

You Write, but Others Read: Common Methodological Misunderstandings in PLS and Related Methods

George A. Marcoulides and Wynne W. Chin

Abstract PLS and related methods are currently enjoying widespread popularity in part due to the availability of easy to use computer programs that require very little technical knowledge. Most of these methods focus on examining a fit function with respect to a set of free or constrained parameters for a given collection of data under certain assumptions. Although much has been written about the assumptions underpinning these methods, many misconceptions are prevalent among users and sometimes even appear in premier scholarly journals. In this chapter, we discuss a variety of methodological misunderstandings that warrant careful consideration before indiscriminately applying these methods.

Key words: Structural equation models, Path models, Confirmatory factor analysis, Multiple regression, Path analysis, Covariance structure analysis, Latent class, Mixture analysis, Equivalent models, Power, Model identification, Formative indicators, Reflective indicators, Mode A, Mode B, Scale invariance

1 Introduction

We begin this section with a short story, which led to the central theme and focus of the chapter. Not long ago, the first author received an action letter from a journal editor concerning a manuscript he had submitted for publication consideration.

G.A. Marcoulides (✉)

Research Methods and Statistics, Graduate School of Education
and Interdepartmental Graduate Program in Management, A. Gary Anderson Graduate
School of Management, University of California, Riverside, CA 925121, USA
e-mail: georgem@ucr.edu

W.W. Chin

Department of Decision and Information Systems, C. T. Bauer College of Business,
University of Houston, Houston, TX 77204-6021, USA
e-mail: wchin@uh.edu

The editor was essentially asking for assistance in tackling a comment provided by one of the manuscript reviewers. The reviewer's comment specifically stated that "... you did not do this analysis correctly you need to follow the procedures outline in Marcoulides [1]..." You can imagine the dismay, especially since the procedures were not outlined in the article in the manner referenced by the reviewer. When the editor was contacted and told that the original writings had been misunderstood by the reviewer, he responded with instructions that "... perhaps you should write more clearly next time..."

Interestingly enough, this experience mirrors one of our favorite passages provided by Galton [2] in his book *Natural Inheritance*:

... some people hate the name statistics, but I find them full of beauty and interest. Whenever they are not brutalized, but delicately handled by the higher methods, and are warily interpreted, their power dealing with complicated phenomena is extraordinary (p. 62).

Galton's words are as *à propos* today as when he wrote them. Whenever issues examined are complex, both theoretical and procedural, people with limited information and/or limited knowledge will likely develop misunderstandings. Like a rumor that contains half-truths, conceptualizations and insights often contain partially correct information. Unfortunately as characterizations of what constitutes good research, they can often lead people away from important understandings (e.g., see the detailed commentary offered by Marcoulides, Chin, and Saunders [3] on comparisons between various modeling techniques).

PLS, and related methods currently enjoy widespread popularity in the behavioral, information, social, and educational sciences. The number of contributions to the literature in terms of books, book chapters, and journal articles applying or attempting to develop extensions to these methods are appearing at an incredible rate. A major reason for their appeal is that these methods allow researchers to examine models of complex multivariate relationships among all types of variables (observed, weighted component, or latent) whereby the magnitude of both direct and indirect effects can be evaluated. Another reason is due to the availability of easy to use computer programs that often require very little technical knowledge about the techniques. For example, programs such as AMOS [4], EQS, [5], LISREL, [6], LVPLS, [7], Mplus, [8], Mx, [9], PLS-Graph, [10, 11], PLS-GUI, [12, 13], RAMONA, [14], SAS PROC CALIS, [15], SEPATH, [16], Smart-PLS, [17], VisualPLS, [18, 19], and XLStat PLSPM, [20], are all broadly available for the analyses of models. Using these programs a variety of complex models can be examined and include confirmatory factor analysis, multiple regression, path analysis, models for time-dependent data, recursive and non-recursive models for cross-sectional, longitudinal data and multilevel data, covariance structure analysis, and latent class or mixture analysis to name a few. A frequent assumption made when using these models is that the relationships among the considered variables are linear (although modeling nonlinear relationships is also becoming increasingly popularity, for details see, [21]).

Unfortunately, once advanced modeling methods become widely available in easy to use software packages, they also tend to be quickly abused. Although these methods are based on a number of very specific assumptions and much has been written about adhering to the assumptions underpinning these techniques, many misconceptions are prevalent among users and sometimes even appear in premier

scholarly journals. Applied researchers conducting analyses based on these modeling methods generally proceed in three stages: (i) a theoretical model is hypothesized, (ii) the overall fit of data to the model is determined, and (iii) the specific parameters of the proposed model are evaluated. A variety of problems might be encountered at each of these stages, especially if assumptions are not kept in mind and examined. Indeed as indicated by Marcoulides et al. [3], applied researchers often do not pay sufficient attention to the stochastic assumptions underlying particular statistical models. This lack of attention to espoused statistical theory can also divert focus from precise statistical statements, analyses, and applications, by purporting to do what cannot be legitimately done with the particular recommended approach and even encouraging others to engage in similar activities. An example that immediately comes to mind is the investigation that attempted to make comparisons between modeling methods provided in Goodhue, Lewis, and Thompson (GLT, [22]). In this investigation, the authors ignored basic issues that most statisticians would consider essential preliminaries to any attempt to apply these methods in practice. So does this suggest that applied investigators (and for that matter, those acting as reviewers, since this manuscript made it through a full review cycle) do not pay much attention to the fundamental assumptions behind the modeling approaches? Is it because they find it difficult to follow the original sources or have misunderstood the writings of the original developers leading to attempts to legitimize bad practices simply because they lack a deep understanding of the mathematical details? In either case, when Marcoulides et al. [3] objected to the inappropriateness of the GLT comparisons, describing them as attempts to compare apples with oranges, the associate editor handling the original manuscript submission stated that “simply pointing out the flaws in GLT is not an effective critique without the inclusion of a direction towards the solution (January 30, 2012, AE Report).” It seems that this associate editor and the overseeing senior editor most likely had not understood the mathematical details of the methodology. Furthermore, they did not realize that in this situation no exact solution was possible and only an approximation was possible (for approximation details see [23]).

The purpose of this chapter, therefore, is to discuss a variety of methodological misunderstandings that warrant careful consideration before applying these methods indiscriminately and obtaining inappropriate interpretations. Naturally many of these issues have been addressed before, in one form or another, either by us or by a number of other researchers. Thus, although in general this paper may be viewed as containing nothing that has not been said before, our approach to the topics is intended to be more informative and didactic. For additional details and more abstruse discussion, we refer readers to the original sources and other essential outlets.

The issues to be addressed here are: (i) modeling perspectives for conducting analyses, (ii) equivalent models, (iii) sample size, (iv) identification issues, (v) myths about coefficient α , (vi) the use of correlation and covariance matrices, and, finally, (vii) comparisons among PLS and related methods. Extensive listing to resources that are readily available in the literature outlining all the issues in detail will be avoided including how one can legitimately use these methods. Also barring inclusion of statements such as “do not try this at home” or “use at your own risk” on all commercially available software programs, in what follows we attempt

to summarize main concerns and provide guidelines towards using PLS and related methods. While many of our comments will occasionally be quite critical and may even come across as inappropriate and derogatory, but, as indicated by Cudeck [24], we believe that “it is good for one’s character, not bad for it, to acknowledge past errors and clearly be capable of learning (p. 317).” Steiger [25] compared entry into the practice of using these modeling methods as akin to trying to merge onto a busy superhighway filled with large trucks and buses driving fast in reverse. Without doubt, the knowledge base required for understanding such analyses is continually expanding, but it is essential if one is to avoid professional embarrassment.

2 Overview of Modeling Perspectives for Conducting Analyses

The fitting and testing of any theoretical model can be considered from three general modeling perspectives or approaches [26]. The first is the so-called strictly confirmatory approach in which a single initially proposed theoretical model is tested against obtained empirical data and is either accepted or rejected. The second situation is one in which a finite number of competing or alternative theoretical models are considered. All proposed models are assessed and the best is selected based upon which model fits the observed data best using any number of currently available fit criteria. The third situation is the so-called model generating approach in which an initially proposed theoretical model is repeatedly modified until some acceptable level of fit is obtained. Of course, we strongly believe that the decision regarding which approach to follow should generally be based on the initial theory. A researcher who is firmly rooted in his or her theory will elect a different approach than one who is quite tentative about the various relationships being modeled or one who acknowledges that such information is what actually needs to be explored and determined. Nevertheless, once a researcher has determined that an initially proposed model is to be abandoned, the modeling approach is no longer confirmatory. Under such circumstance, the modeling approach has clearly entered an exploratory mode in which revisions to the model occur, either by simply adding and/or removing parameters in the model or even completely changing or modifying the initially proposed model both in terms of latent variables, observed variables, and/or their path connections and correlations.

The notion of changing aspects of a PLS model fits quite well with the original ideologies of its founder Wold [27] who indicated that “...PLS is primarily designed for research contexts that are simultaneously data-rich and theory skeletal (p. 26), ...it is an evolutionary process,” one in which “...at the outset the arrow scheme ...is more or less tentative ...” Indeed, Wold [27] saw absolutely nothing wrong with “...getting indications for modifications and improvement, and gradually consolidating the design ...” until a final model is selected. For example, if the values of the loadings for a latent variable show high correlations with an observed variable that has not yet been considered for inclusion as one of its indicators, this might be subsequently deemed worthy of inclusion among its indicators. In a simple confirmatory factor analytic model $\Sigma_{xx} = \Lambda \Phi \Lambda' + \Theta$, where Σ_{xx} is the co-

variance/correlation matrix of the observed $\mathbf{x}$ variables, Λ the factor loading matrix, Φ the factor correlation matrix, and Θ is the error matrix, which is commonly set up by a priori imposing a number of restrictions on Λ and Φ , such an approach would entail changing (fixing or freeing) other additional aspects of these particular matrices (for further details on model restrictions in factor analysis, see [28]).

When considering competing theoretical models, the number of possibilities to compare are feasible for small sets of variables. For example, with only two observed variables, there are only four possible models to examine. For 3 variables, there are now 64 total possible models to examine. However, with more variables in play, the number of possible model combinations can become prohibitively large. For example, even with just 6 observed variables there are 1,073,741,824 possible model combinations to examine. One way to think about the total number of models among P investigated variables is to consider the number of possible ways each pair can be connected, to the power of the number of pairs of variables and is determined by $4^{[P(P-1)/2]}$ [29]. Nonetheless, when examining all possible models becomes impractical, various heuristic optimization or automated search algorithms can be used [30]. Although heuristic search algorithms are specifically designed to determine the best possible model based upon some objective function solution, they do not guarantee that the optimal solution is found—though their performance using empirical testing or worst cases analysis indicates that in many situations they seem to be the only way forward to produce concrete results [31]. So as the models become more complicated, automated procedures can at least make “chaotic situation(s) somewhat more manageable by narrow(ing) attention to models on a recommendation list” ([32], p. 266). Heuristic model search procedures have recently made their way into the general modeling literature. Examples of such numerical heuristic model search procedures include: ant colony optimization ([33–36]), genetic algorithms [37], ruin-and-recreate [38], simulated annealing [39], and Tabu search [30, 40]—and over the years a great variety of modifications have been proposed to these procedures (e.g., [41, 42]). As indicated, all of these methods focus on the evaluation of an objective function, which is usually based upon some aspect of model fit (e.g., the Lagrange multiplier or the Stone-Geisser Criterion; for additional details see [43]).

Model searches and modifications are extremely difficult especially whenever the number of possible variables and potential models are high. Thus, automated algorithms have the potential to be quite helpful for examining models, particularly where all available information has been included in a specified model and when this information is not sufficient to obtain an acceptable model fit. Nevertheless—despite the fact that such searches can usually determine the best models according to a given fit criteria—all final generated models must be cross-validated with new data before any real validity to the final models can be claimed. This is quite important as specification searches are completely “data-driven exploratory model fitting” and, as such, can capitalize on chance [44]. For example, in cases where equivalent models are encountered (see next section), such searches will only lead one to a list of feasible models and then it becomes the responsibility of the researcher to decide which model to accept as the best model. To date, no automated search can make such a decision for a researcher. As noted by Marcoulides et al. [30], as

long as researchers keep in mind that the best use of model searches is to narrow attention to a reduced list (a sort of top-ten list), the algorithms will not be abused in empirical applications. More research on these algorithms is clearly needed to establish which one works best with a variety of models. For now, we believe that the Tabu search is one of the best available automated specification search procedures to provide valuable assistance in modeling applications. Unfortunately, to date no commercially available program offers this automated search option, although many programs do provide researchers with some rudimentary options to conduct specification searches and improve model to data fit.

3 Equivalent Models

While equivalent models has also received considerable methodological attention over the past couple of decades, it does not seem to be well understood or even considered by applied researchers using modeling techniques (e.g., [45–58]). Equivalent models are a set of models that yield identical (a) implied covariance, correlation and other observed variable moment matrices when analyzing the same data, (b) residuals and fitted moment matrices, and (c) fit function and related goodness-of-fit indices (e.g., chi-square values and p -values). Distinguishing between equivalent models cannot be achieved simply by using any currently available fit indices. Model equivalence can only be realistically managed via substantive considerations and/or considerations pertaining to design and data collection features (apart from the case of multiple-population versions of single-group equivalent models, where statistical distinction becomes possible with appropriate group constraints, if substantively correct [54]).

Two hypothesized models (denoted simply as M_1 and M_2), would be considered equivalent if the model implied covariance or correlation matrices are identical (which can be written as $\hat{\Sigma}_{M_1} = \hat{\Sigma}_{M_2}$). Let us consider this notion in the following equation:

$$\Sigma(\underline{\theta}) = \Lambda \Phi \Lambda' + \Theta = \Lambda(I_q - B)^{-1} \Psi (I_q - B')^{-1} \Lambda' + \Theta \quad (1)$$

where $\Sigma(\underline{\theta})$ is the model implied matrix (i.e., either $\hat{\Sigma}_{M_1}$ or $\hat{\Sigma}_{M_2}$), Λ the factor loading matrix, Λ' its transpose, Φ the factor correlation matrix, Θ is the error matrix, B is the matrix of structural regression coefficients relating the latent variables between themselves, Ψ is the covariance matrix of the structural regression residuals, and I_q is the $q \times q$ identity matrix (where q is the number of latent variables in the model, with the usual assumption the matrix $I_q - B$ is full rank). This equation implies that different matrices appearing in its right-hand side may lead to identically reproduced covariance/correlation matrices in its left-hand side. This is because from the sums and products of the matrices one cannot uniquely deduce the individual matrices on the right-hand side of the equation.

This statement highlights the fact that model equivalence is not defined by the data, but rather by an algebraic equivalence between hypothesized model

parameters. In turn, because of this model equivalence, the values of any considered statistical tests or goodness-of-fit indices of model fit will always be identical. Thus, even when a hypothesized model fits well according to the examined fit criteria, there can still be other equivalent models with identical fit—even if the theoretical implications or substantive interpretations of those models are radically different. In fact, as presented by Raykov and Marcoulides [55], there may even potentially be an infinite series of equivalent models to an initially hypothesized one. Identifying equivalent models can be a very difficult and time-consuming task. But there is clearly a compelling reason for undergoing such a difficult activity. Unfortunately, many researchers conducting various modeling activities do not seem to realize that alternative models might exist and that these others need to be considered.

A number of researchers have proposed a taxonomy that can be used to distinguish among several different types of equivalent models: namely, *observationally equivalent* and *covariance equivalent* (see, e.g., [49, 53, 57, 59] to name but a few). Two models are considered *observationally equivalent* only if one model can generate every probability distribution that the other model can generate. Observational equivalence is model equivalence in the broadest sense, and can be shown using data of any type. In contrast, models are considered *covariance equivalent* if every covariance (correlation) matrix generated by one model can be generated by the other. Thus, observational equivalence encompasses covariance (model) equivalence; that is, observational equivalence requires the identity of individual data values, whereas covariance equivalence requires the identity of summary statistics such as covariances and variances. We note that observationally equivalent models are always going to be covariance equivalent, whereas covariance equivalent models might not necessarily be observationally equivalent. Additional distinctions made include the mathematical notions of global and local equivalence, thereby signifying *globally equivalent* models and *locally equivalent* models. For two models to be globally equivalent, a *function* must exist that translates *every* parameter of one model into the parameters of another model. If only a *subset* of one model's parameter set is translatable into the parameter set of another model, the models are then considered *locally equivalent*. Local equivalence does not guarantee that the implied covariance or correlation matrices of the two models will be the same.

Categorizations of strategies for approaching the problem of equivalent models that have been considered in the extant literature include: (i) those that occur either before data collection, and (ii) those after data collection. The strategies consist of the four rules developed by Stelzl [60] and the more general rule by Lee and Hershberger [47], those based on graph theory that translate the model relationships into statistical relations (see [29, 61, 62]), the rank matrix approach that uses the rank of the matrix of correlations among the parameters of the proposed model [63], the data mining type automated heuristic searches [42], the information complexity criterion (ICOMP, [58]) with the one providing the lowest value representing the least complex of models, those that use computational problems associated with model misspecification as a way to distinguish among equivalent models (e.g., [64]), the comparison of the R^2 values among models [6], and the examination of extended individual case residuals (EICR, [65]). Of course, as expected, each of these

strategies has their proponents and opponents. Regardless of which approach is used, we believe that the consideration of equivalent models must become a standard part of the process of defining, testing, and modifying models. Unfortunately, and despite nearly decades of reports arguing convincingly for the importance of model equivalence in the model-fitting enterprise, to date many researchers do not take the extra effort to even consider the potential presence of equivalent models. We strongly believe that researchers must take the initiative and effort required to thoroughly examine the potential presence of equivalent models. We also strongly believe that replication and cross-validation of models are additional essential activities when utilizing such advanced modeling techniques. Perhaps the best affirmation of this ideology was provided by Scherr, who declared that

...the glorious endeavor that we know today as science has grown out of the murk of sorcery, religious ritual, and cooking. But while witches, priests, and chefs were developing taller hats, scientists worked out a method for determining the validity of their results: they learned to ask: Are they reproducible ([66], p. ix)?

4 Sample Size Issues

The issue of sample size and model identification are very different and separate issues. We emphatically declare that even a study using a large fraction from a population of interest (e.g., $N = 10,000$) may still posit a model for which parameters cannot be determined. This is because the issue of model estimation is closely tied to the issue of model identification and not sample size. Sample size is tied to power and stability of estimates whereas model identification is tied to existence and uniqueness of a solution (see details provided in the next section). This appears to be an issue that many researchers regrettably confuse as equivalent when it is in fact not at all comparable.

The recent PLS and related modeling literature is replete with examinations and discussions (some bad, some good) concerning the performance of PLS analyses with various sample sizes (e.g., [67–75]). Indeed, and despite popular belief, the evidence is quite clear that PLS like any other statistical technique is in no way immune to the distributional assumption concerning the need for an adequate sample size. This goes back to Hui and Wold [72] who determined that PLS estimates improved and their average absolute error rates diminished as sample sizes increased. Similarly, Chin and Newsted [70] determined that small sample sizes (e.g., $N = 20$) do not permit a researcher to detect low valued structural path coefficients (e.g., 0.20) until much larger sample sizes (i.e., between $N = 150$ and $N = 200$) are reached. Small sample sizes could only be used with higher valued structural path coefficients (e.g., 0.80), and even then “... with reasonably large standard errors...” ([70], p. 333). Similarly, Marcoulides and Saunders [73], Chin and Dibbern [76] and Chin [77] all noted the deleterious impact of non-normal data on PLS estimates and the need for markedly large sample sizes. Ultimately, a researcher needs to consider the distributional characteristics of the data, the potential presence of missing data, the

psychometric properties of the variables included in the model, and the magnitude of the relationships considered before definitely deciding on an appropriate sample size to use.

These results and recommendations corroborate Wold's earlier writings and theorems in which he indicated that PLS estimates

... are asymptotically correct in the joint sense of consistency (large number of cases) and consistency at large (large number of indicators for each latent variable ... [78], p. 266),

implying in the statistical sense that estimation error decreases as N increases (i.e., as $N \rightarrow \infty$, the estimation error tends to 0), or simply that any estimated PLS coefficients will converge on the parameters of the model as both sample size and number of indicators in the model become infinite (see also Falk and Miller [79]; McDonald [80]). This same statistical interpretation and recommendation is provided by Hui and Wold [72] who indicate that PLS "estimates will in the limit tend to the true values as the sample size N increases indefinitely, while at the same time the block sizes increase indefinitely but remain small relative to N (p. 123)."

Lu [81] and Lu, Thomas, and Zumbo [82] have also warned researchers about the bias that arises from a failure to use large number of indicators for each latent variable (i.e., consistency at large) and labeled it "finite item bias." Dijkstra [83] and Schneeweiss [84] provided some discussion about the magnitude of standard errors for PLS estimators resulting from not using enough observations (consistency) and indicators for each latent variable (consistency at large). Schneeweiss [84] also provided closed form equations that can be used to determine the magnitude of finite item bias relative to the number of indicators used in a model. Using these equations, Schneeweiss ([84], p. 310) indicated that item bias is generally small when many indicators, "each with a sizeable loading and an error which is small and uncorrelated (or only slightly correlated) with other error variables" are used to measure each factor in the model. These warnings clearly echo well established concerns that a determination of the appropriate sample size (which depends on many factors) is also an essential aspect of the whole modeling process.

Although sample size plays an important role in almost every statistical technique applied in practice and there is universal agreement among researchers that the larger the sample the more stable the parameter estimates, there is no agreement as to what constitutes large. This topic has received much attention in the broad statistical literature, but no easily applicable and clear-cut criteria have been determined, only some general rules of thumb have been proposed. For example, some researchers cautiously suggested the general rule of thumb that the sample size should always be more than 10 times the number of free model parameters [85, 86].

To complicate matters, due to the partial nature of the PLS algorithm, the total number of free model parameters should not be the basis for sample size requirements. Being a components based approach, sample size requirements may differ in terms of obtaining stable component weights, measurement paths, and structural model paths. Chin [10] suggested that a researcher using the PLS path weighting scheme should examine the largest of two possibilities: (a) the block

with the largest number of formative indicators (i.e., the largest so-called mode *B* measurement equation) or (b) the dependent variable with the largest number of independent variables impacting it (i.e., the largest so-called structural equation). Chin [10] then concluded by saying

If one were to use a regression heuristic of 10 cases per predictor, the sample size requirement would be 10 times either (a) or (b), whichever is the greater (p. 311, emphasis added).

Here Chin used the example heuristic rule of 10 in conjunction with the path weighting scheme. But the main focus was on considering how to determine the largest regression analysis during the PLS iterative algorithm for estimating required sample size for obtaining stable estimates for either (1) weights for PLS components or (2) model paths (i.e., measurement and structural estimates).

Many researchers seem unaware that the equations for a PLS analysis can change depending on the choice of the inner weighting scheme and that the weight estimates for the PLS components are not necessarily affected by the structural model. Chin [10] noted that

If one is not using a path-weighting scheme for inside approximation, then only the measurement model with formative indicators are considered for the first stage of estimation. At the extreme, we see that a factor- or centroid-weighting scheme with all reflective (mode A) measures will involve only a series of simple regressions. Under this condition, it may be possible to obtain stable estimates for the weights and loadings of each component independent of the final estimates for the structural model (p. 311).

Unfortunately, many applied researchers without adequate statistical understanding of the PLS algorithm have unreflectively applied the example rule of 10 that Chin [10] provided. Beyond identifying the constraining regression equation in a PLS analysis, Marcoulides et al. [74] also noted that it seems there is a “reification of the 10 case per indicator rule of thumb (p. 174)” by most PLS researchers ignoring Chin and Newsted’s [70] statement that

for a more accurate assessment, you would specify the effect size for each regression analysis and look up the power tables provided by Cohen ’[87] or Green’s ’[88] ‘approximation to these tables’ (p. 327).

Clearly, many other researchers (e.g., [3, 70, 73, 74, 89–93]) have indicated that no rule of thumb can be applied indiscriminately to all situations. This is because the appropriate size of a sample depends on the many other factors noted earlier. When these issues are carefully considered, samples of varying magnitude may be needed to obtain reasonable parameter estimates.

In spite of these cautiously proposed rules of thumb available in the PLS literature, there continue to be sweeping claims made by some researchers that PLS modeling can be or should be used (and often, instead of the covariance-based approach) because it makes no sample size assumptions or because “... sample size is less important in the overall model... ([79], p. 93).” Unfortunately, some of these studies even appear in top-tiered journals and frequently report results based on ridiculously low sample sizes, despite the overall inferential intentions of the studies and the actually magnitude of the parent populations of interest. To make things

worse, they also try to legitimize these actions by making references to the original developers of the PLS approach. Even a cursory preview of articles using PLS over the past decade reveal a plethora of problematic comments concerning sample size. Recently Ringle, Sarstedt, and Straub [94] documented a sizable number of articles published in MISQ (one of the top-tiered information systems journals) that reported using PLS due to small sample size. Included in some of those articles, were comments such as: (i) “the PLS approach does not impose sample size restrictions ... for the underlying data ... ([95], p. 237),” (ii) “... PLS, a component based approach that is suitable with smaller data sets ... ([96], p. 685),” and (iii) “... PLS ... provides the ability to model latent constructs even under conditions of non-normality and small- to medium-size samples ... ([97], p. 49).” To be fair to these authors, similar troubling comments concerning sample size in PLS modeling abound in almost every other substantive area we examined, consequently the issue is not unique to the information systems field.

All three of the above mentioned articles reported on results from studies in which they had examined and fulfilled the general 10 cases per indicator rule of thumb mentioned above. Specifically, in the Bhattacharjee and Premkumar’s study [95], the largest number of indicators per construct in the confirmatory factor analysis (CFA) conducted was 4 and the authors reported using samples sizes between 54 and 77 (depending on the specific construct examined, see p. 237). The Basselier and Benbasat’s study [96] also conducted a CFA using 109 observations and 3–4 item scales. Finally, the Subramani’s study [97] used 131 observations in a CFA with 3–4 item scales and the largest number of paths to any construct was 6. Thus, all three above mentioned studies followed the general rule of thumb guidelines regarding sample size.

Nevertheless, as discussed earlier, the generic rule of thumb of 10 cases per indicator does not always ensure accurate and sufficiently stable estimates. So is it the case that many PLS users simply ignore essential preliminaries with regards to sample issues when using these methods in practice? Why is it that many do not seem to carefully examine model parameters along with indexes of their stability across repeated sampling from the studied population? These indexes—the parameter standard errors—also play an instrumental role in constructing confidence intervals for particular population parameters of interest (e.g., [98–101]). Is it not obvious to them that models estimated using questionable sample sizes with extremely unstable estimates and wielding huge standards errors and confidence intervals should be sufficient evidence for an investigator to question the generalizability of results and validity of conclusions drawn? Questionable sample sizes can also cause standard errors to be either overestimated or underestimated. Overestimated standard errors can also result in significant effects being missed, while underestimated standard errors may result in overstating the importance of effects ([73, 93]).

In order to determine the precision of estimation and find standard errors, two approaches can be considered: (a) using analytic approaches (such as the delta or Taylor series expansion method, see, e.g., [98, 101, 102]) or (b) using computer-intensive re-sampling methods (see, e.g., [10, 27]). Unfortunately, finding formulas for standard errors of PLS estimates using the delta or Taylor series expansion

method is not a trivial task [83], but recent work [98, 101, 102] has proved promising. As a consequence, Monte Carlo simulation re-sampling methods continue to dominate the field. The principle behind a Monte Carlo simulation is that the behavior of a parameter estimate in random samples can be assessed by the empirical process of drawing many random samples and observing this behavior.

There are, actually, two kinds of Monte Carlo re-sampling strategies that can be used to examine parameter estimates and related sample size issues. The first strategy can be considered a “reactive” Monte Carlo analysis (such as the popular Jackknife or Bootstrap approaches [10, 27, 73, 98]) in which the performance of an estimator of interest is judged by studying its parameter and standard error bias relative to repeated random samples drawn with replacement from the original observed sample data. This type of Monte Carlo analysis is currently quite popular (particularly the Bootstrap approach), despite the fact that it may often give “an unduly optimistic impression of accuracy or stability” of the estimates ([83], p. 86) and there are no generally applicable results as yet of how good the underlying approximation of sampling by pertinent re-sampling distributions is within the framework of latent variable modeling [102].

The second, a less commonly known strategy, can be considered a “proactive” Monte Carlo simulation analysis [25, 73, 103, 104]. In a proactive Monte Carlo analysis, data are generated from a population with hypothesized parameter values and repeated random samples are drawn to provide parameter estimates and standard errors. The approach can also be used in a reactive manner to judge obtained estimates of parameter values and determine the magnitude of standard errors based upon the sample size actually used in a study. Thus, a proactive Monte Carlo analysis (sometimes also referred to as a power analysis) can be used to both examine parameter estimate precision and the necessary sample size needed to ensure the precision of parameter estimates [73, 93].¹ To date, only a few IS articles have conducted Monte Carlo based PLS power analyses (e.g., [105, 106]).

It is surprising to note that some researchers believe that such power analyses are not useful after a study has been completed. For example, Walden [107] proclaimed that “... no one should ever ask for an after the fact power analysis on a sample that shows results.” He believes this because a

power analysis asks ... how big does a sample need to be to detect an effect of a certain size with some probability This question does not make sense after the sample has been collected and the null hypothesis rejected, for several reasons ... there is no probability to be evaluated ... if an effect is observed, the sample is clearly large enough to observe an effect ... if you detected an effect, you had the power you needed. (March 5, 2012, AIS World).

Walden [107] seems to have forgotten that researchers always conduct statistical hypothesis testing under conditions of uncertainty. In other words, researchers

¹ These generation of Monte Carlo data can easily be done using the statistical analysis program *Mplus* [8]. *Mplus* has a fairly easy-to-use interface and offers researchers a flexible tool to analyze data using all kinds of model choices. Detailed illustrations for using the *Mplus* Monte Carlo simulation options can also be found in [93], in the *Mplus User's Guide* [8], and at the product Web site www.statmodel.com

makes a decision about the “true state of affairs” in a studied population, based on information only from part of it (the sample) which typically is a fairly small fraction of the population of interest. Because one functions in this situation of uncertainty, the decision may be incorrect. This is the reason one may commit one of two types of errors—a Type I or a Type II error (one cannot commit both types of errors as they are mutually exclusive possibilities [108]). Hence, even when there may appear to be overwhelming evidence supporting a null hypothesis, as long as the sample is not identical to the population one can never claim to have definitively proved the validity of the null hypothesis. Such a scenario can only occur when the entire population of interest is exhaustively studied. Any time sample data are used there is no guarantee that a null hypothesis that is rejected based on the observed data, is actually true in the population. Thus, one always runs the risk of committing an error. By at least determining the magnitude of the power of a test of the null hypothesis [i.e., determining $(1 - \beta)$, which is the complement to the probability of making a Type II error] one can gain some probabilistic insight into any decision about the true state of affairs.

Looking at a number of simple power analyses in studies that employed PLS modeling techniques, one can quickly deduce the importance of power analyses and the fallacy of Walden’s [107] argument. To illustrate this point, let us first consider a confirmatory factor analysis (CFA) model in which two correlated factors (ϕ_{21}), each of which has three continuous factor indicators and the following factor loading $\Lambda = [\lambda_{11}, \lambda_{21}, \lambda_{31}, \lambda_{42}, \lambda_{52}, \lambda_{62}]$ and error variance $\Theta = [\theta_{11}, \theta_{22}, \theta_{33}, \theta_{44}, \theta_{55}, \theta_{66}]$ matrix structures. Assume that the data are generated with varying values for the factor inter-correlations (ϕ_{21} between 0.1 and 0.9), each factor loading (λ between 0.4 and 0.9), for the error variances (θ between 0.19 and 0.84) and, consequently, for the indicator reliabilities (between 0.16 and 0.81) and examined with both normal and non-normal distributions. The non-normal data are generated under conditions of moderate non-normality (i.e., skewness set to range between 1 and 1.5, and kurtosis set to range between 1 and 1.5). For ease of presentation, no missing data patterns are considered. To ensure stability of results, the number of sample replications is set at 5,000. To simplify matters further, we focus only on the factor correlation parameter (ϕ_{21}), although any other model parameter could be similarly examined (for additional details see [73]).

Table 1 presents the results of a Monte Carlo simulation based on a pre-selected $N = 100$ sample size. The boldfaced column values correspond to the various factor loadings considered (i.e., the values of λ), while the boldfaced row values correspond to the considered factor inter-correlations (i.e., the values of ϕ_{21}). The entries provided in Table 1 correspond to the computed value of the power of the study to reject the hypothesis that the factor correlation in the population is zero (i.e., the probability of rejecting the null hypothesis when it is actually false). As can be seen by examining the entries provided in Table 1, power remains relatively high when indicators with sizeable factor loadings (and thereby more reliable indicators) are used to measure factors. For example, a power value of 0.97 is achieved when indicators with factor loadings equal to 0.90 are used to examine a 0.50 valued correlation between the two factors. Even so, power estimates deteriorate when examining low valued factor correlations, especially when using poor quality indicators. For exam-

Table 1: Power values determined for normally distributed data with no missing values ($N = 100$)

	λ					
ϕ_{21}	0.9	0.8	0.7	0.6	0.5	0.4
0.1	0.13	0.12	0.13	0.13	0.10	0.06
0.2	0.46	0.31	0.26	0.24	0.18	0.10
0.3	0.85	0.82	0.71	0.48	0.36	0.21
0.4	0.97	0.96	0.90	0.81	0.57	0.27
0.5	0.97	0.96	0.96	0.92	0.77	0.41
0.6	1.00	1.00	1.00	0.96	0.91	0.53
0.7	1.00	1.00	1.00	0.99	0.93	0.66
0.8	1.00	1.00	1.00	1.00	0.97	0.77
0.9	1.00	1.00	1.00	1.00	1.00	0.82

ple, a power value of only 0.10 is achieved when indicators with factor loadings equal to 0.50 are used to examine a 0.10 valued factor inter-correlation.

Table 2: Power values determined for normally distributed data with no missing values ($N = 50$)

	λ					
ϕ_{21}	0.9	0.8	0.7	0.6	0.5	0.4
0.1	0.11	0.11	0.12	0.13	0.15	0.09
0.2	0.29	0.27	0.24	0.22	0.21	0.13
0.3	0.59	0.52	0.45	0.37	0.30	0.22
0.4	0.87	0.78	0.70	0.62	0.46	0.24
0.5	0.97	0.93	0.87	0.72	0.54	0.30
0.6	1.00	0.99	0.94	0.88	0.70	0.46
0.7	1.00	1.00	1.00	0.93	0.74	0.50
0.8	1.00	1.00	1.00	0.97	0.83	0.57
0.9	1.00	1.00	1.00	1.00	0.87	0.58

Table 2 presents the results in which a much smaller pre-selected $N = 50$ sample size is used. As can be seen from these results, power again tends to remain relatively high when psychometrically sound indicators measure factors, particularly when examining very high valued factor correlations. Tables 3 and 4 present the results of Monte Carlo simulations for the same two sample sizes ($N = 50$ and $N = 100$) but instead under conditions of non-normality. Unfortunately, when the power values are examined under such conditions of non-normality, their deterioration is evident and quite disconcerting. In fact, none of the values provided in Tables 3 and 4 are above 0.40, indicating that using these sample sizes a researcher would just not be able to reject false null hypotheses concerning the factor inter-correlation.

So what samples sizes would be needed to achieve a sufficient level of power, say equal to 0.80 (considered by most researchers as acceptable power)? The results of such a Monte Carlo analysis under conditions of normality and non-normality are

Table 3: Power values determined for non-normally distributed data with no missing values ($N = 50$)

	λ					
ϕ_{21}	0.9	0.8	0.7	0.6	0.5	0.4
0.1	0.11	0.09	0.11	0.12	0.09	0.04
0.2	0.12	0.12	0.11	0.12	0.10	0.05
0.3	0.14	0.14	0.13	0.15	0.11	0.06
0.4	0.17	0.15	0.15	0.16	0.12	0.08
0.5	0.19	0.19	0.19	0.19	0.15	0.08
0.6	0.20	0.22	0.21	0.22	0.16	0.11
0.7	0.23	0.23	0.23	0.24	0.20	0.12
0.8	0.25	0.26	0.26	0.27	0.23	0.12
0.9	0.30	0.31	0.30	0.32	0.26	0.18

Table 4: Power values determined for non-normally distributed data with no missing values ($N = 100$)

	λ					
ϕ_{21}	0.9	0.8	0.7	0.6	0.5	0.4
0.1	0.08	0.07	0.08	0.11	0.10	0.04
0.2	0.09	0.07	0.09	0.11	0.11	0.06
0.3	0.14	0.11	0.11	0.12	0.12	0.07
0.4	0.17	0.15	0.12	0.12	0.12	0.09
0.5	0.20	0.21	0.16	0.17	0.15	0.10
0.6	0.25	0.25	0.24	0.19	0.16	0.12
0.7	0.28	0.28	0.26	0.27	0.21	0.14
0.8	0.33	0.34	0.33	0.29	0.24	0.18
0.9	0.40	0.38	0.36	0.36	0.27	0.21

provided in Tables 5 and 6. As can be seen by examining the entries in Table 5, relatively small sample sizes can often be used when psychometrically sound indicators are available to examine high valued factor inter-correlations. However, when trying to examine low valued factor inter-correlations using poor quality indicators, much larger sample sizes are needed. It is important to note that the results presented in Table 5 corroborate those presented by Hui and Wold [72], Chin and Newsted [70], and Schneeweiss [84] that small sample sizes do not permit a researcher to detect low valued model coefficients until much larger sample sizes are reached. However, the problem can be much more disconcerting than these researchers originally reported, as can be seen by examining the entries in Table 6. When moderately non-normal data are considered, the sample sizes needed sometimes become astronomical, despite the inclusion of highly reliable indicators in the model. These results are evidence that determining the appropriate sample size even for a simplistic CFA depends on many model characteristics, including the psychometric properties of the indicators, the strength of the relationships among the factors, and the distributional characteristics of the data. It is also important to note that for only a limited number of normally distributed data conditions would the PLS rule of thumb of 10

cases per indicator really suffice in these example models considered. As indicated previously, a researcher must consider the distributional characteristics of the data, potential missing data, the psychometric properties of the variables examined, and the magnitude of the relationships considered before deciding on an appropriate sample size to use or to ensure that a sufficient sample size is actually available to study the phenomena of interest.

Table 5: Sample sizes needed to achieve power = 0.80 with normally distributed data and no missing values

	λ					
ϕ_{21}	0.9	0.8	0.7	0.6	0.5	0.4
0.1	916	1,053	1,261	1,806	2,588	4,927
0.2	256	292	371	457	764	1,282
0.3	96	99	147	223	317	672
0.4	46	57	71	98	186	343
0.5	25	34	43	66	111	220
0.6	16	20	23	44	78	175
0.7	15	15	17	33	61	134
0.8	15	15	17	25	46	109
0.9	15	15	17	25	42	99

Table 6: Sample sizes needed to achieve power = 0.80 with non-normally distributed data and no missing values

	λ		
ϕ_{21}	0.9	0.8	0.7
0.1	15,646	24,574	31,381
0.2	4,922	5,766	6,251
0.3	2,357	2,623	2,817
0.4	1,331	1,536	1,715
0.5	931	1,018	1,203
0.6	653	707	864
0.7	467	545	639
0.8	386	433	486
0.9	345	351	407

It is also quite useful to examine in detail the previously mentioned studies and determine the extent to which they exhibited a sufficient level of power to support the results and validity of conclusions drawn. We note that all of these studies indicated they had examined and fulfilled the 10 cases per indicator rule of thumb. Table 7 presents a power analysis of the CFA model from the study by Bhattacharjee and Premkumar [95]. Two values are provided in each cell of Table 7, the inter-construct correlation reported in the published study (for full details see [95], Table 2, p. 239), and the determined level of power for each examined inter-correlation.

Because no specific details were provided in the published study about the distributional characteristics of the data or any apparent missing data patterns, the power analyses were conducted by assuming normally distributed data with no missing data patterns. The factor loadings reported for the scaled items in the study were all in the 0.80–0.96 range and the sample sizes were between 54 and 77 observations, depending on the construct examined (see Table 1, p. 238). As can be seen by examining the entries in Table 7, and assuming normally distributed data with no missing data patterns, power would be considered quite high for all the inter-construct correlations examined in the Bhattacharjee and Premkumar’s study [95].

Table 8 presents a power analysis of the CFA from the study by Bassellier and Benbasat [96]. Once again, two values are provided in Table 8. The inter-correlations among the constructs reported in the published study (see [96], Table

Table 7: CFA inter-construct correlations and power values for Bhattacharjee and Premkumar [95] (N = 54)

CBT study (time t2 – t3)	U2	A2	D3	S3	U3	A3
Attitude (A2)	0.74 ^a (0.94) ^b					
Disconfirmation (D3)	0.45 (1.00)	0.44 (0.94)				
Satisfaction (S3)	0.59 (1.00)	0.58 (0.100)	0.46 (1.00)			
Usefulness (U3)	0.65 (1.00)	0.61 (1.00)	0.60 (1.00)	0.64 (1.00)		
Attitude (A3)	0.63 (1.00)	0.69 (1.00)	0.54 (0.99)	0.80 (1.00)	0.71 (1.00)	
Intention (I3)	0.63 (1.00)	0.58 (1.00)	0.52 (0.97)	0.53 (0.98)	0.79 (1.00)	0.61 (1.00)

^aInter-construct correlation

^bPower

Table 8: CFA inter-construct correlations and power values for Bassellier and Benbasat [96] (N = 109)

CBT study	1	2	3	4	5	6	7
1. Intentions for partnerships							
2. Organizational overview	0.406 ^a (0.97) ^b						
3. Organizational unit	0.352 (0.89)	0.809 (1.00)					
4. Organizational responsibility	0.504 (1.00)	0.601 (1.00)	0.676 (1.00)				
5. IT-business integration	0.516 (1.00)	0.628 (1.00)	0.589 (1.00)	0.595 (1.00)			
6. Knowledge networking	0.283 (0.73)	0.453 (0.99)	0.405 (0.96)	0.308 (0.81)	0.341 (0.87)		
7. Interpersonal communication skills	0.381 (0.94)	0.485 (0.99)	0.386 (0.95)	0.345 (0.88)	0.478 (0.98)	0.474 (0.98)	
8. Leadership skills	0.427 (0.97)	0.589 (1.00)	0.600 (1.00)	0.516 (1.00)	0.652 (1.00)	0.517 (1.00)	0.582 (1.00)

^aInter-construct correlation

^bPower

6, p. 688), and the appropriately determined level of power for each examined inter-correlation. Because this study also did not provide any specific details about the distributional characteristics of the data or any apparent missing data patterns, the analyses were conducted under the assumption of normality and no missing data patterns. The factor loadings reported for scaled items in the study were all in the 0.71–0.89 range and the sample size was 109 observations (see Table 5, p. 687). As can be seen from the entries in Table 8, and assuming normally distributed data with no missing data patterns, power would also be considered quite high for almost all the inter-construct correlations examined in the Bassellier and Benbasat's study [96]. The only power value that is lower (0.73) than the commonly accepted cutoff point of 0.80 is for the correlation between the constructs of "knowledge networking (#6)" and "intentions for partnerships (#1)."

Finally, Table 9 presents a power analysis of the CFA from the study by Subramani [97]. The same two values are provided in the table; the inter-construct correlations reported in the published study (see [97]; Table 2, p. 61) and the determined level of power for each examined correlation. As with the previously examined studies, this study also did not provide any specific details about the distributional characteristics of the data or missing data patterns. As such, the analyses were again conducted under the assumption of normality and no missing data patterns. The factor loadings were not specifically reported for scaled items in the study, but were apparently "uniformly high (p. 59)" and between 0.71 to "above 0.80 (p. 59)," using a sample with 131 observations. As can be seen by examining the entries in Table 9, and assuming normally distributed data with no missing data patterns, power would be considered quite low (and in some cases well below the commonly accepted cutoff point of 0.80), even for many of the reported statistically significant inter-construct relationships. For example, although the relationship between the Operational Benefits (#5) and IT USE for Exploitation (#1) was reported as being statistically significant (0.179, $p < 0.05$), its power value on the basis of the proactive Monte Carlo simulation analysis is determined to be equal to 0.40 (assuming of course normally distributed data with no missing values—if in fact, the data were not normally distributed a much lower power value would be expected). In other words, the computed value of the power of the study to reject the hypothesis that the factor inter-correlation in the population is zero was determined to be quite low. It is particularly important to note that, despite the fact that Subramani's study [97] utilized a larger sample size than either of the other two previously examined studies (which as we saw exhibited sufficiently high levels of power), the low valued factor correlations examined in Subramani's study [97] actually deteriorated the power of the statistical tests conducted. And, although Subramani's study [97] clearly fulfilled the frequently used rule of thumb of using 10 cases per indicator, it appears that the generalizability of some of the results and the validity of the conclusions drawn from this study may be questionable.

As indicated by Marcoulides and Saunders [73] the selection of an appropriate sample size that will ensure an adequate level of power clearly depends on many factors. These include the psychometric properties of the variables considered, the strength of the relationship among the variables, the complexity and size of the

Table 9: CFA inter-construct correlations and power values for Subramani [97] ($N = 131$)

CBT study	1	2	3	4	5	6	7	8	9	10
1. IT use for exploitation										
2. IT use for expiration	0.188 ^a (0.41) ^b									
3. Business process specificity	0.321* (0.89)	0.039 (0.09)								
4. Domain-knowledge specificity	0.329* (0.89)	0.468* (1.00)	0.200 (0.49)							
5. Operational benefits	0.179* 0.40	0.343* (0.96)	0.163 (0.32)	0.550* (1.00)						
6. Strategic benefits	0.258* (0.70)	0.352* (0.96)	0.257* (0.69)	0.410* (0.99)	0.489* (1.00)					
7. Competitive performance	0.086 (0.16)	0.005 (0.07)	0.013 (0.07)	0.158 (0.30)	0.173 (0.36)	0.274* (0.77)				
8. Uncertainty	0.049 (0.09)	0.198 (0.14)	0.077 (0.14)	0.197 (0.46)	-0.028 (0.08)	-0.044 (0.08)	0.131 (0.23)			
9. Retailer replaceability	0.028 (0.09)	0.070 (0.14)	-0.100 (0.22)	-0.111 (0.23)	-0.230** (0.72)	-0.310** (0.90)	-0.670** (1.00)	0.130 (0.23)		
10. Size	0.132 (0.24)	0.165 (0.32)	-0.152 (0.26)	0.205* (0.52)	-0.019 (0.07)	-0.015 (0.07)	-0.01* (0.07)	0.328** (0.89)	0.190** (0.42)	
11. Years of association	0.137 (0.24)	0.117 (0.23)	-0.187* (0.56)	0.097 (0.21)	-0.002 (0.08)	0.055 (0.10)	0.007 (0.09)	0.133 (0.25)	-0.069 (0.14)	0.320** (0.89)

Shaded cells are correlations reported significant (but below the 0.80 power threshold)

* $p < 0.05$, ** $p < 0.01$

^aInter-construct correlation

^bPower

model, the amount of missing data, and the distributional characteristics of the variables. Examining all these issues using a proactive Monte Carlo simulation analysis will at least provide researchers with some insight concerning the stability and power of the parameter estimates that would be obtained across repeated sampling from the studied population. Ignoring these issues could lead to important effects being completely missed in a study or lead to overstating the importance of effects in a study.

5 Model Identification Issues

The examination of issues related to model identification began in the early part of the last century with the work of Albert [109, 110], Koopmans and Reiersol [111], and Ledermann [112]. Identification basically consists of two specific aspects: existence and uniqueness (e.g., [63, 113, 114]). For example, in the context of a factor analysis model, *existence* and *uniqueness* would imply the following: (1) *Existence*: Does a factor decomposition (e.g., $\Sigma = \Lambda\Lambda' + \Psi$) exist in the population (for a given number of factors m), where Λ is a $p \times m$ factor loading matrix with rank m , and Ψ is a unique variance diagonal (error) matrix with positive elements? (2) *Uniqueness*: Assuming the existence of a factor decomposition, is it the *only* decomposition possible? In other words, are there no other matrices (e.g., some other matrices such as Λ_2 and Ψ_2 different from Λ and Ψ) that can give the same matrix Σ (i.e., $\Sigma = \Lambda_2\Lambda_2' + \Psi_2$, with rank of Λ_2 not greater than m)?

Model identification can also be categorized in one of two ways: (1) *global identification*, in which *all* of a model's parameters are identified, or in contrast, (2) *local identification*, in which at least one—but not all—of a model's parameters is identified. Globally identified models are locally identified, but locally identified models may or may not be globally identified. Global identification is a prerequisite for drawing inferences about an entire model. When a model is not globally identified, local identification of some of its parameters permits inferential testing in only that section of the model. Some researchers have referred to this as *partial identification* [28]. Kano [115] has suggested that we focus more attention on examining what he referred to as *empirical identification*, rather than the more commonly considered notion of *mathematical identification*. At least in theory (although somewhat debatable in practice) parameters that are not identified do not influence the values of ones that are identified.

In addition to categorizing models as either globally or locally identified, they can also be classified as (1) *under-identified*, (2) *just-identified*, or (3) *over-identified*. Consider for example a model involving four ($P = 4$) observed variables. Such a model would be determined as having a correlation or covariance data matrix with altogether $\frac{1}{2}P(P + 1) = \frac{4 \times 5}{2} = 10$ nonredundant elements. Now let us consider the case of an under-identified model. Under-identification occurs when not enough relevant data are available to obtain unique parameter estimates. Using the notion of the degrees of freedom of any hypothesized model as the difference between the

number of nonredundant elements in the data matrix and the number of parameters in the model, an under-identified model will have negative degrees of freedom. We note that when the degrees of freedom of a model are negative, at least one of its parameters is under-identified. Having positive degrees of freedom with any proposed model is a necessary but not a sufficient condition for identification. That is because having positive degrees of freedom does not guarantee that every parameter is identified. There can in fact be situations in which the degrees of freedom for a model are quite high (the so-called over-identified case) and yet some of its parameters remain under-identified [100]. Conversely, having negative degrees of freedom is a sufficient but not a necessary criterion for showing that a model is globally under-identified.

Two additional and frequently interchangeably used concepts are those of a “saturated model” and of a “just identified” model. A just identified model can be defined as an identified model that has zero degree of freedom, while a saturated model can be defined as a model that has zero degree of freedom [116, 117]. Nevertheless, as noted by Raykov et al. [117], the distinction between these two models is quite important since using them interchangeably can lead to consequential theoretical and empirical confusion, with potentially misleading substantive conclusions. Because a saturated model need not be (just) identified, the two concepts must be kept separate. Raykov et al. [117] proposed that (a) the notion of “the saturated model” be reserved for a particular saturated model (the one with unconstrained variable variances and covariances), and (b) that the reference “a saturated model” be used when the pertinent statement would be correct for any saturated model for that set of observed variables.

If theory testing is the main objective, the most desirable identification status of a model is *over-identification*, where the number of available data elements is more than those needed to obtain a unique solution. Although as indicated above, having positive degrees of freedom does not guarantee that every parameter in the model is identified. An over-identified model thereby implies that, for at least one parameter, there is more than one equation the estimate of a parameter must satisfy; only under these circumstances—the presence of multiple solutions—are models provided with the opportunity to be rejected by the data.

Although identification issues have major implications with respect to model fitting, they are frequently ignored due to their challenging technical intricacies [28]. To simplify matters, some researchers often make a specific assumption about existence and focus mainly on uniqueness aspects. For example, existence in factor analysis implies that factor decomposition exists for a given number of factors, whereas uniqueness assumes it is the only decomposition possible—in other words, it is commonly assumed that a factor decomposition does exist in the population of interest. The topic of model uniqueness has generally followed two early lines of research: one originating in Albert [109, 110] and Anderson and Rubin [118] and the other based on the work of Ledermann [112]—for a detailed overview see [28] and the references therein.

Anderson and Rubin [118] specifically proposed the following theorem (the so-called *Theorem 5.1*) in factor analysis for a *sufficient* condition of *uniqueness*:

Theorem 5.1. If any single row of a factor loading matrix Λ is deleted, there still remain two disjoint (i.e., non-overlapping) submatrices of rank m . Then the FA decomposition is *unique* (for a detailed proof of this theorem see [119]).

To illustrate, consider for example the following matrix:

$$\Sigma = \begin{pmatrix} 1 & 0.340 & 0.310 & 0.100 & 0.095 \\ 0.340 & 1 & 0.285 & 0.095 & 0.090 \\ 0.310 & 0.285 & 1 & 0.090 & 0.085 \\ 0.100 & 0.095 & 0.090 & 1 & 0.150 \\ 0.095 & 0.090 & 0.085 & 0.150 & 1 \end{pmatrix}$$

and a FA decomposition leading to a factor loading (Λ) matrix with rank $m = 2$:

$$\Lambda = \begin{pmatrix} 0.60 & 0.10 \\ 0.55 & 0.10 \\ 0.50 & 0.10 \\ 0.10 & 0.40 \\ 0.10 & 0.35 \end{pmatrix}$$

and a unique variance (Ψ) matrix equal to:

$$\Psi = \begin{pmatrix} 0.63 & 0 & 0 & 0 & 0 \\ 0 & 0.69 & 0 & 0 & 0 \\ 0 & 0 & 0.74 & 0 & 0 \\ 0 & 0 & 0 & 0.83 & 0 \\ 0 & 0 & 0 & 0 & 0.87 \end{pmatrix}.$$

If the first row in the factor loading matrix Λ were deleted to provide

$$\Lambda = \begin{pmatrix} 0.55 & 0.10 \\ 0.50 & 0.10 \\ 0.10 & 0.40 \\ 0.10 & 0.35 \end{pmatrix},$$

then the two possible disjoint submatrices

$$\begin{pmatrix} 0.55 & 0.10 \\ 0.50 & 0.10 \end{pmatrix} \text{ and } \begin{pmatrix} 0.10 & 0.40 \\ 0.10 & 0.35 \end{pmatrix}$$

would provide nonzero determinants, thereby signifying that there are two disjoint submatrices whose rank is $m = 2$ (we note that similar rank results would be ob-

tained for the remaining possible submatrices; for complete details and a step by step analysis for conducting such examination, including a SAS PROC IML subroutine, see Table 1 in [28]). Consequently, the factor analysis decomposition $\Sigma = \Lambda\Lambda' + \Psi$ is unique. In other words, based upon these results there is no alternative factor decomposition (such as $\Sigma = \Lambda_2\Lambda_2' + \Psi_2$) of the matrix Σ .

The above Anderson and Rubin [118] theorem essentially requires that the relationship between the number of observed variables (p) and number of factors (m) be satisfied as $p \geq 2m + 1$ (i.e., the number of observed variables p has to be greater than twice the number of factors m [115]). In other words, what *Theorem 5.1* implies is that if the number of factors (m) selected is greater than $(p - 1)/2$ observed variables, it will be difficult for the solution to be identified [120]. For example, if four factors were selected in a study with only eight observed variables, it will be difficult to get the solution to be identified. A number of other researchers (e.g., [121, 122]) have also provided alternative conditions to those proposed by Anderson and Rubin [118] and much research continues to date on this topic (e.g., [114, 120]).

Anderson and Rubin [118] also proposed other important theorems for *necessary* conditions of *uniqueness*, which they called *Theorems 5.5–5.7*. A particular theorem that has very important practical implications to the practice of factor analysis and related models is *Theorem 5.6*. This theorem states: If any rotated factor loading matrix (rotated by a nonsingular matrix) has a column with *at most two non-zero elements*, then the FA decomposition is *not unique* (and therefore is *not identified*).

In other words, if a researcher extracts a factor whose factor loading estimates are quite small and do not differ significantly from zero except for at most two elements, then it may be reasonable to suspect that the factor is not *uniquely* identified. The example, the population factor loading matrix Λ and its estimate $\hat{\Lambda}$ would be illustrative of such a not *uniquely* identified third factor (see [115], p. 143):

$$\Lambda = \begin{pmatrix} 0.57 & 0.13 & 0.00 \\ 0.57 & 0.33 & 0.00 \\ 0.21 & 0.37 & 0.54 \\ 0.75 & 0.07 & 0.00 \\ 0.73 & 0.07 & 0.00 \\ 0.29 & 0.25 & 0.54 \\ 0.19 & 0.45 & 0.00 \\ 0.18 & 0.33 & 0.00 \\ 0.10 & 0.71 & 0.00 \\ 0.31 & 0.43 & 0.00 \\ 0.02 & 0.42 & 0.00 \\ 0.03 & 0.53 & 0.00 \end{pmatrix} \quad \hat{\Lambda} = \begin{pmatrix} 0.58 & 0.13 & 0.02 \\ 0.59 & 0.33 & 0.03 \\ 0.24 & 0.35 & 0.05 \\ 0.74 & 0.61 & 0.01 \\ 0.72 & 0.07 & 0.01 \\ 0.31 & 0.23 & 0.55 \\ 0.20 & 0.45 & 0.02 \\ 0.17 & 0.33 & 0.02 \\ 0.12 & 0.71 & 0.03 \\ 0.30 & 0.40 & 0.01 \\ 0.01 & 0.43 & 0.02 \\ 0.03 & 0.45 & 0.02 \end{pmatrix}.$$

Consequently, a researcher should be very cautious whenever an estimated factor loading matrix looks anything like the one displayed on the right-hand side above.

Fortunately, most general statistics programs provide options to output the standard errors for rotated factor loadings and, consequently, a researcher can at least select to conduct hypothesis tests to determine whether the rotated factor loadings in

the population significantly differ from zero (although once again the issue of power and sample size considered in the above section would again come into play). For simultaneous hypothesis testing, it may be wise to employ Bonferroni adjustments to control the overall Type I error however, it is a difficult decision because such smaller overall alpha levels often result in lowering statistical power. Alternatively, Kano [115] proposed that the Lagrangian multiplier test be used (this test is also quite commonly provided in some commercially available modeling programs; see, e.g., EQS [85]) to investigate such cases.

Despite the fact that the issues of model identification have been well documented, these do not appear to be well known or commonly considered by applied researchers using modeling methodologies. We strongly believe that researchers must become cognizant of the potential consequences of ignoring these issues and at least understand some of the basics involved in accordance to the model being tested.

6 Myths About the Coefficient α

Coefficient alpha is frequently used in empirical research as an index that informs about measurement instrument reliability. Most measurement instruments (e.g., inventories, questionnaires, self-reports or tests), are typically developed to provide an overall assessment (an overall score) of an underlying latent dimension by accumulating information about various aspects of the latent dimension across their components (e.g., questions or items). Coefficient alpha is applicable when the components of a given measurement instrument are dichotomous or polytomous and capitalizes on the interrelationships among the instrument components (specifically their covariance) to provide a reliability estimation index. The estimate is readily available in most statistical packages and can be easily obtained with them for use in any empirical research setting. Unfortunately, and despite its availability and widespread use, a number of troubling myths about coefficient alpha appear prevalent among researchers [123]. For example, many researchers incorrectly use it as an index of dimensionality, often declaring that a set of items can be judged to be unidimensional when $\alpha > 0.70$. Coefficient alpha assumes unidimensionality, but it is not a test of it. As we highlight below, however, such interpretations and reliance on alpha can be quite problematic and misleading (for complete details, see [54, 123–134]). We address here two specific myths held about the coefficient and clarify some inaccuracies and inconsistencies commonly encountered in the literature (for more details see [123]):

1. Alpha is only an index of internal consistency. In other words, it is an index of the degree to which a set of instrument components are interrelated (in terms of inter-item covariance). The higher the magnitude of this covariance, the higher the value of coefficient alpha. Such evidence, however, does not imply unidimensionality of the set of components of a considered instrument (see also [135], for an insightful discussion and counter-examples, as well as [124]). The coefficient

alpha merely assumes unidimensionality, it does not test for it. If a researcher is interested in assessing unidimensionality, alpha cannot provide such information. In order to examine the unidimensionality hypothesis itself, one should prefer the use of an exploratory or a confirmatory factor analysis. With a confirmatory factor analysis one can essentially statistically test this hypothesis and evaluate the extent to which it may be viewed as supported for a measurement instrument in a given data set.

2. Alpha is not in general a lower bound of reliability. Alpha is a lower bound of reliability only under certain specific measurement circumstances. For example, Raykov and Marcoulides [123] stressed that this property only holds with uncorrelated errors among a set of components (for further details see also [54, 134]). With correlated errors, the underestimation feature of alpha does not generally hold. If they are correlated, alpha may or may not be a lower bound of composite reliability, regardless of the number of underlying dimensionality of the measuring instrument under consideration.

7 The Use of Correlation and Covariance Matrices

Although much literature has addressed the issue of the potential differences that can occur when analyzing correlation matrices as covariance matrices (and vice versa), it does not seem to be well understood that applying a covariance structure to a correlation matrix can produce some combination of incorrect standard errors, parameter estimates, or test statistics, and may even alter the studied model and results (see also [24] and references therein). Researchers applying PLS and related methods often appear to arbitrarily analyze the matrix of choice (or perhaps even convenience) without realizing that it is possible and probably most likely that incorrect conclusions may be drawn because of this choice. Such choices are especially important when flawed attempts are made to compare the various methods and their performance under supposedly varying distributional characteristics (see, e.g., [22] in which PLS was compared to other commonly used modeling techniques—see detailed discussion in next section).

The message is quite simple. The model must be scale invariant [24]. In other words, a scale invariant covariance matrix is one that can be transformed into the associated correlation matrix by rescaling the model parameters by functions of standards deviations. Simply standardizing the covariance matrix may or may not affect the analysis, but it really depends on the model being considered. For example, using both the correlation and covariance matrices computed for a set of eight variables ($n = 72$) collected in a clinical setting originally reported in Jolliffe ([136]; see p. 40 for the observed data matrix), Marcoulides et al. [74] showed that the weighted composite $w_1 = 0.2, 0.4, 0.4, 0.4, -0.4, -0.4, -0.2, -0.2$ (explaining 35% of the total variation in the variables) would be obtained when the correlation matrix is used, compared to the obtained weighted composite $w_2 = 0.0, 1.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0$ (explaining 99% of the total variation) when the

covariance matrix is used. The main reason for this disparity is the considerable difference in the standard deviations caused by the differences in scale for each of the eight variables (which are 0.371, 41.253, 1.935, 0.077, 0.071, 4.037, 2.732, and 0.297, respectively). Cudeck [24] also presented three simple factor analytic models for the observed matrix $\Sigma = \Sigma(\gamma) = \Lambda\Phi\Lambda' + \Theta$, (where Λ is the factor loading matrix, Φ the factor correlation matrix, Θ is the error matrix, and γ is the model defined parameter vector), and showed that, unless the model under examination is indeed appropriate for scale changes, any rescaling that occurs modifies the model completely (in the example used two of the models were scale invariant and one was not). Thus, if a factor analysis model is invariant, it is always possible to obtain estimates of the parameters. The original model structure will not be modified, only the elements of the parameter vector γ .

Although an algebraic derivation is the best way to determine whether a model is scale invariant, unfortunately this is often quite cumbersome and particularly difficult with complex models. An easy to follow practical approach is to simply fit the proposed model structure twice: first to the observed covariance matrix and then again to the observed matrix of correlations. The model structure is likely invariant if at the minima the discrepancy function of each obtained model is equal. We emphasize that this equality is not in and of itself sufficient evidence [24]. Nevertheless, the structure is categorically not invariant if the two are not equal. Assuming that either a correlation or a covariance matrix may be interchangeably examined can prove to be tricky.

8 Comparisons Among Modeling Methods

Numerous researchers have attempted studies comparing the efficacy of PLS with that of other modeling approaches, often without ever addressing the issue of the legitimacy of these comparisons. For example, if the comparison is between multiple regression, PLS, or other related modeling technique, then it is trivial. This is because an analysis of the same data and model based on a single regression equation using multiple regression, PLS or other modeling approach will always result in identical estimates. Obtaining such identical estimates is due to the well known fact that a single regression equation is a just-identified model and fits the data in the exact same way irrespective of the minimized fit function. For instance, Goodhue et al. [22, 137] attempted such trivial comparisons and then reported on supposed differences in the methods without ever realizing that their comparisons were wrong (for complete details see [3]). In contrast, Hwang et al. [138] in their comparison study carefully stipulated the precise conditions of their analyses and fully acknowledge the limitations of their comparisons. They openly acknowledged the differences in the setup of the approaches in terms of model specification and parameter estimation ahead of any analyses conducted. They subsequently indicated that

...this leads to the specification of different sets of model parameters for latent variables (i.e., factor means and/or variances in covariance structure analysis versus compo-

nent weights in partial least squares) ... The algebraic formulations underlying the three approaches seem to result in substantial difference in the procedures of parameter estimation.

They go on to point out again that the

... approaches estimate different sets of model parameters ... Thus, in this study we evaluate and report the recovery of the estimates of a common set of parameters ... (p. 703).

They conclude by acknowledging their inability to provide correctly parameterized comparisons among the approaches and indicate that

... we generated simulated data on the basis of covariance structure analysis ... we adopted the procedure because it was rather difficult to arrive at an impartial way of generating synthetic data for all three approaches ... (p. 710).

An appropriate approach for correctly parameterized comparisons between PLS and other methods was recently proposed by Treiblmaier et al. [23]. This approach begins by distinguishing between models with observed variables ($\mathbf{x}$), composite variables ($\mathbf{F}$) and latent variables (F), and unambiguously implements an $\mathbf{F}$ that closely approximates an F for comparison purposes. Doing so, however, requires a two-step approach that splits the determinate part of the composite into two or more composites and then models them as latent variables. This method can be readily contrasted with other inappropriate comparisons that simply create substitute estimates of latent variables (as was done in [22]). As explained by Marcoulides et al. [3], specifying models in this manner does not eliminate the fact that they are differentially parameterized models (in other words, an $\mathbf{x} \rightarrow \mathbf{F}$ path is not the same as a $\mathbf{x} \rightarrow F$ path). Although substitution of estimates for $\mathbf{F}$ is routinely done when conducting such comparisons, there are well-known and clear consequences (see complete details provided in [23]), not the least of which that "... not all parameters will be estimated consistently" ([139], p. 37).

The above reasons summarize why Marcoulides et al. [3, 74] emphatically warned researchers that

... the comparison of PLS to other methods cannot and should not be applied indiscriminately

and referred to any inappropriate evaluations between methods as "comparing apples with oranges."

The central issue dictating the legitimacy of such comparisons revolves around the notion of differentially parameterized models. Ignoring the legitimacy of this concern can lead to incorrect conclusions or may lead to overstating the importance of observed results [74]. Goodhue, Lewis, and Thompson to some extent acknowledged this fact when they stated that

We owe you all an apology! We were so certain that we were right in the equivalence of the methods, but now we see that the issue is more complicated than we thought (You are probably not surprised, at least by this last phrase!). We were focusing on how the techniques were used in practice, and didn't see that how they are used in practice is, in fact, not equivalent (May 1, 2008, personal communication).

But, unfortunately, Goodhue et al. [22, 140] somehow ultimately disregarded this fact and attempted to report on results from incorrect comparisons. It is essential that researchers ensure that any observed differences encountered between methods are not merely a function of differentially parameterized models being analyzed. Ignoring this matter can and has repeatedly lead to the unfortunate incidence of overstating the importance of the outcomes observed as in the case of Goodhue et al. [22, 140].

References

- [1] G.A. Marcoulides, Structural equation modeling for scientific research. *Journal of Business and Society*, **2**, pp. 130–138, 1989.
- [2] F. Galton, *Natural inheritance*, New York: MacMillan, 1889.
- [3] G.A. Marcoulides, W.W. Chin, and C. Saunders, When imprecise statements become problematic: A response to Goodhue, Lewis, and Thompson. *MIS Quarterly*, **36**, pp. 717–728, 2012.
- [4] J.L. Arbuckle and W. Wothke, *Amos 4.0 User's guide*, Chicago, IL: SPSS, 1999.
- [5] P.M. Bentler, *EQS structural equations program manual*. Encino, CA: Multivariate Software, 2004.
- [6] K. G. Jöreskog, and D. Sörbom, *LISREL 8 user's reference guide*. Chicago, IL: Scientific Software International, Inc, 2005.
- [7] J.B. Lohmoller, *Latent variable path modeling with partial least squares*, Heidelberg: Physica-Verlag, 1989.
- [8] L.K. Muthén, and B.O. Muthén, *Mplus user's guide* (5th ed.). Los Angeles, CA: Muthén and Muthén, 2011.
- [9] M.C. Neale, S.M. Boker, G. Xie, and H.H. Maes, *Mx: Statistical modeling* (5th ed.). Richmond, VA: Virginia Commonwealth University, 1999.
- [10] W.W. Chin, The partial least squares approach for structural equation modeling. In G.A. Marcoulides (Ed.), *Modern methods for business research* (pp. 295–336), Mahwah, NJ: Lawrence Erlbaum, 1998.
- [11] W.W. Chin, *PLS Graph User's Guide—Version 3.0*, Soft Modeling Inc, 1993–2003.
- [12] Y. Li, *PLS-GUI: Graphic user interface for partial least squares (PLS-PC 1.8)—Version 2.0.1 beta*, University of South Carolina, Columbia, SC, 2005.
- [13] N. Sellin, *PLSPATH—Version 3.01. Application manual*. Universität Hamburg, Hamburg, 1989.
- [14] M.W. Browne, and G. Mels, Path analysis (RAMONA). In SPSS Inc. *SYSTAT 10 statistics II*, Chapter 7, (pp. 233–291), Chicago: Author, 2000.
- [15] SAS Institute, *SAS PROC CALIS User's guide*, Cary, NC: Author, 1989.
- [16] Statistica, *User's guide*, Tulsa, OK: Statistica Inc, 1998.
- [17] C.M. Ringle, S. Wende and A. Will, *SmartPLS—Version 2.0*, Universität Hamburg, Hamburg, 2005.
- [18] J.-R. Fu, *VisualPLS, Partial Least Square (PLS) Regression: An enhanced GUI for Lvpls (PLS 1.8 PC) Version 1.04*, National Kaohsiung University of Applied Sciences, Taiwan, ROC, 2006.
- [19] J.-R. Fu, *VisualPLS: Partial least square (PLS) regression, an enhanced GUI for LVPLS (PLS 1.8 PC) Version 1.04*. <http://www2.kuas.edu.tw/prof/fred/vpls/index.html>, 2006.
- [20] Addinsoft XLSTAT 2012, Data analysis and statistics software for Microsoft Excel, <http://www.xlstat.com>, Paris, France, 2012.
- [21] R.E. Schumacker and G.A. Marcoulides, *Interaction and nonlinear effects in structural equation modeling*, Mahwah, NJ: Lawrence Erlbaum, 1998.

- [22] D. Goodhue, W. Lewis, and R. Thompson, Comparing PLS to regression and LISREL: A response to Marcoulides, Chin, and Saunders. *MIS Quarterly*, **36**(3), pp. 703–716, 2012.
- [23] H. Treiblmaier, P.M. Bentler, and P. Mair, Formative constructs implemented via common factors. *Structural Equation Modeling*, **18**, pp. 1–17, 2010.
- [24] R. Cudeck, Analysis of correlation matrices using covariance structure models. *Psychological Bulletin*, **105**, pp. 317–327, 1989.
- [25] J.H. Steiger, Driving fast in reverse: The relationship between software development, theory, and education in structural equation modeling. *Journal of the American Statistical Association*, **96**, pp. 331–338, 2001.
- [26] K.G. Jöreskog, Testing structural equation models. In K.A. Bollen and J.S. Long (Eds.), *Testing structural equation models* (pp. 294–316). Newbury Park, CA: Sage, 1993.
- [27] H. Wold, Soft modeling: Intermediate between traditional model building and data analysis. *Mathematical statistics*, **6**, pp. 333–346, 1982.
- [28] K. Hayashi, and G.A. Marcoulides, Examining identification issues in factor analysis. *Structural Equation Modeling*, **13**, pp. 631–645, 2006.
- [29] C. Glymour, R. Scheines, R. Spirtes, and K. Kelly, *Discovering causal structure: Artificial intelligence, philosophy of science, and statistical modeling*, Orlando, FL: Academic Press, 1987.
- [30] G.A. Marcoulides, Z. Drezner, and R.E. Schumacker, Model specification searches in structural equation modeling using Tabu search. *Structural Equation Modeling*, **5**, pp. 365–376, 1998.
- [31] S. Salhi, Heuristic search methods. In G.A. Marcoulides (Ed.), *Modern methods for business research* (pp. 147–175). Mahwah, NJ: Lawrence Erlbaum, 1998.
- [32] G.A. Marcoulides, and Z. Drezner, Specification searches in structural equation modeling with a genetic algorithm. In G.A. Marcoulides and R.E. Schumacker (Eds.), *New developments and techniques in structural equation modeling* (pp. 247–268). Mahwah, NJ: Lawrence Erlbaum, 2009.
- [33] W.L. Leite, I.C. Huang, and G.A. Marcoulides, Item selection for the development of short-form of scaling using an ant colony optimization algorithm. *Multivariate Behavioral Research*, **43**, pp. 411–431, 2008.
- [34] W.L. Leite, and G.A. Marcoulides, *Using the ant colony optimization algorithm for specification searches: A comparison of criteria*, Paper presented at the Annual Meeting of the American Education Research Association, San Diego: CA, 2009, April.
- [35] G.A. Marcoulides, and W.L. Leite, Exploratory data mining algorithms for conducting searches in structural equation modeling: A comparison of some fit criteria. In J.J. McArdle, and G. Ritschard (Eds.), *Contemporary Issues in Exploratory Data Mining in the Behavioral Sciences*, London: Taylor & Francis, 2013.
- [36] G.A. Marcoulides, and Z. Drezner, Model specification searches using ant colony optimization algorithms. *Structural Equation Modeling*, **10**, pp. 154–164, 2003.
- [37] G.A. Marcoulides, and Z. Drezner, A model selection approach for the identification of quantitative trait loci in experimental crosses: Discussion on the paper by Broman and Speed. *Journal of the Royal Statistical Society, Series B*, **64**, pp. 754, 2002.
- [38] G.A. Marcoulides, *Conducting specification searches in SEM using a ruin and recreate principle*, Paper presented at the Annual Meeting of the American Psychological Society, San Francisco, CA, 2009, May.
- [39] G.A. Marcoulides, and Z. Drezner, Using simulated annealing for model selection in multiple regression analysis. *Multiple Regression Viewpoints*, **25**, pp. 1–4, 1999.
- [40] G.A. Marcoulides, and Z. Drezner, Tabu search variable selection with resource constraints. *Communications in Statistics: Simulation & Computation*, **33**, pp. 355–362, 2004.
- [41] Z. Drezner, and G.A. Marcoulides, A distance-based selection of parents in genetic algorithms. In M. Resenda and J.P. Sousa (eds.), *Metaheuristics: Computer decision-making* (pp. 257–278). Boston, MA: Kluwer Academic Publishers, 2003.
- [42] G.A. Marcoulides, *Using heuristic algorithms for specification searches and optimization*, Paper presented at the Albert and Elaine Brochard Foundation International Colloquium, Missillac, France, 2010, July.

- [43] G.A. Marcoulides, and M. Ing., Automated structural equation modeling strategies. In R. Hoyle (Ed.), *Handbook of structural equation modeling*, New York: Guilford Press, 2012.
- [44] R. MacCallum, Specification searches in covariance structure modeling. *Psychological Bulletin*, **100**, pp. 107–120, 1986.
- [45] S.J. Breckler, Applications of covariance structure modeling in Psychology: Cause for concern? *Psychological Bulletin*, **107**, pp. 260–273, 1990.
- [46] S.L. Hershberger, The specification of equivalent models before the collection of data. In A. von Eye and C. Clogg (Eds.), *The analysis of latent variables in developmental research* (pp. 68–108). Beverly Hills, CA: Sage, 1994.
- [47] S. Lee., and S.L. Hershberger, A simple rule for generating equivalent models in covariance structure modeling. *Multivariate Behavioral Research*, **25**, pp. 313–334, 1990.
- [48] S.L. Hershberger, and G.A. Marcoulides, The problem of equivalent models. In G.R. Hancock and R.O. Mueller (Eds.), *Structural equation modeling: A second course* (2nd Ed.). G. R. Greenwich, CT: Information Age Publishing, 2012.
- [49] T.C.W. Luijben, Equivalent models in covariance structure analysis. *Psychometrika*, **56**, pp. 653–665, 1991.
- [50] R.C. MacCallum, D.T. Wegener, B.N. Uchino, and L.R. Fabrigar, The problem of equivalent models in applications of covariance structure analysis. *Psychological Bulletin*, **114**, pp. 185–199, 1993.
- [51] R. Levy, and G.R. Hancock, A framework of statistical tests for comparing mean and covariance structure models. *Multivariate Behavioral Research*, **42**, pp. 33–66, 2007.
- [52] K.A. Marcus, Statistical equivalence, semantic equivalence, eliminative induction and the Raykov-Marcoulides proof of infinite equivalence. *Structural Equation Modeling*, **9**, pp. 503–522, 2002.
- [53] R.P. McDonald, What can we learn from the path equations?: Identifiability, constraints, equivalence. *Psychometrika*, **67**, pp. 225–249, 2002.
- [54] T. Raykov, Equivalent structural equation models and group equality constraints. *Multivariate Behavioral Research*, **32**, pp. 95–104, 1997.
- [55] T. Raykov, and G.A. Marcoulides, Can there be infinitely many models equivalent to a given covariance structure model? *Structural Equation Modeling*, **8**, pp. 142–149, 2001.
- [56] T. Raykov, and G.A. Marcoulides, Equivalent structural equation models: A challenge and responsibility. *Structural Equation Modeling*, **14**, pp. 527–532, 2007.
- [57] T. Raykov, and S. Penev, On structural equation model equivalence. *Multivariate Behavioral Research*, **34**, pp. 199–244, 1999.
- [58] L. J. Williams, H. Bozdogan, and L. Aiman-Smith, Inference problems with equivalent models. In G.A. Marcoulides and R.E. Schumacker (Eds.), *Advanced structural equation modeling: Issues and techniques* (pp. 279–314). Mahwah, NJ: Erlbaum, 1996.
- [59] C. Hsiao, Identification. In Z. Griliches and M.D. Intriligator (Eds.), *Handbook of econometrics, Vol. 1* (pp. 224–283). Amsterdam: Elsevier Science, 1983.
- [60] I. Stelzl, Changing a causal hypothesis without changing the fit: Some rules for generating equivalent path models. *Multivariate Behavioral Research*, **21**, pp. 309–331, 1986.
- [61] J. Pearl, *Causality: Models, reasoning, and inference* (2nd ed.). New York: Cambridge University Press, 2009.
- [62] B. Shipley, *Cause and correlation in biology*, New York: Cambridge University Press, 2000.
- [63] P.A. Bekker, A. Merckens, and T.J. Wansbeek, *Identification, equivalent models, and computer algebra*, Boston: Academic Press, 1994.
- [64] S.B. Green, M.S. Thompson, and J. Poirier, Exploratory analyses to improve model fit: Errors due to misspecification and a strategy to reduce their occurrence. *Structural Equation Modeling*, **6**, pp. 113–126, 1999.
- [65] T. Raykov, and S. Penev, The problem of equivalent structural equation models: An individual residual perspective. In G.A. Marcoulides and R.E. Schumacker (Eds.), *New developments and techniques in structural equation modeling*, (pp. 297–321). Mahwah, NJ: Lawrence Erlbaum, 2001.
- [66] G.H. Scherr, Irreproducible science: Editor's introduction. *The best of the Journal of Irreproducible Results*, New York: Workman Publishing, 1983.

- [67] H. Apel, and H. Wold, *Simulation experiments on a case value basis with different sample lengths, different sample sizes, and different estimation models, including second dimension of latent variables*. Unpublished manuscript, Department of Statistics, University of Uppsala, Sweden, 1978.
- [68] H. Apel, and H. Wold, Soft modeling with latent variables in two or more dimensions: PLS estimation and testing for predictive relevance. In K.G. Jöreskog and H. Wold (Eds.), *Systems under indirect observation, Part II*, (pp. 209–248). Amsterdam: North Holland, 1982.
- [69] B.E. Areskoug, The first canonical correlation: Theoretical PLS analysis and simulation experiments. In K.G. Jöreskog and H. Wold (Eds.), *Systems under indirect observation, Part II* (pp. 95–118). Amsterdam: North Holland, 1982.
- [70] W.W. Chin, and P.R. Newsted, 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, 1999.
- [71] B.S. Hui, *The partial least squares approach to path models of indirectly observed variables with multiple indicators*. Unpublished doctoral dissertation, University of Pennsylvania, Philadelphia, PA, 1978.
- [72] B.S. Hui, and H. Wold, Consistency and consistency at large in partial least squares estimates. In K. G. Jöreskog and H. Wold (Eds.), *Systems under indirect observation, Part II* (pp. 119–130). Amsterdam: North Holland, 1982.
- [73] G.A. Marcoulides, and C. Saunders, PLS: A Silver Bullet? *MIS Quarterly*, **30**, pp. iv–viii, 2006.
- [74] G.A. Marcoulides, W.W. Chin, and C. Saunders, A critical look at partial least squares modeling. *MIS Quarterly*, **33**, pp. 171–175, 2009.
- [75] R. Noonan, and H. Wold, PLS path modeling with indirectly observed variables: A comparison of alternative estimates for the latent variable. In K. G. Jöreskog and H. Wold (Eds.), *Systems under indirect observation, Part II* (pp. 75–94). Amsterdam: North Holland, 1982.
- [76] W.W. Chin, and J. Dibbern, A permutation based procedure for multi-group PLS analysis: Results of tests of differences on simulated data and a cross cultural analysis of the sourcing of information system services between Germany and the USA, In V.E. Vinzi, W.W. Chin, J. Henseler and H. Wang (Eds.), *Handbook of partial least squares concepts, methods and applications* (pp. 171–193), New York, NY: Springer Verlag, 2010.
- [77] W.W. Chin, A permutation procedure for multi-group comparison of PLS models. Invited presentation, In M. Valares, M. Tenenhaus, P. Coelho, V.E. Vinzi, and A. Morineau (Eds.), *PLS and related methods, proceedings of the PLS-03 International Symposium: "Focus on Customers"*, Lisbon, September 15th to 17th, pp. 33–43, 2003.
- [78] K.G. Jöreskog, and H. Wold, *Systems under indirect observation, Part I & II*, North Holland: Amsterdam, 1982.
- [79] R.F. Falk, and N.B. Miller, *A primer of soft modeling*. Akron, OH: The University of Akron Press, 1992.
- [80] R.P. McDonald, Path analysis with composite variables. *Multivariate Behavioral Research*, **31**, pp. 239–270, 1996.
- [81] I.R.R. Lu, *Latent variable modeling in business research: A comparison of regression based on IRT and CTT scores with structural equation models*, Doctoral dissertation, Carleton University, Canada, 2004.
- [82] I.R.R. Lu, D.R. Thomas, and B.D. Zumbo, Embedding IRT in structural equation models: A comparison with regression based on IRT scores. *Structural Equation Modeling*, **12**, pp. 263–277, 2005.
- [83] T. Dijkstra, Some comments on maximum likelihood and partial least squares methods. *Journal of Econometrics*, **22**, pp. 67–90, 1983.
- [84] H. Schneeweiss, Consistency at large in models with latent variables. In K. Haagen, D.J. Bartholomew, and M. Deistler (Eds.), *Statistical modelling and latent variables*, Elsevier: Amsterdam, 1993.
- [85] P.M. Bentler, *EQS structural equation program manual*, Encino, CA: Multivariate Software, Inc., 1995.

- [86] L.-T. Hu, P.M. Bentler, and Y. Kano, Can test statistics in covariance structure analysis be trusted? *Psychological Bulletin*, **112**, pp. 351–362, 1992.
- [87] J. Cohen, *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum, 1988.
- [88] S.B. Green, How many subjects does it take to do a regression analysis? *Multivariate Behavioral Research*, **26**, pp. 499–510, 1991.
- [89] A. Boomsma, The robustness of LISREL against small sample sizes in factor analysis models. In K.G. Jöreskog and H. Wold (Eds.), *Systems under indirect observation, Part I* (pp. 149–174). Amsterdam: North Holland, 1982.
- [90] R. Cudeck, and S.J. Hensly, Model selection in covariance structure analysis and the “problem” of sample size: A clarification. *Psychological Bulletin*, **109**, pp. 512–519, 1991.
- [91] D.L. Jackson, Revisiting sample size and the number of parameter estimates: Some support for the $N : q$ Hypothesis. *Structural Equation Modeling*, **10**, pp. 128–141, 2003.
- [92] R.C. MacCallum, M.W. Browne, and H.M. Sugawara, Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods*, **1**, pp. 130–149, 1996.
- [93] L.K. Muthén, and B.O. Muthén, How to use a Monte Carlo study to decide on sample size and determine power. *Structural Equation Modeling*, **9**, pp. 599–620, 2002.
- [94] C. Ringle, M. Sarstedt, and D. Straub, A critical look at the use of PLS-SEM in MIS Quarterly. *MIS Quarterly*, **36**, pp. iii–xiv, 2012.
- [95] A. Bhattacharjee, and G. Premkumar, Understanding changes in belief and attitude toward information technology usage: A theoretical model and longitudinal test. *MIS Quarterly*, **28**, pp. 229–254, 2004.
- [96] G. Bassellier, and I. Benbasat, Business competence of information technology professionals: Conceptual development and influence on IT-business partnerships. *MIS Quarterly*, **28**, pp. 673–694, 2004.
- [97] M. Subramani, How do suppliers benefit from information technology use in supply chain relationships? *MIS Quarterly*, **28**, pp. 45–73, 2004.
- [98] M.C. Denham, Prediction intervals in Partial Least Squares. *Journal of Chemometrics*, **11**, pp. 39–52, 1997.
- [99] W. L. Hays, *Statistics*, Fort Worth, TX: Harcourt Brace Jovanovic, 1994.
- [100] T. Raykov, and G.A. Marcoulides, *A first course in structural equation modeling* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum, 2006.
- [101] S. Serneels, P. Lemberge, and P.J. Van Espen, Calculation of PLS prediction intervals using efficient recursive relations for the Jacobian matrix. *Journal of Chemometrics*, **18**, pp. 76–80, 2004.
- [102] T. Raykov, and G.A. Marcoulides, Using the delta method for approximate interval estimation of parameter functions in SEM. *Structural Equation Modeling*, **11**, pp. 621–637, 2004.
- [103] G.A. Marcoulides, Evaluation of confirmatory factor analytic and structural equation models using goodness-of-fit indices. *Psychological Reports*, **67**, pp. 669–671, 1990.
- [104] P. Paxton, P.J. Curran, K.A. Bollen, J. Kirby, and F. Chen, Monte Carlo experiments: Design and implementation. *Structural Equation Modeling*, **8**, pp. 287–312, 2001.
- [105] A. Majchrak, C. Beath, R. Lim, and W.W. Chin, Managing client dialogues during information systems design to facilitate client learning. *MIS Quarterly*, **29**, pp. 653–672, 2005.
- [106] J. Dibbern, W.W. Chin, A. Heinzl, Systemic determinants of the information systems outsourcing decision: A comparative study of German and United States firms. *Journal of the Association for Information Systems*, **13**, pp. 466–497, 2012.
- [107] E. Walden, AIS World: Power analysis after the fact. Aisworld-bounces@listss.aisnet.org. [Monday, March 5, 2012 6:51pm], 2012.
- [108] T. Raykov, and G.A. Marcoulides, *Basic statistics: An introduction with R*. London, UK: Rowman & Littlefield Publishers, Inc., 2012.
- [109] A.A. Albert, The matrices of factor analysis. *Proceedings of the National Academy of Science*, **30**, pp. 90–95, USA, 1944.

- [110] A.A. Albert, The minimum rank of a correlation matrix. *Proceedings of the National Academy of Science*, **30**, pp. 144–146, USA, 1944.
- [111] T.C. Koopmans, and O. Reiersol, The identification of structural characteristics. *Annals of Mathematical Statistics*, **21**, pp. 165–181, 1950.
- [112] W. Ledermann, On the rank of reduced correlation matrices in multiple factor analysis. *Psychometrika*, **2**, pp. 85–93, 1937.
- [113] M. Sato, A study of identification problem and substitute use in principal component analysis in factor analysis. *Hiroshima Mathematical Journal*, **22**, pp. 479–524, 1992.
- [114] M. Sato, On the identification problem of the factor analysis model: A review and an application for an estimation of air pollution source profiles and amounts. In *Factor analysis symposium at Osaka* (pp. 179–184). Osaka: University of Osaka, 2004.
- [115] Y. Kano, Exploratory factor analysis with a common factor with two indicators. *Behaviormetrika*, **24**, pp. 129–145, 1997.
- [116] K.A. Bollen, *Structural equations with latent variables*, New York, NY: Wiley, 1989.
- [117] T. Raykov, G.A. Marcoulides, and T. Patelis, Saturated versus just identified models: A note on their distinction. *Educational and Psychological Measurement*, **73**, pp. 162–168, 2013.
- [118] T.W. Anderson, and H. Rubin, Statistical inferences in factor analysis. In J. Neyman (Ed.), *Proceedings of the third Berkeley symposium on mathematical statistics and probability* (Vol. 5, pp. 111–150). Berkeley: University of California, 1956.
- [119] M. Ihara, and Y. Kano, A new estimator of the uniqueness in factor analysis. *Psychometrika*, **51**, pp. 563–566, 1986.
- [120] H. Yanai, K. Shigemasu, S. Maekawa, and M. Ichikawa, *Factor analysis: Its theory and methods*, Tokyo: Asakura-shoten (in Japanese), 1990.
- [121] J.S. Williams, A note on the uniqueness of minimum rank solutions in factor analysis. *Psychometrika*, **46**, pp. 109–110, 1981.
- [122] Y. Tumura, and M. Sato, On the identification in factor analysis. *TRU Mathematics*, **16**, pp. 121–131, 1980.
- [123] T. Raykov, and G.A. Marcoulides, *Introduction to psychometric theory*. New York, NY: Routledge, 2011.
- [124] R.P. McDonald, The dimensionality of tests and items. *British Journal of Mathematical and Statistical Psychology*, **34**, pp. 100–117, 1981.
- [125] R.P. McDonald, *Test theory. A unified treatment*, Mahwah, NJ: Lawrence Erlbaum, 1999.
- [126] P.M. Bentler, Alpha, dimension-free, and model-based internal consistency reliability. *Psychometrika*, **74**, pp. 137–144, 2009.
- [127] L. Crocker, and J. Algina, *Introduction to classical and modern test theory*, Fort Worth, TX: Harcourt College Publishers, 1986.
- [128] S.B. Green, and Y. Yang, Commentary on coefficient alpha: A cautionary tale. *Psychometrika*, **74**, pp. 121–136, 2009.
- [129] S.B. Green, and Y. Yang, Reliability of summed item scores using structural equation modeling: An alternative to coefficient alpha. *Psychometrika*, **74**, pp. 155–167, 2009.
- [130] W. Revelle, and R. Zinbarg, Coefficients alpha, beta, and the GLB: Comments on Sijsma. *Psychometrika*, **74**, pp. 145–154, 2009.
- [131] K. Sijsma, On the use, the misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika*, **74**, pp. 107–120, 2009.
- [132] K. Sijsma, Reliability beyond theory and into practice. *Psychometrika*, **74**, pp. 169–174, 2009.
- [133] T. Raykov, Cronbach's alpha and reliability of composite with interrelated nonhomogenous items. *Applied Psychological Measurement*, **22**, pp. 375–385, 1998.
- [134] T. Raykov, Bias of coefficient alpha for congeneric measures with correlated errors. *Applied Psychological Measurement*, **25**, pp. 69–76, 2001.
- [135] S.B. Green, R.W. Lissitz, and S.A. Mulaik, Limitations of coefficient alpha as an index of test unidimensionality. *Educational and Psychological Measurement*, **37**, pp. 827–838, 1977.
- [136] I.T. Jolliffe, *Principal component analysis* (2nd Ed.). New York: Springer, 2002.

- [137] D. Goodhue, W. Lewis, and R. Thompson, PLS, Small Sample Size, and Statistical Power in MIS Research. *HICSS '06 Proceedings of the 39th Annual Hawaii Conference on System Sciences*, pp. 202b, 2006.
- [138] H. Hwang, N.K. Malhotra, Y. Kim, M.A. Tomiuk, and S. Hong, A comparative study of parameter recovery of the three approaches to structural equation modeling. *Journal of Marketing Research*, **67**, pp. 699–712, 2010.
- [139] T. Dijkstra, Latent variables and indices: Herman Wold's basic design and partial least squares. In V.E. Vinzi, W.W. Chin, J. Henseler, and H. Wang (Eds.), *Handbook of partial least squares: Concepts, methods, and applications, computational statistics* (pp. 23–46). New York: Springer Verlag, 2010.
- [140] D. Goodhue, W. Lewis, and R. Thompson, Does PLS have advantages for small sample size or non-normal data. *MIS Quarterly*, **36**, pp. 981–1001, 2012.

New Perspectives in Partial Least Squares and Related
Methods

Abdi, H.; Chin, W.W.; Esposito Vinzi, V.; Russolillo, G.;
Trinchera, L. (Eds.)

2013, XXIII, 344 p. 96 illus., 49 illus. in color., Hardcover

ISBN: 978-1-4614-8282-6