Evaluation of Partial Least Squares Parameter Recovery.

Completed Research Paper

Frances Chumney, PhD
University of West Georgia
fchumney@westga.edu

Abstract

The purpose of this study was to evaluate the performance of Partial Least Squares under less-than-ideal conditions selected to imitate real-world data. A simulation study with a 3×3×2×2 design was conducted. The design and data features of interest were sample size (50, 300, 1000), number of items per latent variable (3, 5, 7), degree of model misspecification (correctly specified, misspecified), and nature of the relationships between items and latent variables in the measurement models (reflective, formative). Bias of model parameter estimates, and bias and accuracy of standard error estimates were examined to evaluate the performance of PLS for each experimental condition.

Keywords: *partial least squares, structural equation modeling, small samples, reflective indicators, formative indicators, model complexity, misspecification*

1. Introduction

In response to increasing expectations from funding agencies, trends in educational research require scientists to investigate increasingly complex phenomena with regard for the contexts in which they occur. These additional layers of exploration and understanding lead to increasingly complex hypotheses and require advanced statistical techniques. Structural equation modeling (SEM) is a common analytic approach for dealing with complex systems of information. Despite their flexibility (Zhu, Walter, Rosenbaum, Russell, & Raina, 2006), traditional SEM methods require large samples in general, and even larger samples for estimating complex models. For applied researchers, large samples are often difficult and sometimes impossible to obtain.

The most common estimation method used with SEM is maximum likelihood (ML; Hoyle, 2000). ML has been studied across myriad contexts and data conditions, and its limitations are well documented. One context in which ML does not perform well is in the presence of small samples (Kline, 2011). Due to this limitation, it is imperative that researchers investigate the utility of alternative approaches to recovering parameter estimates (e.g., partial least squares (PLS), generalized structural components analysis (GSCA), Markov Chain Monte Carlo (MCMC)). If the strengths and weaknesses of each alternative method in the context of small sample research were more fully understood, researchers would be better equipped to make informed decisions with regard to selecting appropriate estimation methods and interpreting results.

2. Literature Review and Background

As the field of methodology has advanced, alternative estimation methods have developed and include generalized least squares, weighted least squares, PLS, GSCA, and MCMC approaches. Unfortunately, the performance of these alternatives is not well understood, and their performance with real data is often difficult to predict (Henseler, 2012; Hwang, Ho, & Lee, 2010; Hwang Malhotra, Kim, Tomiuk, & Hong, 2010). Although estimation methods other than those described here have been developed for use with SEMs when the assumptions of ML are violated (e.g., robust ML, weighted least squares), it is not feasible to compare and evaluate the performance of all such alternatives in a single study. Thus, the present study will focus solely on the performance of PLS for addressing the problem of estimating SEMs with small samples.

2.1. Sample size

Inherent to traditional estimation methods (i.e., ML) is the expectation of large samples. Specifically, the parameter estimates produced by ML are based on asymptotic theory, which implies large samples (Tanaka, 1987). Therefore, as sample size decreases, methods such as ML do not perform as well (e.g., Lee & Song, 2004). Proponents of PLS often promote it as performing well in instances of small samples (e.g., Chin & Newsted, 1999), but the method has been found to perform inconsistently at times (e.g., Hwang, Malhotra, et al., 2010), which indicates that more work is needed to understand the interactions between sample size and other design features.

2.2. Number of items

Just as the performance of estimation methods is expected to improve with increased sample size, estimation methods are expected to produce more reliable parameter estimates as the number of items per latent factor increases (e.g., Boomsma, 1982; Velicer & Fava, 1998). As illustrated by Marsh, Hau, Balla, and Grayson (1998), however, increasing the number of items does not necessarily improve the ability of an estimation method to recover parameter estimates. The relationship between quality of parameter estimates and number of items per latent variable has not been studied at length in the context of PLS.

3.3. Model misspecification

In both substantive and methodological research endeavors that utilize SEM, inferences and conclusions are the result of the model used. Although it is difficult to know whether or not theoretical models are specified correctly in applied research, simulation-based research has illustrated the impact of misspecification on parameter recovery across estimation methods (e.g., Asparouhov & Muthén, 2010; Hwang, Malhotra, et al., 2010). The extent to which estimates are impacted by the misspecification of the model depends on design features such as sample size (e.g., Henseler, 2010; Tanaka, 1987) and overall complexity of the model (e.g., Tanaka, 1987).

4.4. Nature of indicator-latent variable relationships

Whether the relationships between observed variables and latent constructs are formative or reflective in nature is as important to methodological study as it is to theory-driven, applied research. In the context of SEM, latent variables can be modeled as the cause of those

observed values (reflective; Bollen & Lennox, 1991), or as a representation of the combined values of those observed values (formative; Curtis & Jackson, 1962). SEMs should be specified to reflect the correct theoretical relationships, but estimation methods sometimes vary in their performance depending on the type of relationship specified. Until recent years, it was held that SEMs including formative measurement models were inappropriate for traditional ML approaches altogether (Chin, 1998; Ringle, Götz, Wetzels, & Wilson, 2009). In contrast to ML, Ringle et al. found that PLS is likely to underestimate parameters in formative models and overestimate parameters in reflective models.
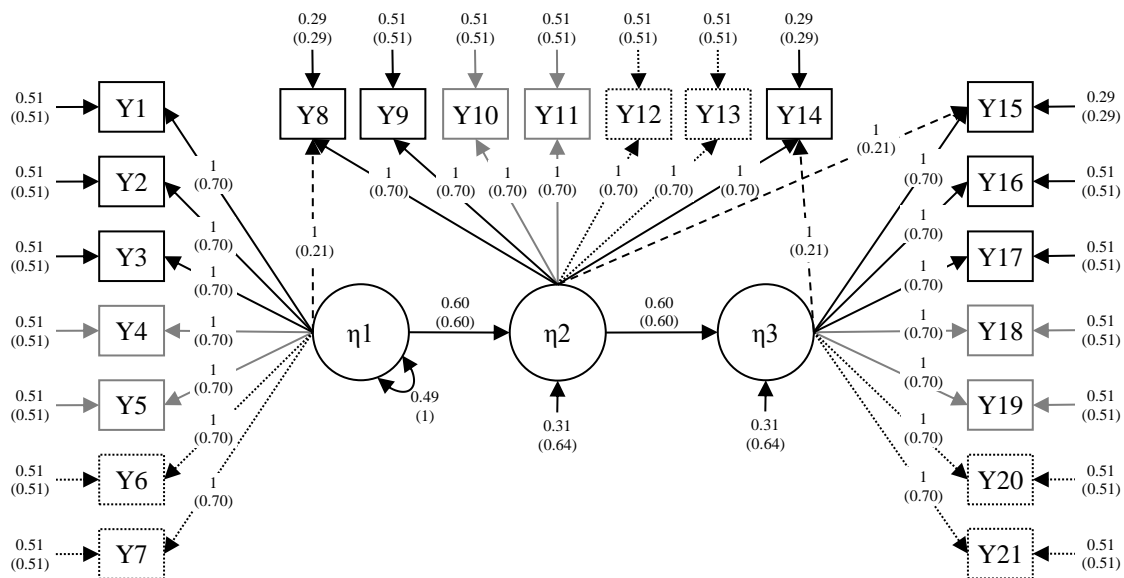
2.5. Present study

The purpose of the present study is to evaluate the performance of PLS under sub-ideal data conditions. The overarching goal of this study is to understand the effects of sample size, number of items per latent variable, model misspecification, and the nature of the latent variable-indicator relationships on the ability of PLS to recover measurement and structural model parameter estimates and their standard errors. No specific research questions were posed a priori.

3. Method

3.1. Simulation design

The experimental conditions of interest in the present study as are follows: sample size (N = 50, 300, 1000), direction of relationship between items and factors (reflective, formative), number of items per latent variable ($i = 3, 5, 7$), and degree of model misspecification (correctly specified, misspecified). These experimental conditions were selected because they are representative of conditions common to both simulation-based research on structural equation modeling and researchers dealing with applied datasets.
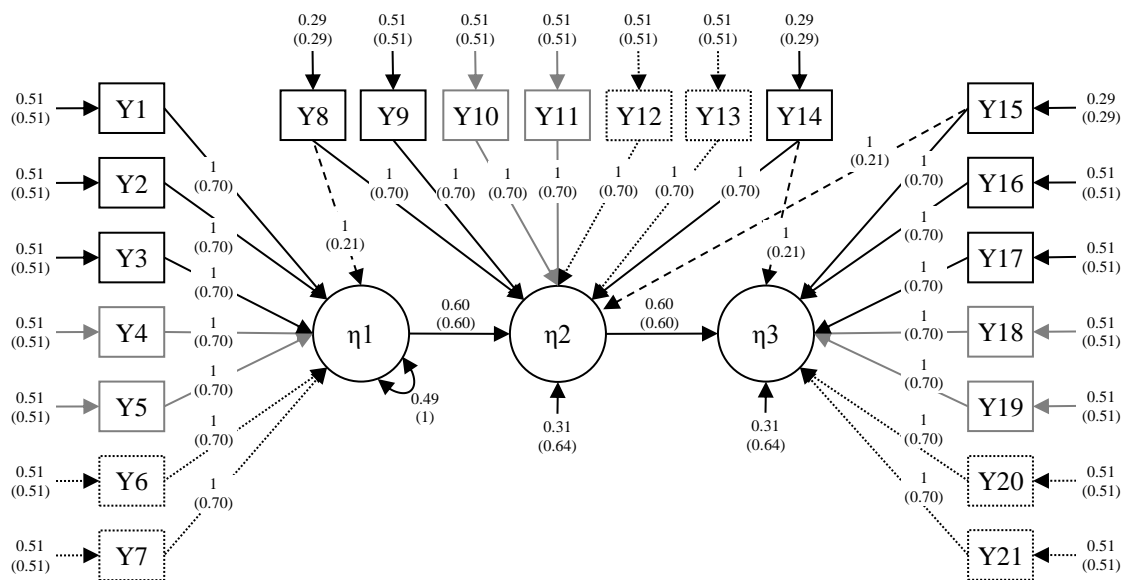

Note: Manifest variables and relationships indicated by gray lines are included in conditions that include more than 3 items; dotted lines are included only in conditions that include 7 items. Dashed lines indicate cross-loading relationships included only in the generation of data for misspecification conditions.

Figure 1: Population model for conditions including reflective indicators

The data for this study were simulated to reflect SEMs common to both applied and simulation-based research. The population model used for this study includes three latent variables, an equal number of items per latent variable, and no cross-loadings. This model was selected for its simplicity and similarity to population models employed by previous research (e.g., Henseler, 2012; Hwang, Malhotra, et al., 2010; Paxton, Curran, bollen, Kirby, & Chen, 2001; Tomás, Hontangas, Oliver, 2000). The population model used for conditions featuring reflective indicator-latent variable relationships is identical to model used by Hwang, Malhotra, et al. (2010) and Paxton et al. (2001), and is displayed in Figure 1.

As noted by Vinzi, Trinchera, and Amato (2010), the relationship between reflective and formative measurement models is essentially the same relationship that exists between factor models with high reliability among the indicators (reflective models with measurement model error) and factor models with low reliability among the indicators (formative models with essentially no measurement model error; e.g., Diamantopoulos, Riefler, & Roth, 2008). Thus, for experimental conditions under which formative indicator-latent variables are of interest in the present study, reflective models with low reliability among indicators was used for data generation (see Figure 2).

Note: Manifest variables and relationships indicated by gray lines are included in conditions that include more than 3 items; dotted lines are included only in conditions that include 7 items. Dashed lines indicate cross-loading relationships included only in the generation of data for misspecification conditions.

Figure 2: Population model for conditions including formative indicators

This approach for modeling formative indicator-latent variable relationships was selected to establish a greater level of consistency across experimental conditions, as it allowed for all data sets to be simulated in the same manner. Further, conceptualizing the formative conditions as error-free reflective models makes it possible to apply each estimation method in the same manner across all data sets regardless of population model (i.e., Mode A estimation was implemented for the PLS approach across all conditions). Ultimately, this allows for the differences in the performance of PLS across experimental conditions to be attributed to the nature of the indicator-latent variable relationships instead of to differences between Mode A and Mode B estimation.

M*plus* (version 6; Muthén & Muthén, 1998-2010) was used to simulate 150 replications of each unique condition. Several studies have been completed which rely on fewer replications and an ANOVA approach to analysis (e.g., 100 replications as reported in Kankaraš, Vermunt, & Moors, 2011; Lee, Song, & Lee, 2003; Lee & Xia, 2008; Lee & Zhu, 2002; Song

& Lee, 2002; Song, Lee, & Hser, 2008; and 200 replications as reported in Fan, Thompson, & Wang, 1999; Hu & Bentler, 1999; Jackson, 2003, 2007). The plsSEM package (Monecke & Leisch, 2012) developed for R (R Development Core Team, 2012) was used to obtain model parameter and standard error estimates. For each replication, the maximum number of iterations allowed was set to 1,000, and the number of bootstrap samples used to recover standard error estimates was set to 500.

3.2. Outcomes of interest

To evaluate the performance of each estimation approach for recovering parameter estimates under the varying experimental conditions, six characteristics of the recovered estimates were examined: average measurement model bias, average structural model bias, mean absolute differences of the measurement model standard error estimates, mean absolute differences of structural model standard error estimates, accuracy of standard error estimates for the measurement model, and accuracy of standard error estimates for the structural model. Standardized estimates were used in the calculation of all outcomes.

3.2.1. Parameter estimate bias

For this study, parameter estimate bias was defined as the proportion of the difference between the sample and population values, relative to the population values (Enders & Bandalos, 2001), and was calculated

$$\%BIAS = \left[\frac{|\theta_i - \theta_B|}{\theta_B}\right] \times 100$$

where θ_i is the recovered parameter estimate and θ_B is the known population parameter. Average bias was calculated separately for the measurement and structural models in each replication data set.

3.2.2. Mean absolute difference of standard error estimates

The mean absolute difference between standard error estimates and their corresponding empirical standard errors (MAD; Hwang, Malhotra, et al., 2010) were calculated as

$$MAD = \frac{\sum_{j=1}^{P} |\text{SE}(\hat{\theta}_j) - \text{SE}(\theta_j)|}{P}$$

where $\text{SE}(\hat{\theta}_j)$ is the recovered standard error estimate, $\text{SE}(\theta_j)$ is the true value for that standard error, and P is the number of parameters. The true values for $\text{SE}(\theta_j)$ were obtained empirically via a Monte Carlo simulation (conducted in M*plus*, version 6, Muthén & Muthén. 1998-2010) which included 500 replications and 2,000 bootstrap resamples per replication for each experimental condition. True (empirical) standard errors were calculated as

$$\text{SE}(\theta_j) = \sqrt{\frac{\sum_{i=1}^{B}\left(\hat{\theta}_j - \bar{\bar{\theta}}_j\right)^2}{B-1}}$$

where $\hat{\theta}_j$ is the parameter estimate obtained for a single replication, and $\bar{\bar{\theta}}_j$ is the mean parameter estimate obtained for B replications (Hwang, Malhotra, et al., 2010; Sharma,

Durvasula, & Dillon, 1989; Srinivasan & Mason, 1986). MAD was calculated separately for the measurement and structural models in each replication data set.

3.2.3. Accuracy of standard error estimates

The ability of the estimation methods to produce standard errors was also evaluated by constructing a confidence interval around each parameter estimate and determining whether the corresponding population parameter falls within this confidence interval (i.e., accuracy of the standard error estimate; Gerbing & Anderson, 1985). For this purpose, the confidence interval was defined as ±1.96 standard errors around the parameter estimate, and the value of interest is the proportion of parameter estimates for which the population parameter falls within the appropriate confidence interval. This value was calculated for each replication to reflect the accuracy of the standard errors associated with the measurement and structural models separately.

3.3. Analytic approach

A multivariate analysis of variance (MANOVA) was calculated to evaluation the performance of the PLS across experimental conditions. The MANOVA analysis included the four design factors as independent variables (i.e., sample size, number of items per latent variable, degree of misspecification, type of latent variable-indicator relationships), and the six outcomes of interest as dependent variables (i.e., bias in measurement model estimates, bias in structural model estimates, MAD and accuracy measurement model standard error estimates, MAD and accuracy of structural model standard error estimates). All interaction effects were included in the MANOVA. Effect sizes (partial η^2) were calculated for each direct and interaction effect. This method is consistent with recommendations and practices in this field (e.g., Hwang, Malhotra, et al., 2010; Paxton et al., 2001), and strengthens the connection between this and previous work.

4. Results

The four-factor MANOVA was computed as the first step toward understanding the effects of sample size, number of items per latent variable, degree of misspecification, and type of indicator-latent variable relationships within the present study. The results of the MANOVA are displayed in Table 1. It is important to note that the significant effects may be a reflection of the large number of observations included in the complete data set for this study (a total of 3,600 observations representing 150 replications for each of experimental design condition). For this reason, only significant results for which the tests of between-subjects effects were characterized by a medium or large effect size (i.e., partial $\eta^2 \geq .06$) are presented (Hwang, Malhotra, et al., 2010; Paxton et al., 2001). In instances where pairwise comparisons are made, only significant results for which the differences are characterized by a medium or large effect size (i.e., $d \geq .50$) are presented (Cohen, 1988). Accordingly, direct and interaction effects are described as moderate or large and not as significant or not significant. Where relevant, p values are reported as a matter of standard practice, not for the purpose of interpreting effects.

Table 1: Multivariate effects

Effect	Wilks' Λ	df	F
Intercept	0.004	6	212972.42[‡]
Sample Size (n)	0.047	12	3221.96[‡]
Number of Items per Latent Variable (items)	0.080	12	2273.11[‡]
Degree of Misspecification (spec)	0.030	6	29383.39[‡]
Type of Measurement Model Relationships (iLV)	0.044	6	19441.61[‡]
n×items	0.178	24	499.29[‡]
n×spec	0.344	12	629.07[‡]
n×iLV	0.081	12	2239.75[‡]
items×spec	0.242	12	920.96[‡]
items×iLV	0.251	12	890.19[‡]
spec×iLV	0.032	6	26874.92[‡]
n×items×spec	0.311	24	310.08[‡]
n×items×iLV	0.191	24	473.17[‡]
n×spec×iLV	0.382	12	552.47[‡]
items×spec×iLV	0.146	12	1443.50[‡]
n×items×spec×iLV	0.331	24	290.58[‡]

Notes: all $p < 0.001$; [‡] partial $\eta^2 > 0.13$

4.1. Bias in measurement model parameter estimates

Sample size (partial $\eta^2 = 0.13$), number of items per latent variable (partial $\eta^2 = 0.89$), specification (partial $\eta^2 = 0.95$), and type of indicator-latent variable relationships included in the measurement portion of the model (partial $\eta^2 = 0.80$) were all found to have a large effect on the bias of parameter estimates for the measurement portion of the models. The four-way (sample size × number of items per latent variable × specification × indicator-latent variable relationships) interaction was not found to have an effect on the bias of parameter estimates for the measurement portion of the models ($p > 0.05$; partial $\eta^2 = 0.00$).

Follow-up, pairwise comparisons indicate that PLS recovers measurement model parameter estimates with less bias for models with reflective indicators compared to models with formative indicators when the model is correctly specified (all $d > 0.50$). When the model is misspecified, PLS yields less biased measurement model parameter estimates for formative models (all $d > 0.50$) than reflective models. For correctly specified models that include formative indicator-latent variable relationships, the bias of measurement model parameter estimates increases as the number of items per latent variable increases (all $d > 0.50$). The bias of recovered measurement model parameter estimates decreases as the number of items increases (all $d > 0.50$) for correctly specified models with reflective measurement structures as well as for misspecified models (regardless of the type of indicator-latent variable relationships included in the measurement model).

4.2. Bias in structural model parameter estimates

Sample size (partial $\eta^2 = 0.18$) and number of items per latent variable (partial $\eta^2 = 0.25$) were each found to have a large effect on the bias of parameter estimates for the structural portion of the models; neither degree of misspecification nor type of indicator-latent variable relationships included in the measurement portion of the model were found to effect the bias of structural model estimates. The four-way (sample size × number of items per latent

variable × specification × indicator-latent variable relationships) interaction was not found to be significant.

Follow-up analyses indicated that PLS resulted in more biased structural model estimates for reflective models than formative models when the model was correctly specified; when the model was misspecified, PLS produced more biased structural model estimates for formative models than reflective models. Across all misspecified models and correctly specified models with reflective indicators, bias of structural model parameter estimates was found to decrease as the number of items per latent variable increased. No change in the bias of observed structural model estimates was observed as number of items increased for correctly specified, formative models.

4.3. Mean absolute difference of measurement model standard error estimates

Sample size (partial $\eta^2 = 0.87$), number of items per latent variable (partial $\eta^2 = 0.72$), specification (partial $\eta^2 = 0.59$), and type of indicator-latent variable relationships included in the measurement portion of the model (partial $\eta^2 = 0.87$) were found to have a large effects on the MAD of measurement model standard error estimates. The effect of the four-way (sample size × number of items per latent variable × specification × indicator-latent variable relationships) interaction was found to be large (partial $\eta^2 = 0.58$). PLS was found to yield less biased standard error estimates for the measurement model for formative models than for reflective models. MAD of measurement model standard error estimates was found to decrease (improve) as the number of items per latent variable increased for misspecified models as well as for correctly specified, reflective models (all $d > 0.50$).

4.4. Accuracy of measurement model standard error estimates

Sample size (partial $\eta^2 = 0.70$), number of items per latent variable (partial $\eta^2 = 0.27$), and type of indicator-latent variable relationships included in the measurement portion of the model (partial $\eta^2 = 0.75$) were found to have a large effects on the accuracy of standard error estimates for the measurement model parameters. The four-way (sample size × number of items per latent variable × specification × indicator-latent variable relationships) interaction was not found to be significant. Follow-up analyses indicated that PLS yielded more accurate estimates of measurement model standard errors for reflective models than formative models, regardless of degree of misspecification. Accuracy of standard error estimates for the measurement model was observed to increase and the number of items increased for reflective models regardless of degree of misspecification (all $d > 0.50$). No changes in accuracy of standard error estimates were observed for formative models as the number of items increased.

4.5. Mean absolute difference of structural model standard error estimates

Sample size (partial $\eta^2 = 0.84$), number of items per latent variable (partial $\eta^2 = 0.25$), degree of misspecification (partial $\eta^2 = 0.35$), and type of indicator-latent variable relationships included in the structural portion of the model (partial $\eta^2 = 0.43$) were found to have a large effects on the MAD of structural model standard error estimates. The four-way (sample size × number of items per latent variable × specification × indicator-latent variable relationships) interaction was not found to be significant. Follow-up analyses indicated that PLS yields more precise estimates of structural model standard errors for formative models than for reflective models, regardless of degree of misspecification. The MAD of structural model standard estimates was found to decrease and the number of items per latent variable

increased for misspecified models as well as for correctly specified models with reflective indicators (all $d > 0.50$).

4.6. Accuracy of structural model standard error estimates

Sample size (partial $\eta^2 = 0.35$), number of items per latent variable (partial $\eta^2 = 0.08$), and degree of misspecification (partial $\eta^2 = 0.15$) were found to have a large effects on the accuracy of standard error estimates for the structural model parameters; no effect of type of indicator-latent variable relationships included in the measurement portion of the model (partial $\eta^2 = 0.01$) was found. The four-way (sample size $\times$ number of items per latent variable $\times$ specification $\times$ indicator-latent variable relationships) interaction was not found to be significant. Follow-up analyses indicated that PLS yielded more accurate estimates of structural model standard errors for formative models than for reflective models when correct specification was used and the sample size was moderate or large ($n = 300, 1000$; all $d > 0.50$); no difference in accuracy of structural model estimates between reflective and formative models was observed when applied to a small sample ($n = 50$). When applied to misspecified models, PLS yielded more accurate standard error estimates for the structural model with reflective indicators than formative indicators (all $d > 0.50$). The accuracy of structural model standard error estimates increased as the number of items per latent variable increased for misspecified, formative models as well as for correctly specified, reflective models. No change in accuracy was observed as number of items increased for misspecified, reflective models or correctly specified, formative models.

5. Brief Discussion

This study attempted to replicate and extend previous research by evaluating the performance of PLS under varying data conditions. Generally speaking, it was found that PLS does not perform consistently across experimental conditions. Not surprisingly, the performance of PLS was found to vary between reflective and formative models, as well as between correctly specified and misspecified models. Further, the performance of PLS for measurement model parameter recovery was not equivalent to the performance of PLS for structural model parameter recovery. Table 2 (omitted from this document due to length) indicates the conditions under which PLS produced the best estimates for each outcome of interest.

5.1. Limitations and future research

Part of the discrepancy in performance between correctly specified models and misspecified models may be a result of the methods used for data simulation and estimation. First, all data were generated for reflective models, with parameter values varied to simulate formative relationships for formative models. Second, all parameter estimates were recovered using PLS Mode A estimation. Future research should consider implications of a different method of data simulation, as well as the relative performance of PLS when models are estimated using the Mode B approach. Even though it is not expected that the sole use of Mode A negatively influenced the estimation process or recovered estimates, it is a question worth empirical investigation.

A second limitation of the present study is the simplicity of the population and analytic models. The population models used for the present study were relatively simple compared to some models employed by substantive researchers. Specifically, all data were generated as

normally distributed representations of their respective variables, but typical data is rarely normally distributed. Both the population and analytic models used for the present study were relatively simple: each latent variable was related to an equal number of indicators in the measurement models, and the structural models included only a minimal number of latent variables and relationships between those latent variables. The simplicity of the models examined were appropriate for the investigatory nature of the present study. Future research, however, should examine the performance of PLS when applied to more complex models (e.g., cross-loadings as part of the analytic model, combination of reflective and formative indicators in the measurement model, misspecification in the structural portion of the model, multiple group analyses, etc.).

5.2. Implications and conclusion

The driving force behind the need for this type of research is to provide a more complete understanding of the performance of PLS under conditions common to applied research data. The findings presented herein are best interpreted as guidance for the development of additional methodological work to extend this research and delve deeper into the issues at hand. Applied researchers are cautioned to remember that these findings are contingent upon the characteristics of the data generated for this study (i.e., normally distributed variables throughout the measurement and structural models) - characteristics sometimes uncommon in substantive, "real-world" research endeavors.

6. References

Asparouhov, T., & Muthén, B. (2010). *Bayesian analysis using Mplus: Technical implementation*. Retrieved December 18, 2012 from http://www.statmodel.com

Bollen, K. A., & Lennox, R. (1991). Conventional wisdom on measurement: A structural equation perspective. *Psychological Bulletin, 110*(2), 305-314.

Boomsma, A. (1982). Robustness of LISREL against small sample sizes in factor analysis models. In K. G. Joreskog & H. Wold (Eds.), *Systems under indirect observation: Causality, structure, prediction* (Part 1), pp. 149-173. Amsterdam: North Holland.

Chin, W. W. (1998). Issues and opinion on structural equation modeling. *Management Information Systems Quarterly, 22*(1).

Chin, W. W., & Newsted, P. R. (1999). Structural equation modeling analysis with small samples using partial least squares. In R. H. Hoyle (Ed.), *Statistical strategies for small sample research* (pp. 307-341). Thousand Oaks, CA: Sage.

Cohen, J. (1988). *Statistical power analyses for the behavioral sciences*. Hillsdale, NJ: Erlbaum Associates.

Curtis, R. F., & Jackson, E. F. (1962). Multiple indicators in survey research. *American Journal of Sociology, 68*, 195-204.

Diamantopoulos, A., Riefler, P., & Roth, K. P. (2008). Advancing formative measurement models. *Journal of Business Research, 61*, 1203-1218.

Enders, C. K., & Bandalos, D. L. (2001). The relative performance of full information maximum likelihood estimation for missing data in structural equation models. *Structural Equation Modeling, 8*(3), 430-457.

Fan, X., Thompson, B., & Wang, L. (1999). Effects of sample size, estimation methods, and model specification on structural equation modeling fit indexes. *Structural Equation Modeling, 6*(1), 56-83.

Gerbing, D. W., & Anderson, J. C. (1985). The effects of sampling error and model characteristics on parameter estimation for Maximum Likelihood confirmatory factor analysis. *Multivariate Behavioral Research, 20*, 255-271.

Henseler, J. (2010). On the convergence of the partial least squares path modeling algorithm. *Computational Statistics, 25*, 107-120.

Henseler, J. (2012). Why generalized structured component analysis is not universally preferable to structural equation modeling. *Journal of the Academy of Marketing Science, 40*, 402-413.

Hoyle, R. H. (2000). Confirmatory factor analysis. In H. E. A. Tinsley & S. D. Brown (Eds.), *Handbook of applied multivariate statistics and mathematical modeling*. New York: Academic Press.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*, 1-55.

Hwang, H., Ho, M. H. R., & Lee, J. (2010). Generalized structured component analysis with latent interactions. *Psychometrika, 75*, 228-242.

Hwang, H., Malhotra, N. K., Kim, Y., Tomiuk, & Hong (2010). A comparative study on parameter recovery of three approaches to structural equation modeling. *Journal of Marketing Research, 47*, 699-712.

Hwang, H., & Takane, Y. (2004). Generalized structured component analysis. *Psychometrika, 69*(1), 81-99.

Jackson, D. L. (2003). Revisiting sample size and number of parameter estimates: Some support for the N:q hypothesis. *Structural Equation Modeling, 10*(1), 128-141.

Jackson, D. L. (2007). The effect of the number of observations per parameter in misspecified confirmatory factor analytic models. *Structural Equation Modeling, 14*(1), 48-76.

Kankaraš, M., Vermunt, J. K., & Moors, G. (2011). Measurement equivalence of ordinal items: A comparison of factor analytic, item response theory, and latent class approaches. *Sociological Methods & Research, 40*(2), 279-310.

Kline, R. B. (2011). Principles and practice of structural equation modeling. New York, NY: Guilford.

Lee, S. Y., & Song, X. Y. (2004). Evaluation of the Bayesian and maximum likelihood approaches in analyzing structural equation models with small sample sizes. *Multivariate Behavioral Research, 39*, 653-686.

Lee, S. Y., & Song, X. Y., & Lee, J. C. K. (2003). Maximum likelihood estimation of nonlinear structural equation models with ignorable missing data. *Journal of Educational and Behavioral Statistics, 28*(2), 111-134.

Lee, S. Y., & Xia, Y. M. (2008). A robust Bayesian approach for structural equation models with missing data. *Psychometrika, 73*(3), 343-364.

Lee, S. Y., & Zhu, H. T. (2002). Maximum likelihood estimation of nonlinear structural equation models. *Psychometrika, 67*(2), 189-210.

Marsh, H. W., Hau, K.-T., Balla, J. R., & Grayson, D. (1998). Is more ever too much? The number of indicators per factor in confirmatory factor analysis. *Multivariate Behavioral Research, 33*(2), 181-220.

Monecke, A., & Leisch, F. (2012). semPLS: Structural equation modeling using partial least squares. *Journal of Statistical Software, 48*(3), 1-32.

Muthén, L. K., & Muthén, B. O. (1998-2010). M*plus* User's Guide. Sixth Edition. Los Angeles, CA: Muthén & Muthén.

Paxton, P., Curran, P. J., Bollen, K. K., Kirby, J. B., & Chen, F. (2001). Monte Carlo experiments: Design and implementation. *Structural Equation Modeling, 8*(2), 287-312.

R Development Core Team (2012). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL *http://www.R-project.org.*

Ringle, C. M., Götz, O., Wetzels, M., & Wilson, B. (2009). On the use of formative measurement specifications in structural equation modeling: A Monte Carlo simulation study to compare covariance-based and partial least squares model estimation methodologies. In *METEOR Research Memoranda (RM/09/014)*: Maastricht University.

Sharma, S., Durvasula, S., & Dillon, W. R. (1989). Some results on the behavior of alternate covariance structure estimation procedures in the presence of non-normal data. *Journal of Marketing Research, 26*, 214-221.

Song, X. Y., & Lee, S. Y. (2002). Analysis of structural equation model with ignorable missing continuous and polytomous data. *Psychometrika, 67*(2), 261-288.

Song, X. Y., & Lee, S. Y. (2005). Maximum likelihood analysis of nonlinear structural equation models with dichotomous variables. *Multivariate Behavioral Research, 40*(2), 151-177.

Song, X. Y., & Lee, S. Y. (2006). Model comparison of generalized linear mixed models. *Statistics in Medicine, 25*, 1685-1698.

Song, X. Y., Lee, S. Y., & Hser, Y. I. (2008). A two-level structural equation model approach for analyzing multivariate longitudinal responses. *Statistics in Medicine, 27*, 3017-3041.

Srinivasan, V., & Mason, C. H. (1986). Nonlinear least squares estimation of new product diffusion models. *Marketing Science 5*, 169-178.

Tanaka, J. S. (1987). "How big is big enough?": Sample size and goodness of fit in structural equation models with latent variables. *Child Development, 58*(1), 134-146.

Tomás, J. M., Hontangas, P. M., & Oliver, A. (2000). Linear confirmatory factor models to evaluate multitrait-multimethod matrices: The effects of number of indicators and correlation among methods. *Multivariate Behavioral Research, 35*(4), 469-499.

Velicer, W. F., & Fava, J. L. (1998). Effects of variable and subject sampling on factor pattern recovery. *Psychological Methods, 3*(2), 231-251.

Vinzi, V. E., Trinchera, L., & Amato, S. (2010). PLS path modeling: From foundation to recent developments and open issues for model assessment and improvement. In V. E. Vinzi et al. (Eds.), *Handbook of Partial Least Squares*. Springer-Verlag: Berlin.

Zhu, B., Walter, S. D., Rosenbaum, P. L., Russell, D. J., & Raina, P. (2006). Structural equation and log-linear modeling: A comparison of methods in the analysis of a study on caregivers' health. *BMC Medical Research Methodology, 6*, 49-62.