.010
			Unequal	0.062	0.043	0.026	0.020	0.015	0.012

First, average RMSD between groups decreases as the total sample size increases. Second, average RMSD decreases as the strength of the loadings relating the indicators to their latent variables increases. Third, average RMSD is higher when the groups are of different size than when they are of equal size. Fourth, average RMSD also decreases as the strength of the structural relationships, evidenced by the changing paths across the different models, increases. Finally (not shown in Table 2), average RMSD decreases as the number of indicators per latent variable increases. All these results can be traced back to the effect of the different experimental factors on the sampling variability present in the data used by PLS to estimate the models in the two groups under comparison. It is a well-known characteristic of the PLS algorithm that it will take advantage of all sampling variability in order to obtain its estimates (Rönkkö & Evermann, 2013). Therefore, the larger the variability in the sample for each group, the more likely it is that PLS will estimate weights that will differ between the groups even if, in the population, both groups were equivalent by design, as different weights within each group would result in the local (that is, intra-group) maximization of explained variance in the endogenous composites.

In this case, stronger loadings and structural relationships result in indicators that are more strongly correlated in the population. As it is also well-known, the sampling variability in strong correlations is lower than in weaker ones. Both of these conditions, then, lead to reduced variability in the sample data used by PLS to estimate the models. Larger total samples have a similar effect, in that sampling variability is inversely related to the size of the obtained sample. As well, when the groups are unequal, which was operationalized in this simulation as the first group being three times larger than the second group, the sampling variability in the latter is naturally greater than in the former due to differences in sample sizes, leading to more variability in the weights obtained by PLS for the second group. Finally, RMSD decreases as the number of indicators per latent variable increases. While this is not directly related to sampling variability, an increasing number of indicators involved in the creation of each composite means that capitalization on sampling variability for any given indicator is less likely to affect the composite, as each indicator is a relatively smaller part of the composite when there are more indicators than when there are few.

While which particular value of the RMSD statistic is sufficient to support the assumption of equal measurement across groups that is required for the conduct of comparisons in structural paths is a matter of judgment, it is nonetheless clear that there are also several conditions shown in Table 2 where the use of PLS is not advisable as, even when the models under consideration were fully invariant across the groups in the population, the estimates for these scenarios show PLS will assign quite different weights to the items in the separate analyses for each group.

It is important for researchers, however, to not only rely on the average values across the simulation conditions to assess the scenarios in which PLS can be relied upon to provide meaningful comparisons. In addition, researchers should consider the likelihood of observing values much different from the means reported in Table 2, even if the groups were fully invariant in the population. As researchers collect data in a single study and not over thousands of replications, it is important to understand the likelihood that, for any given condition, PLS may result in groups with a high value of the RMSD statistic, which would cast doubts on the results of that particular research study. With that goal in mind, we also include here the standard deviations of the RMSD averages reported earlier. These can be found in Table 3 (only the case of three indicators is shown).

To make this issue more evident, we include here an ordered plot of the RMSD values – from highest to lowest – for the one thousand replications in the most problematic condition of those shown in Table 2 (that is, for the lowest values of all experimental factors examined here: 3 indicators per latent variable, total sample of 100 subjects, loadings at 0.70, and unequal group sizes) for each of the six models. The vertical lines depict the quintiles of the data. As Figure 2 shows, there is an important likelihood that researchers operating in this scenario would come across much higher values in any given study. It is clear from this figure that the models with stronger structural paths not only have a much lower likelihood to result in highly unacceptable RMSD values, but also the majority of the observations for those models occur within a relatively narrow range of values. Models with weaker structural paths, on the other hand, not only exhibit some very high values of the RMSD statistic, but also a great degree of variability in results, which further supports our arguments about the effects of sampling variability. For example, more than 40% of the results for the base model in this particular scenario exhibit RMSD values above 0.30. Researchers working with theoretical models where latent variables are only weakly correlated, should consider these results as an input to their sample size planning and choice of modeling approach.

TABLE 3. Standard Deviations of RMSD of Weights across Groups – Study 1 (3 indicators only)

Indicators	Loadings	Total Size	Groups	Model B	Model C1	Model C2	Model C3	Model C4	Model C5
3	0.70	100	Equal	0.130	0.114	0.074	0.045	0.026	0.022
			Unequal	0.131	0.119	0.095	0.079	0.054	0.038
		200	Equal	0.109	0.078	0.038	0.024	0.016	0.014
			Unequal	0.116	0.094	0.060	0.035	0.023	0.019
		300	Equal	0.091	0.060	0.026	0.017	0.013	0.011
			Unequal	0.103	0.077	0.037	0.027	0.016	0.013
		400	Equal	0.076	0.043	0.021	0.014	0.011	0.009
			Unequal	0.093	0.066	0.030	0.019	0.014	0.011
	0.80	100	Equal	0.135	0.109	0.056	0.029	0.018	0.012
			Unequal	0.135	0.118	0.079	0.051	0.029	0.021
		200	Equal	0.098	0.065	0.024	0.014	0.010	0.007
			Unequal	0.114	0.085	0.044	0.024	0.015	0.010
		300	Equal	0.075	0.041	0.015	0.010	0.007	0.006
			Unequal	0.085	0.068	0.026	0.017	0.010	0.007
		400	Equal	0.054	0.027	0.012	0.008	0.006	0.005
			Unequal	0.080	0.052	0.018	0.012	0.008	0.006
	0.90	100	Equal	0.137	0.096	0.037	0.017	0.009	0.006
			Unequal	0.136	0.111	0.071	0.035	0.015	0.009
		200	Equal	0.086	0.056	0.012	0.007	0.005	0.004
			Unequal	0.104	0.071	0.024	0.015	0.007	0.005
		300	Equal	0.060	0.028	0.009	0.006	0.004	0.003
			Unequal	0.075	0.055	0.016	0.008	0.005	0.003
		400	Equal	0.039	0.017	0.007	0.005	0.003	0.002
			Unequal	0.070	0.041	0.010	0.006	0.004	0.003

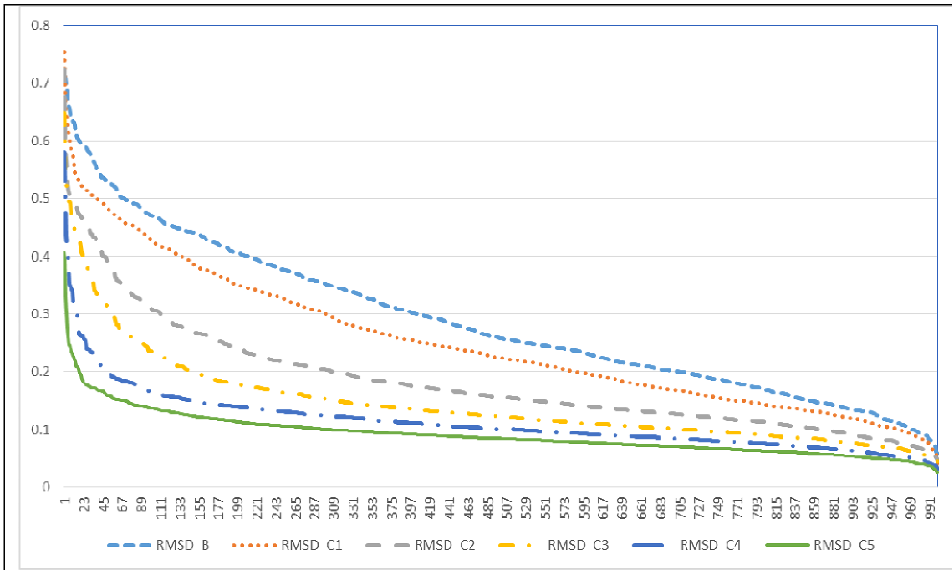

Figure 2. Plot of Ordered RMSD Values Across Models

To summarize, we examined in this first simulation study the performance of PLS when it comes to reaching equal weights across fully-invariant groups. Our results highlight the influence of sampling variability in the ability of the technique to do so and also show that, under conditions of larger sampling variability, the apparently acceptable average results may mask a high likelihood of coming across extreme statistics that should be taken into consideration by researchers when planning to conduct such studies. Our results also show that the structural portion of the research model has an important effect on these issues, and thus are not solely the outcome of concerns about the measurement aspects of the model. Note that here loadings were not examined as, given equal loadings in the population, equal weights across groups will result in equal loading estimates as well. In our next simulation study we focus on the question of invariant structural relationships that exhibit differences in loading strength across groups, and whether PLS is able to detect and estimate those differences.

Study 2: Lack of Measurement Invariance

In this second simulation study we sought to evaluate the performance of between-group comparisons in PLS when there is full structural invariance between the groups – that is when, at the population level, the structural portions of the research models are identical – but there are violations of measurement equivalence, in the form of unequal loadings. This is an important consideration, as absence of measurement equivalence is generally considered to represent a major obstacle to making any inferences about the equality – or not – of the structural relationships, which are typically those of prime interest in the conduct of research. This was also the major concern expressed by Carte and Russell (2003) with regards to PLS in their review of moderation research practices.

For this study we considered only Models B and C5 in Table 1, which are the two research models with the weakest and strongest structural paths, respectively. Doing so provides some insight into the effects that strength of structural relationships have, if any, on results when measurement equivalence is not obtained. The following changes were made from the setup of the first simulation study; every other aspect of the research design remained the same in this second study. We generated data using the same structural model in both groups – either Model B or Model C5 – under the following four different conditions: (1) in the first group all items loaded at 0.70, whereas in the second group they loaded at 0.50, (2) in the first group all items loaded at 0.70, whereas in the second group they loaded at 0.60, (3) in the first group all items loaded at 0.70, whereas in the second group they loaded at 0.80, and (4) in the first group all items loaded at 0.70, whereas in the second group they loaded at 0.90. This design, replicated across different sample sizes, number of indicators per latent variable and group size equality, as in the first study, allows us to examine both different degrees of discrepancy in the loadings as well as the effects of having, in the unequal group size conditions, the smaller group loading more or less strongly than the first group, on the results. As before, the RMSD statistic was used to analyze our findings.

Results

Table 4 presents results for RMSD across weight vectors for the two structural models and four discrepancy conditions under consideration here (only the scenarios with three or twelve indicators are shown). As was the case in the first study, variability in weight estimates across groups, evidenced by higher values of the RMSD statistic, is a function of the degree to which sampling variability is present in the experimental conditions and the number of indicators measuring the latent variables, which also contribute to more or less stable results (depending on whether the number of indicators is higher or lower in each condition). It is clear, across both structural models, that the results with more variability in the estimates are those where the second group had a lower value (e.g., Conditions 1 and 2), particularly when the general loading level was relatively weaker (that is, in Condition 1, with loadings of 0.7 for the first group and only 0.50 for the second group). This was further exacerbated by unequal group sizes, as the smaller size and weaker loadings would reinforce the effect. Therefore, the pattern of results with respect to the presence of sampling variability and number of indicators per latent variable, and their joint effect in the variability of weight estimates across groups, observed in the first study is validated by these results as well.

TABLE 4. Average RMSD of Weights across Groups – Models B and C5 (3 and 12 indicators only)

Indic.	Total Size	Groups	Model B				Model C5			
			Cond. 1 (-.2)	Cond. 2 (-.1)	Cond. 3 (+.1)	Cond. 4 (+.2)	Cond. 1 (-.2)	Cond. 2 (-.1)	Cond. 3 (+.1)	Cond. 4 (+.2)
3	100	Equal	0.368	0.307	0.244	0.220	0.157	0.105	0.067	0.076
		Unequal	0.396	0.341	0.247	0.218	0.221	0.145	0.068	0.071
	200	Equal	0.292	0.235	0.173	0.163	0.114	0.073	0.050	0.065
		Unequal	0.331	0.268	0.184	0.158	0.151	0.095	0.050	0.062
	300	Equal	0.247	0.192	0.134	0.129	0.099	0.062	0.043	0.061
		Unequal	0.288	0.226	0.142	0.124	0.124	0.077	0.044	0.060
	400	Equal	0.220	0.162	0.113	0.111	0.090	0.055	0.040	0.060
		Unequal	0.262	0.197	0.122	0.109	0.110	0.067	0.040	0.058
12	100	Equal	0.125	0.103	0.080	0.075	0.053	0.032	0.021	0.027
		Unequal	0.133	0.110	0.080	0.072	0.064	0.040	0.022	0.026
	200	Equal	0.099	0.076	0.054	0.053	0.045	0.025	0.017	0.025
		Unequal	0.112	0.087	0.059	0.054	0.051	0.029	0.017	0.024
	300	Equal	0.083	0.060	0.041	0.042	0.041	0.022	0.016	0.024
		Unequal	0.099	0.073	0.046	0.043	0.046	0.026	0.015	0.024
	400	Equal	0.073	0.050	0.034	0.036	0.040	0.021	0.015	0.024
		Unequal	0.090	0.064	0.039	0.037	0.044	0.023	0.015	0.023

It is also evident from Table 4 that, for both structural models, the RMSD of the differences between weight vectors in each condition converges to zero as the simulation conditions improve, from a sampling standpoint. That is, as the sample size increases, the loadings are stronger, there are more indicators measuring each latent variable, and the groups are equal. It is also clear that this convergence occurs at different rates for the two structural models, which also differ in their overall levels of estimate variability (e.g., RMSD) to begin with. In particular, results from Model C5 exhibit much lower levels overall, and otherwise low RMSD values for many of the scenarios included here, which points to the effects of stronger structural paths on the results – also commensurate with our findings from the first study. On the other hand, it would seem puzzling that, given differences in the measurement portions of the models between each of the two groups, as discussed before, the weights would converge across them; that is, that models with different loadings for the same items would result in similar weights given to those items across the two groups.

To see why this is the case here, it is important to remember that between-group comparisons in PLS are conducted by separately estimating each model in different PLS analyses, and then comparing the results from those. The weights assigned by PLS to each item capture how important that item is to the composite of which it is part, which is in turn a proxy for a latent variable of interest. Items that are more heavily weighted are those that contribute more, based on the postulated research model, to explaining variance on the endogenous composites. Conversely, items that are less important in this respect would be assigned a lower weight by the algorithm. These weights, however, are estimated separately for each group, as they are the result of separate PLS runs. That is, the algorithm has no knowledge of what the results are for the other group when it estimates the weights in an analysis. Weights, therefore, represent the relative importance of each individual item with regards to the composite of which it is part.

In the particular simulation design here, all items measuring a latent variable have the same loading within each of the conditions examined. That is, each item represents its respective latent variable as well as the other items loading on the same latent variable and it would thus be the case, aside from sampling variability, that all items measuring a latent variable would be assigned the same weight by the PLS algorithm. While all items measuring the same latent variable receive the same weights, the weights can be different between latent variables if the strength of the loadings varies; a latent variable with weakly loading indicators receives higher weights than a latent variable with strongly loading items. This occurs because the weights are scaled to produce a standardized composite and the variance of the raw composite depends on the indicator correlations. More specifically, three items loading at 0.70 on a latent variable would be assigned, at the limit, the same relative weights of 0.17 by the PLS iterative process. Similarly, three items loading on a latent variable at 0.50 would also be assigned, at the limit, the same weights by the algorithm, but in this scenario the weight would be 0.22 (note that this is representative of our Condition 3 above).

One necessary, but not sufficient, condition for conducting between-group comparisons in PLS, then, is that weights for the same items are similar across the groups. For example, if an item had a weight in one group twice as large as the weight, for the same item, in another group, that would imply that, in the first group, the item is a much more important determinant of the composite representing the latent variable of interest than in the second group. To the extent that a latent variable was measured with the same items on the two groups, if the relative importance of the indicators within each group is the same, then the weights estimated by PLS will converge to the same values, with the caveat noted before about sampling variability and number of indicators per latent variable. When compared across these conditions, differences in weights across groups, evidenced by high values of the RMSD statistic, are indicative of a lack of between-group invariance with respect to the relative importance of each item within its own measure, which is problematic for the conduct of comparisons in structural paths between groups (Carte & Russell, 2003).

This does not necessarily imply, however, that the loadings relating each item to its latent variable will be the same across the two groups. As shown in the example above, identical weights can occur even if the loadings relating each item to its latent variable are different across the groups, as weights and loadings are representing different aspects of the measurement model. In particular, the loading estimates obtained by PLS are determined by the weights estimated by the algorithm and the true values of the loadings in the population – recall that loadings in PLS are equal to the correlation of an item with a composite of which the item itself is part; the population value of the loadings determine how much each item is related to the other items in the composite, whereas the weight determines the importance of the focal item within the composite. Therefore, after the equivalence of the item weights across the groups has been established, it is necessary to examine the loading estimates obtained from the PLS analyses. Even if those are known to be biased when compared to their population counterparts (Aguirre-Urreta, Marakas, & Ellis, 2013), the issue here is not so much individual accuracy in their estimation, but rather equivalence of the estimates – albeit biased – across the two groups of interest.

Table 5 (only the scenarios with three or twelve indicators are shown) shows RMSD values for the comparison of the vectors of loadings estimated by PLS for the two groups. Recall that, by design, the difference between the loadings should be 0.10 in Conditions 2 and 3 (where the loadings were 0.70/0.60 and 0.70/0.80 across the two groups, respectively) and 0.20 in Conditions 1 and 4 (where the loadings were 0.70/0.50 and 0.70/0.90 across the two groups, respectively). There are several different effects that can be seen in these results.

First, the RMSD average tends to approximate its population value as the sample size and number of indicators per latent variable increase. This is most evident by comparing values for the same sample size across number of indicators, and for the same number of indicators across different sample sizes. In some cases it may appear that cells with lower sample sizes show closer estimates to the population difference than with larger samples; for an example, see the results for Model C5, Condition 4, 12 indicators per latent variable as the total sample size increases from 100 to 400. While this may appear to go against expectations based on statistics and the "*consistency-at-large*" theorem underlying the PLS algorithm (Wold, 1982), this not actually the case. Rather, smaller samples lead to more sampling variability, which features into PLS estimates. As the total sample size increases, estimates become closer to what they would be for the experimental conditions tested here when evaluated with an infinite sample (Aguirre-Urreta et al., 2013).

Second, though increasing total sample size and number of indicators per latent variable lead to the estimates converging on the population value of the difference between loadings, this approximation is clearly faster for some conditions than for others. In particular, all Conditions under Model B show a much greater beginning discrepancy in RMSD against the expected population values than those for Model C, though the differences all but disappear for the most beneficial scenarios (e.g., large samples, high number of indicators, etc.). As before, we attribute this to the strength of the structural relationships, which lead to the indicators from Model B to be, in general, much weakly correlated across latent variables than those for Model C5. Given the relationship between correlation size and sampling variability previously discussed, models with stronger structural paths, even at the lowest value of our experimental conditions, start reasonably close to the actual population values for the discrepancies. On the other hand, results from Model B exhibit quite a severe bias in many cases; for example, the RMSD for 3 indicators, total sample of 100, and unequal groups under Condition 1 is more than twice as large as the actual population discrepancy.

TABLE 5. Average RMSD of Loadings across Groups – Models B and C5 (3 and 12 indicators only)

Indic.	Total Size	Groups	Model B				Model C5			
			Cond. 1 (-.2)	Cond. 2 (-.1)	Cond. 3 (+.1)	Cond. 4 (+.2)	Cond. 1 (-.2)	Cond. 2 (-.1)	Cond. 3 (+.1)	Cond. 4 (+.2)
3	100	Equal	0.371	0.267	0.185	0.209	0.194	0.121	0.091	0.138
		Unequal	0.448	0.334	0.181	0.190	0.281	0.173	0.093	0.132
	200	Equal	0.289	0.194	0.132	0.172	0.148	0.090	0.076	0.130
		Unequal	0.360	0.244	0.130	0.155	0.190	0.112	0.076	0.127
	300	Equal	0.246	0.158	0.106	0.153	0.134	0.079	0.070	0.127
		Unequal	0.306	0.200	0.103	0.141	0.160	0.094	0.070	0.125
	400	Equal	0.221	0.133	0.094	0.144	0.126	0.073	0.068	0.126
		Unequal	0.280	0.173	0.092	0.136	0.146	0.084	0.068	0.125
12	100	Equal	0.276	0.189	0.152	0.218	0.221	0.145	0.122	0.194
		Unequal	0.345	0.242	0.153	0.206	0.258	0.174	0.125	0.190
	200	Equal	0.231	0.142	0.118	0.196	0.197	0.119	0.105	0.186
		Unequal	0.273	0.175	0.118	0.189	0.215	0.134	0.106	0.184
	300	Equal	0.210	0.122	0.106	0.189	0.188	0.108	0.100	0.184
		Unequal	0.245	0.150	0.105	0.185	0.202	0.121	0.099	0.181
	400	Equal	0.199	0.112	0.101	0.186	0.184	0.103	0.097	0.182
		Unequal	0.229	0.134	0.101	0.183	0.194	0.113	0.097	0.181

Third, even though the discrepancies between loadings are the same for Conditions 2 and 3 (e.g., 0.10), on the one hand, and for Conditions 1 and 4 (e.g., 0.20), on the other, both the initial values for the discrepancies and the rate at which those approximate their population values are different, for the case of Model B; this effect is not as evident for Model C5. In particular, the results for Conditions 1 and 2, where general loading level was lower than for conditions 3 and 4 – 0.70/0.60 and 0.70/0.50 compared to 0.70/0.80 and 0.70/0.90 – exhibit higher initial discrepancies, and these differences largely remain across number of indicators when the total sample is on the smaller side. This is also consistent with generally lower loadings in these scenarios resulting in higher sampling variability, which leads to more variance in the estimated loadings. As the total sample size increases the differences are less evident.

Finally, there is a marked group size effect, in that scenarios where there were group size differences – that is, where the total sample was split 75/25 between the two groups – show higher RMSD values when compared to the same experimental conditions but with equal sample sizes in the two groups. This effect, however, appears only when the group with the smaller sample is also the one with the lower loadings (Conditions 1 and 2) but not when the smaller group has higher loadings (Conditions 3 and 4). When the smaller group also has the lower loadings, sampling variability is much higher than for the larger group, which leads to the effects previously discussed. On the other hand, when the smaller group has higher loadings, the two effects appear to cancel out to some extent, resulting in much less noticeable differences. The effect is stronger for Model B than for Model C5, but is nonetheless present in both cases.

To summarize, this second study shows that several different factors affect the degree to which estimates from separate PLS runs for each group converge on the expected values based on the populations from which the data are draw. It also shows that equality of weights across groups is a necessary but not sufficient condition to the determination of measurement equivalence that must be supported before path comparisons can be conducted. Equality of weights across groups indicates that, within each latent variable, the items measuring it have the same relative importance when it comes to the composite that represents the latent variable in the PLS analyses. While the same item having a different relative importance for the creation of the composites across groups would be an indicator of an underlying difference in the importance of that item – and thus a violation of measurement equivalence – the equality of weights does not necessarily imply equality of loadings. Rather, once it has been established that weights are comparable across groups, as indicated by RMSD values close to zero, then loadings should be examined, as it is possible – and shown here with results from our simulations –for items to have the same relative importance when it comes to determining the composites, while still have different loading estimates. Our results show that, as simulation conditions improve, the values of the RMSD statistic approximate the expected difference between the loadings based on the simulation design. The range of values for the experimental factors where the RMSD values for both weights and loadings convergence on the expected values provides a boundary for the applicability of PLS for effecting these comparisons.

Study 3: Lack of Structural Invariance

Our last study focuses on scenarios where there is measurement equivalence but there are differences in the structural paths across groups. Together with our first simulation study, these two analyses represent the most desirable scenarios by researchers, where the lack of measurement differences across the groups allows for the examination of differences in the structural portion of the model which, from a theoretical standpoint, are most interesting. Similar to past research by Qureshi and Compeau (2009), on which our population models are based, we here compare the equivalence of weights and loadings when the measurement portions of the research models are identical, but not so in their structural aspects. In particular, conducting such an analysis is a precondition to the tests by Qureshi and Compeau (2009), who relied on measurement invariance by design in their simulation but did not assess the extent to which PLS was able to achieve it for the various examined conditions.

In this last simulation we compare the Base model against each of the other five models created by Qureshi and Compeau (2009). Within each of these comparisons we vary, as in the first simulation, the number of items measuring each latent variable (3, 6, 9 or 12), the total sample size (100, 200, 300 or 400), the relative size of the two groups (50/50 or 75/25 split) and the strength of the loadings relating the items to the indicators (0.70, 0.80, or 0.90). All measurement aspects are identical across the two models under comparison. In all the simulations, the first group is drawn the Base model (B), and the second group is drawn from the corresponding Comparison model (e.g., C1, C2, C3, C4 or C5). Also, as before, all samples are drawn from a multivariate normal distribution, with one thousand replications in each condition.

Results

We begin by examining the results with regards to weight estimates across the two groups, which are reported in Table 6 (only the case of three indicators is shown). In each of the comparisons reported here, the base model B is contrasted again alternative models, which are increasingly more different in their values for the structural paths, though the measurement portion of the models – that is, their loadings – remains identical across them. As discussed before, for identical loadings across groups, models with stronger structural paths will result in stronger correlations among the indicators, which in turn affects the sampling variability present, for any given sample size. As a result, the base model, which has the weakest structural paths and thus the greatest sampling variability of those included here, is compared against alternative models with increasingly stronger paths and thus increasingly smaller sampling variability.

This effect can be seen in Table 6 by comparing RMSD estimates for any given condition across the columns in the table. As one moves from the comparison between B and C1 and into the stronger alternative models, it is evident that the RMSD decreases markedly. This occurs as a result of reduction in overall sampling variability due to the alternative models, as the base model remains the same across all the simulations. The first effect of interest in these results, then, is that comparisons where the structural paths of the models are stronger are more likely to result in acceptable RMSD estimates that will allow the researchers to conduct those comparisons than those scenarios where one or both of the models have weaker paths, which results in much greater variability in the estimates and makes it more difficult to obtain similar composites across the two models, which is required before comparisons of the relationships among them can be performed.

TABLE 6. Average RMSD of Weights across Groups – Study 3 (3 indicators only)

Indicators	Loadings	Total Size	Groups	B vs. C1	B vs. C2	B vs. C3	B vs. C4	B vs. C5
3	0.70	100	Equal	0.247	0.216	0.202	0.197	0.192
			Unequal	0.271	0.230	0.209	0.190	0.178
		200	Equal	0.175	0.154	0.146	0.142	0.141
			Unequal	0.190	0.152	0.135	0.126	0.120
		300	Equal	0.134	0.120	0.114	0.112	0.110
			Unequal	0.151	0.116	0.104	0.097	0.092
		400	Equal	0.111	0.101	0.095	0.093	0.092
			Unequal	0.124	0.097	0.086	0.081	0.077
	0.80	100	Equal	0.178	0.153	0.144	0.140	0.136
			Unequal	0.198	0.161	0.139	0.127	0.120
		200	Equal	0.117	0.101	0.097	0.094	0.092
			Unequal	0.129	0.098	0.085	0.079	0.075
		300	Equal	0.085	0.076	0.072	0.070	0.070
			Unequal	0.098	0.073	0.064	0.059	0.056
		400	Equal	0.069	0.062	0.059	0.057	0.057
			Unequal	0.081	0.062	0.055	0.051	0.049
	0.90	100	Equal	0.122	0.103	0.098	0.095	0.094
			Unequal	0.133	0.100	0.083	0.076	0.071
		200	Equal	0.069	0.059	0.057	0.055	0.054
			Unequal	0.078	0.055	0.048	0.044	0.042
		300	Equal	0.049	0.043	0.041	0.040	0.040
			Unequal	0.058	0.041	0.036	0.033	0.031
		400	Equal	0.039	0.035	0.033	0.033	0.032
			Unequal	0.047	0.034	0.030	0.028	0.026

This also leads to a puzzling finding with regards to the effects that equality of group sizes has on the RMSD estimates for the weights, which can be seen more clearly in the lower levels of the experimental factors employed here (e.g., 3 indicators per latent variable, 0.70 loadings) but that remains throughout the results shown in Table 6. In the comparisons between the base model and those closer to it in terms of structural path differences (e.g., B vs. C1, B vs. C2) the effects of group equality are consistent with those described earlier. As the group with the stronger paths, and thus the relatively lower sampling variability, is smaller than the base model, with the larger sampling variability, the weight estimates are more likely to differ across them even if, by design, they are identical in the population. This effect, however, reverses as the groups are more dissimilar (e.g., B vs. C4, B vs. C5) which we interpret as follows, aided by an examination of the standard deviation of the estimates (not reported here).

Consider the scenario with only 3 indicators measuring each latent variable, loadings at 0.70 throughout, and a total sample size of 100. In the comparison between the B and C1 models the RMSD estimates for the difference between the weights are .247 and .271 for the case of equal (50/50 split of the total sample size) and unequal (75/25 split of the total sample size) groups, respectively. In this first case, we are comparing a sample of 50 drawn from the B model and another sample of 50 drawn from C1 (equal groups) with a sample of 75 drawn from the B model and a sample of 25 drawn from the C1 model. Though the sample drawn from model B is larger than in the first case, which would be beneficial as this model has the most sampling variability of those examined, the sample of only 25 drawn from C1, which is close to B in terms of correlation strength and thus sampling variability, leads to, on average, very variable estimates of the weights across the groups, which is reflected in a higher RMSD.

On the other hand, consider the estimates for the same conditions but when comparing the B and C5 models, which are .192 and .178 for the equal and unequal groups, respectively. Here, the comparison is between a sample of 50 from B and 50 from C5 for the equal groups condition, and a sample of 75 from B and 25 from C5, for the unequal groups condition. Though both the total sample size (100) as well as the relative size of the two groups (50/50 or 75/25) are the same as before, the unequal groups condition has a lower RMSD value than the equal groups condition, which is indicative of an average lower variability in the weight estimates. This occurs because the overall sampling variability is lower with a larger sample from the more variable model (that is, model B) and a smaller sample from the less variable model (model C5), than when the two samples are of the same size. The advantage in variability obtained from a larger sample from B outweighs the decrease resulting from a smaller sample from C5. This is because samples from the C5 model, unlike those from C1, exhibit much lower variability to begin with, as a result of the stronger structural paths[1].

Other effects shown in the results reported in Table 6 are consistent with expectations based on sampling theory. The values of the RMSD statistic approximate their population value of zero – given the design of this simulation study – as the number of indicators per latent variable increases, as the total size of the samples increases, and as the strength of the loadings relating the indicators to the latent variable increases. All these are consistent with findings from the previous two simulations. Taken together, these effects outline the boundaries of the result space in which the PLS algorithm, even when the measurement portions of the two groups under consideration are equivalent across groups, is able to obtain estimates that are similar enough to support the conduct of comparisons between the structural paths, which are of main interest in this kind of research design. We now turn to an examination of the behavior of the loadings for these same models and conditions.

TABLE 7. Average RMSD of Loadings across Groups – Study 3 (3 indicators only)

Indicators	Loadings	Total Size	Groups	B vs. C1	B vs. C2	B vs. C3	B vs. C4	B vs. C5
3	0.70	100	Equal	0.186	0.167	0.158	0.155	0.151
			Unequal	0.223	0.187	0.173	0.159	0.149
		200	Equal	0.124	0.111	0.107	0.105	0.105
			Unequal	0.139	0.112	0.101	0.098	0.094
		300	Equal	0.090	0.082	0.079	0.079	0.078
			Unequal	0.105	0.082	0.076	0.073	0.071
		400	Equal	0.073	0.068	0.066	0.065	0.065
			Unequal	0.085	0.068	0.063	0.061	0.059
	0.80	100	Equal	0.111	0.098	0.095	0.093	0.092
			Unequal	0.134	0.110	0.096	0.091	0.087
		200	Equal	0.066	0.059	0.058	0.057	0.057
			Unequal	0.076	0.060	0.056	0.054	0.052
		300	Equal	0.046	0.043	0.042	0.042	0.042
			Unequal	0.055	0.044	0.042	0.040	0.040
		400	Equal	0.036	0.034	0.034	0.034	0.033
			Unequal	0.045	0.037	0.035	0.034	0.034
	0.90	100	Equal	0.049	0.044	0.043	0.043	0.043
			Unequal	0.058	0.047	0.042	0.039	0.039
		200	Equal	0.027	0.024	0.024	0.024	0.024
			Unequal	0.031	0.026	0.024	0.024	0.024
		300	Equal	0.019	0.018	0.018	0.018	0.018
			Unequal	0.023	0.019	0.018	0.018	0.018
		400	Equal	0.015	0.015	0.014	0.014	0.014
			Unequal	0.019	0.016	0.016	0.016	0.015

Results for discrepancies between loading estimates across groups are reported in Table 7 (only the case of three indicators is shown), and are consistent with the effects just discussed for the weights. In particular, the discrepancy between loading estimates across groups, as evidenced by the values of the RMSD statistic, decreases as the base model is compared with models with increasingly stronger – and

---

[1] We verified this to be the case by rerunning this comparison (results not reported due to space limitations) where the B model only had 25% of the total sample with the C5 model the 75% remaining. In this case, the RMSD for the unequal groups was much higher than that presented here as well as higher than that for the equal groups.

different – structural paths. Second, the same reversal of the discrepancies between equal and unequal groups that was previously discussed, due to a compensation effect between reduction in sampling variability from the base model due to a larger sample and a stronger, if smaller, sample from the less variable group, was observed here as well. Third, also as before, the discrepancies approximate their population value of zero – given the simulation design with equivalent measurement across groups – as the total size of the sample, the number of indicators per latent variable, and the strength of the loadings increase, all of which are to be expected in light of the known behavior of statistical sampling.

To summarize, this third simulation study examined the performance of PLS for between-group comparisons in those scenarios where the measurement portions of the models are equivalent across groups, but those differ in the population values of the structural paths. Results for the discrepancies between weight estimates and between loading estimates across groups indicate that PLS is able to achieve stable and comparable estimates only under certain conditions. While the particular value of the RMSD statistic with which researchers would feel comfortable conducting between-group comparisons of the structural paths is a matter of judgment, results from this simulation show that the technique only approximates the population design of equal measurement for the higher levels of the experimental factors examined here. For the lower level of those, on the other hand, it is quite evident PLS is not an appropriate technique and caution is advised.

Discussion, Limitations and Conclusions

In this research we examined the performance of PLS weights and loadings in multi-group analyses under three different conditions: population models that are fully equivalent between two-groups (Study 1), population models with identical structural parameters but which differ in the measurement characteristics of the indicators, in the form of invariant loadings (Study 2), and population models that were fully equivalent from a measurement perspective but which exhibited differences in the structural relationships among latent variables (Study 3) – the latter represent the scenario that is most desirable by researchers, where between-group differences in paths can be attributed to true underlying differences and not to differential measurement. Specifically, we were interested in examining the technique in these scenarios from the perspective of equivalence – or not – of measurement across both groups, which has not received attention in the literature. Table 8 summarizes the design of each of the three studies.

TABLE 8. Summary of Design Conditions

Study 1: Full Equivalence
Sample size: 100, 200, 300, 400 Number of indicators per construct: 3, 6, 9, 12 Loading strength: 0.7, 0.8, 0.9 Group composition: 50%/50%, 75%/25%
Study 2: Lack of Measurement Invariance
Sample size: 100, 200, 300, 400 Number of indicators per construct: 3, 6, 9, 12 Base loading strength: 0.7 Loading differences: -0.2, -0.1, +0.1, +0.2 Group composition: 50%/50%, 75%/25%
Study 3: Lack of Structural Invariance
Sample size: 100, 200, 300, 400 Number of indicators per construct: 3, 6, 9, 12 Loading strength: 0.7, 0.8, 0.9 Group composition: 50%/50%, 75%/25%

Our results highlight several interesting findings. First, even when the models are fully equivalent across groups, multi-group analyses only lead to comparable results – those where measurement of the latent variables of interest is similar across the two groups – when sampling variability is relatively more limited, which leads to both groups in the analysis to converge on relatively similar measurement models. It is important to note that the strength of the structural relationships play a role here as well; results are better when the latent variables are more strongly related. Second, when measurement is not invariant across groups in the population, results from PLS runs only lead to relatively accurate estimates of the measurement differences also when sampling variability is relatively limited. Findings from our second study also point to the need to first examine weights across groups and, when those are deemed comparable, only then compare loading estimates. Finally, results from the scenario with equivalent measurement but differences in structural relationships across groups indicate that PLS can achieve comparable measurement models – which are a prerequisite to making comparisons of the structural relationships – again only under conditions where sampling variability is relatively more limited. Taken together, results from our research strongly underscore the importance of sampling variability in the ability of PLS to obtain accurate estimates of the magnitude of the differences, if any, in measurement models across groups. These results suggest that after measurement invariance of the items has been established, a good strategy for applying PLS would be to calculate the weights for a pooled sample and only then estimate the final regressions separately for each group. These conclusions apply for also the newly introduced consistent PLS (PLSc, e.g., Dijkstra & Henseler, 2015) because this new estimator uses the PLS weights as input for both the loadings estimation as well as defining the composites which are then used in regressions after correcting for attenuation. Table 9 summarizes the findings of this research.

TABLE 9. Summary of Results by Study

Study 1: Full Equivalence – The RMSD of both weights and loadings is heavily influenced by sampling variability, even when all simulated data come from models that are completely invariant by design. The larger the variability in the sample for each group, the more likely it is that PLS will estimate weights that will differ between the groups. – Researchers operating under conditions of high sampling variability should be cautious about using PLS to model this data, unless they can show the RMSD of their model is reasonable.
Study 2: Lack of Measurement Invariance – Highlights necessary but not sufficient condition for between-group comparisons in PLS – weights can be comparable even when loadings are not; both are needed before meaningful path comparisons can be made. – Even though both models here had the same structural paths, the strength of those affects the comparability of results. In general, working with models where the latent variables are more strongly related in the population makes it easier to detect violations of invariance.
Study 3: Lack of Structural Invariance – Populations that differ in the values of the structural parameters but not with regards to measurement can only be compared under certain conditions. The RMSD statistic indicates that estimates from separate models are only comparable in the upper-range of the simulation conditions (where sampling variability is relatively low). – The strength of the structural relationships in the population affects the ability of PLS to conduct effective group comparisons. Stronger models are easier to compare than weaker ones.

There are several limitations to our research, all of which provide fruitful avenues for future research. First, our results are to some extent specific to the particular research models and conditions examined here. Models with a different number of latent variables, different number of indicators, and different strength of the relationships, to name a few conditions, could be examined in order to extend the validity of the findings to other conditions. Second, in this research we examined the performance of PLS in only three distinct conditions: full equivalence, lack of measurement invariance with regards to loadings, and lack of structural invariance. There are, however, other invariance conditions that could be examined, such as intercept, mean, or factor variance invariance (e.g., Vandenberg & Lance, 2000). Given the common practice in PLS analyses of standardizing both the indicators and the composites, these differences are likely to be obscured in the processing of the data. More research is needed on how to identify these possible violations of measurement invariance and how to best assess their presence and magnitude. Third, though we examined the performance of the RMSD statistic as a measure of the extent to which measurement invariance was obtained in multi-group PLS analyses, we did not attempt to develop a formal, statistical test that can be used by researchers to provide evidence of measurement invariance in the analysis prior to interpreting the results. This also remains an important open avenue for future research. Some work in this regard has been performed by Henseler, Ringle and Sarstedt (forthcoming) in their MICOM procedure, but with a different definition of compositional invariance – that is, invariance of the weights – than employed here, one based on the correlation of composite scores. In particular, the MICOM procedure proposed a three step process of *configural*, *compositional*, and invariance of *composite means and variances*. In particular, their compositional invariance addressed the same issues as discussed in this research – that is, establishing the comparability of composites prior to performing structural comparisons. In their framework, however, composites can be deemed comparable if they are highly correlated, even if the weights used to construct them are very different. Further work is required to bring both these definitions into a common framework.

We believe our results have important implications for the conduct of research that seeks to compare two distinct groups by using a multi-group analysis in PLS (and also with PLSc, as noted above). Given that researchers would only be interested in making comparisons between groups when the composites representing the latent variables of interest are indeed comparable to begin with, we would argue such assumption must be carefully examined, which is not currently common practice. In this research we have highlighted the variable performance of PLS in this regard, and noted scenarios – particularly those relating to less-than-ideal conditions – where researchers should be cautious about interpreting their results. Moreover, we have also outlined a two-step process – examination of weights first, and then loadings, for equivalence – which researchers can follow prior to attempting interpretation of structural results. From the perspective of study and research design, researchers are advised to consider their intended analytical technique for conducting between-group comparisons in light of these results. Given that the statistics employed here can only be calculated after the data have been collected, researchers operating under less than ideal conditions, as discussed in this research, should consider the possibility that PLS will not be able to construct groups that are comparable, and thus reflect on the possibility that an alternative approach, such as covariance-based SEM, would need to be employed. Given that this approach has its own requirements researchers are encouraged to consider these issues as they perform sample size calculations and decide on their research design.

References

Aguirre-Urreta, M., Marakas, G., & Ellis, M. (2013). Measurement of Composite Reliability in Research Using Partial Least Squares: Some Issues and an Alternative Approach. *The DATA BASE for Advances in Information Systems*, *44*(4), 11–43.

Ahuja, M., & Thatcher, J. (2005). Moving beyond intentions and toward the theory of trying: Effects of work environment and gender on post-adoption information technology use. *MIS Quarterly*, *29*(3), 427–459.

Carte, T., & Russell, C. (2003). In Pursuit of Moderation: Nine Common Errors and Their Solutions. *MIS Quarterly*, *27*(3), 479–501.

Chai, S., Das, S., & Rao, R. (2011). Factors Affecting Bloggers' Knowledge Sharing: An Investigation Across Gender. *Journal of Management Information Systems*, *28*(3), 309–342.

Dijkstra, T. K., & Henseler, J. (2015). Consistent partial least squares path modeling. *MIS Quarterly*, *39*(2), 297–316.

Henseler, J., Ringle, C., & Sarstedt, M. (forthcoming). Testing Measurement Invariance of Composites Using Partial Least Squares. *International Marketing Review*.

Keil, M., Tan, B., Wei, K., Saarinen, T., Tuunainen, V., & Wassenaar, A. (2000). A cross-cultural study on escalation of commitment behavior in software projects. *MIS Quarterly*, *24*(2), 299–325.

Millsap, R. (2011). *Statistical Approaches to Measurement Invariance*. East Sussex: Routledge.

Pornprasertmanit, S., Miller, P., & Schoemann, A. (2013). *simsem: SIMulated Structural Equation Modeling*. Retrieved from http://CRAN.R-project.org/package=simsem

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

Rönkkö, M. (2014). *matrixpls: Matrix-based Partial Least Squares Estimation*. Retrieved from https://github.com/mronkko/matrixpls

Rönkkö, M., & Evermann, J. (2013). A Critical Examination of Common Beliefs About Partial Least Squares Path Modeling. *Organizational Research Methods*, *16*(3), 425–448. http://doi.org/10.1177/1094428112474693

Sia, C., Lim, K., Leung, K., Lee, M., Huang, W., & Benbasat, I. (2009). Web Strategies to Promote Internet Shopping: Is Cultural-Customization Needed? *MIS Quarterly*, *33*(3), 491–512.

Vandenberg, R. (2002). Toward a Further Understanding of and Improvement in Measurement Invariance Methods and Procedures. *Organizational Research Methods*, *5*(2), 139–158.

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

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

Venkatesh, V., Morris, M., Davis, G., & Davis, F. (2003). User Acceptance of Information Technology: Toward a Unified View. *MIS Quarterly*, *27*(3), 425–478.

Wold, H. (1982). Soft modeling - The Basic Design And Some Extensions. In K. G. Jöreskog & S. Wold (Eds.), *Systems under indirect observation□: causality, structure, prediction* (pp. 1–54). Amsterdam: North-Holland.