

Partial Least Squares Path Modeling: Time for Some Serious Second Thoughts¹

Mikko Rönkkö
Aalto University School of Science
PO Box 15500
FI-00076 Aalto, Finland
phone: + 358 50 387 8155
email: mikko.ronkko@aalto.fi

Cameron N. McIntosh
Public Safety Canada

John Antonakis
Faculty of Business and Economics
University of Lausanne

Jeffrey R. Edwards
Kenan-Flagler Business School
University of North Carolina at Chapel Hill

Accepted for publication in Journal of Operations Management
DOI: 10.1016/j.jom.2016.05.002

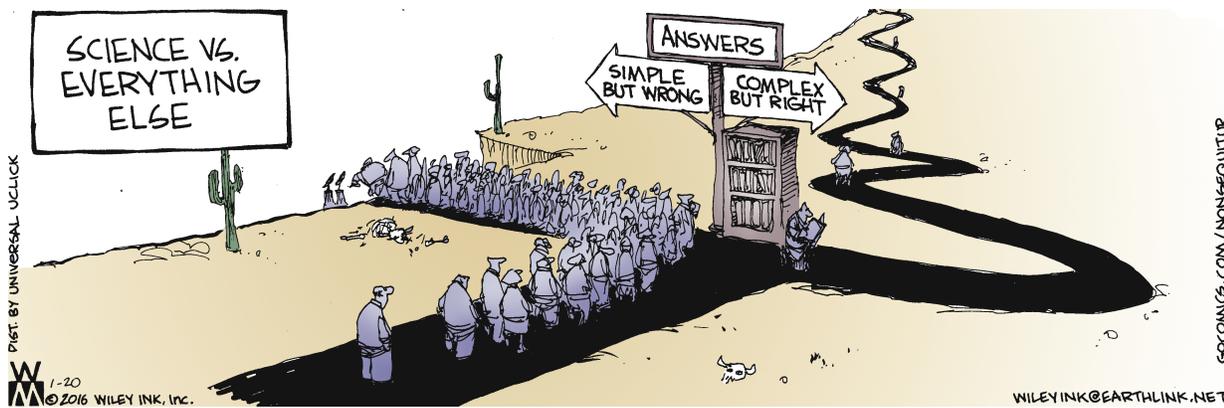
¹ We thank Mikko Ketokivi for giving us the data used in this article. We thank John Cadogan, Joerg Evermann, Dale Goodhue, Philipp Holtkamp, Mikko Ketokivi, Nick Lee, and Ron Thompson for their comments on earlier versions of the article.

The authors have no financial interests in any of the software packages mentioned. The first authors is developer of the free and open source *matrixpls* R package and the *pls* module for Stata.

Abstract

Partial least squares (PLS) path modeling is increasingly being promoted as a technique of choice for various analysis scenarios, despite the serious shortcomings of the method. The current lack of methodological justification for PLS prompted the editors of this journal to declare that research using this technique is likely to be desk-rejected (Guide & Ketokivi, 2015). To provide clarification on the inappropriateness of PLS for applied research, we provide a non-technical review and empirical demonstration of its inherent, intractable problems. We show that although the PLS technique is promoted as a structural equation modeling (SEM) technique, it is simply regression with scale scores and thus has very limited capabilities to handle the wide array of problems for which applied researchers use SEM. To that end, we explain why the use of PLS weights and many rules of thumb that are commonly employed with PLS are unjustifiable, followed by addressing why the touted advantages of the method are simply untenable.

Keywords: Partial least squares, structural equation modeling, formative measurement, composite variables, capitalization on chance, significance testing, model fit, causality, statistical and methodological myths and urban legends.



NON SEQUITUR © 2016 Wiley Ink, Inc.. Dist. By UNIVERSAL UCLICK. Reprinted with permission. All rights reserved.

1. Introduction

Partial least squares (PLS) has become one of the techniques of choice for theory testing in some academic disciplines, particularly marketing and information systems, and its uptake seems to be on the rise in operations management (OM) as well (Peng & Lai, 2012; Rönkkö, 2014b). The PLS technique is typically presented as an alternative to structural equation modeling (SEM) estimators (e.g., maximum likelihood), over which it is presumed to offer several advantages (e.g., an enhanced ability to deal with small sample sizes and non-normal data).

Recent scrutiny suggests, however, that many of the purported advantages of PLS are not supported by statistical theory or empirical evidence, and that PLS actually has a number of disadvantages that are not widely understood (Goodhue, Lewis, & Thompson, 2015; McIntosh, Edwards, & Antonakis, 2014; Rönkkö, 2014b; Rönkkö & Evermann, 2013; Rönkkö, McIntosh, & Antonakis, 2015). As recently concluded by Henseler (2014) “like a hammer is a suboptimal tool to fix screws, PLS is a suboptimal tool to estimate common factor models”, which are the kind of models OM researchers use (Guide & Ketokivi, 2015, p. vii). Unfortunately, whereas a person attempting to hammer a screw will quickly realize that the tool is ill-suited for that

purpose, the shortcomings of PLS are much more insidious because they are not immediately apparent in the results of the statistical analysis. Although PLS promises simple solutions to complex problems and often produces plausible statistics that are seemingly supportive of research hypotheses, both the technical and applied literature on the technique seems to confound two distinct notions: (1) something can be done; and (2) doing so is methodologically valid (Westland, 2015, Chapter 3). As stated in a recent editorial by Guide and Ketokivi (2015): “Claiming that PLS fixes problems or overcomes shortcomings associated with other estimators is an indirect admission that one does not understand PLS” (p. vii). However, the editorial provides little material aimed at improving the understanding of PLS and its associated limitations.

Although there is no shortage of guidelines on how the PLS technique should be used, many of these are based on conventions, unproven assertions, and hearsay, rather than rigorous methodological support. Although OM researchers have followed these guidelines (Peng & Lai, 2012), such works do not help readers gain a solid and balanced understanding of the technique and its shortcomings. This state of affairs makes it difficult to justify the use of PLS, beyond arguing that someone has said that using the method would be a good idea in a particular research setting (Guide & Ketokivi, 2015, p. vii) Therefore, in order to mitigate common misunderstandings, we clarify issues concerning the usefulness of PLS in a non-technical manner for applied researchers. In light of these issues, it becomes apparent that the findings of studies employing PLS are ambiguous at best and at worst simply wrong, leading to the conclusion that PLS should be discontinued until the methodological problems explained in this article have been fully addressed.

2. What is PLS and What Does It Do?

A PLS analysis consists of two stages. First, indicators of latent variables are combined as weighted sums (composites); second, the composites are used in separate regression analyses, applying null hypothesis significance testing by comparing the ratio of a regression coefficient and its bootstrapped standard error against Student's t distribution. In a typical application, the composites are intended to measure theoretical constructs that are measured with multiple indicators. In this type of analysis, the purpose of combining the indicators into composites is to produce aggregate measures that can be expected to be more reliable than any of their components, and can therefore be used as reasonable proxies for the constructs. Thus the only difference between PLS and more traditional regression analyses using summed scales, factor scores, or principal components, is how the indicators are weighted to create the composites. Moreover, instead of applying traditional factor analysis techniques, the quality of the measurement model is evaluated by inspecting the correlations between indicators and composites that they form, summarized as the composite reliability (CR) and average variance extracted (AVE) indices.

Although PLS is often marketed as a SEM method, a better way to understand what the technique actually does is to simply consider it as one of many indicator weighting systems. The broader methodological literature provides several different ways to construct composite variables. The simplest possible strategy is taking the unweighted sum of the scale items, with a refined version of this approach being the application of unit weights to standardized indicators (Cohen, 1990). The two most common empirical weighting systems are principal components, which retain maximal information from the original data, and factor scores that assume an underlying factor model (Widaman, 2007), with various different calculation techniques

producing scores with different qualities (Grice, 2001b). Commonly used prediction weights include regression, correlation, and equal weights (Dana & Dawes, 2004). Although not linear composites, different models based on item response theory produce scale scores that take into account both respondent ability and item difficulty (Reise & Revicki, 2014). Outside the context of research, many useful indices are composites, such as stock market indices that can weight individual stocks based on their price or market capitalization. Given the large number of available approaches for constructing composites variables, two key questions are: (1) Does PLS offer advantages over more well-established procedures?; and (2) What is the purpose of the PLS weights used to form the composites? We address these questions next.

2.1. On the “Optimality” of PLS Weights

Most introductory texts on PLS gloss over the purposes of the weights, arguing that PLS is SEM and therefore it must provide an advantage over regression with composites (e.g., Gefen, Rigdon, & Straub, 2011); however, such works often do not explicitly point out that PLS itself is also simply regression with composites. Other authors suggest the weights are optimal (e.g., Henseler & Sarstedt, 2013, p. 566), but do not explain why and for which specific purpose. As noted by Krämer (2006): “In the literature on PLS, there is often a huge gap between the abstract model [...] and what is actually computed by the PLS path algorithms. Normally, the PLS algorithms are presented directly in connection with the PLS framework, insinuating that the algorithms produce optimal solutions of an obvious estimation problem attached to PLS. This estimation problem is however never defined” (p. 22).

The purpose of PLS weights remains ambiguous (Rönkkö et al., 2015, p. 77), as various rather *different* explanations abound (see Table 1 for some examples). However, none of these works (or their cited literature) provide mathematical proofs or simulation evidence to support

their arguments. Perhaps the most common argument is that the indicator weights maximize the R^2 values of the regressions between the composites in the model (e.g. Hair, Hult, Ringle, & Sarstedt, 2014, p. 16). However, this claim is problematic, for two main reasons: (a) why maximizing the R^2 values is a good optimization criterion is unclear; and (b) PLS has not been shown to be an optimal algorithm for maximizing R^2 . In contrast, Rönkkö (2016a, sec. 2.3) demonstrates a scenario where optimizing indicator weights directly with respect to R^2 produces a 180% larger R^2 value than PLS, thus demonstrating that if the purpose of the analysis is to maximize R^2 , the PLS algorithm is not an effective algorithm for this task.

----- Insert Table 1 about here -----

Another common claim is that PLS weights reduce the impact of measurement error (e.g., Chin, Marcolin, & Newsted, 2003, p. 194; Gefen et al., 2011, p. v). The traditional approach to adjusting for measurement error is to combine multiple “noisy” indicators into a composite measure (i.e., a sum), and then use the composite in the statistical analyses. This procedure typically assumes that the measurement errors are independent in the population, in which case combining multiple measures reduces the overall effect of the errors (cf, Rigdon, 2012). Whether the indicators should be weighted when forming the composite received considerable attention in the early literature on factor analysis, and indeed, the problem of how to generate maximally reliable composites was already solved in the 1930’s (Thomson, 1938) by the introduction of the regression method for calculating factor scores, a technique which is the default means of calculating factor scores in most general purpose statistical packages. Nevertheless, a central conclusion of the factor score literature is that, in general, the advantages of factor scores or other forms of indicator weighting over unit-weighted summed scales are typically vanishingly small, and that calculating weights based on sample data can even lead to substantially worse

outcomes in terms of reliability; hence the recommendation to use unit weights instead as a general-purpose solution (e.g., Bobko, Roth, & Buster, 2007; Cohen, 1990; Cohen, Cohen, West, & Aiken, 2003, pp. 97–98; Grice, 2001a; McDonald, 1996).

We demonstrate this general result in Figure 1, which shows four sets of estimates calculated from data simulated from a known population model². The regression estimates based on summed scales are negatively biased, which is a direct consequence of the presence of measurement errors in the composites (Bollen, 1989, Chapter 5). The same bias is visible in the PLS estimates as well, but there is also a clear bias away from zero that we will address later in the article. The ideally-weighted composites, calculated by regressing the latent variable scores, which were used to generate the data, on the indicators, provide a theoretical upper limit on reliability against which other composites can be compared. The differences between summed scales and ideal weights are small, demonstrating that indicator weighting cannot meaningfully reduce the effect of measurement error in the composites. In contrast to biased composite-based estimates, ML SEM estimates are unbiased, which is the expected result because the technique does not create composite proxies for the factors, but rather explicitly models different sources of variation in the indicators, including the errors.

---- Insert Figure 1 about here ----

This general finding is also showcased by a recent study demonstrating that, even in highly favorable scenarios, the value-added of PLS weights over unit weights is trivial (only about a 0.6% increase in reliability, on average), whereas in less favorable scenarios, the

² We simulated 1000 samples of 100 observations each from a population where two latent variables were each measured with three indicators loading at 0.8, 0.7, and 0.6, varying the correlation between the latent variables between 0-0.65. Next, using the data sampled from each population scenario, the pathway between the latent variables was estimated using each of the four techniques. The R code for the simulation is available in Appendix A, along with a replication using Stata in Appendix B.

associated loss is much more striking (an average decrease of around 16.8%) (Henseler et al., 2014, table 2). Therefore, the common claim that the PLS indicator weighting system would minimize the effect of measurement error (e.g., Chin et al., 2003; Fornell & Bookstein, 1982; Gefen et al., 2011), or more generally, that indicator weighting can meaningfully improve reliability (Rigdon, 2012), is simply untenable (McIntosh et al., 2014; Rönkkö & Evermann, 2013; Rönkkö et al., 2015).

2.2. What Do the PLS Weights Actually Accomplish?

Although it is not clear what purpose the PLS weights serve, or what their advantages are over other modeling approaches, this confusion does not automatically make the weights invalid. Nevertheless, understanding how the PLS weights behave in sample data is critical for assessing their merits. Thus, we will now explain the PLS algorithm and its outcome by using a simple example: Consider two blocks of indicators: (a) X, which is a weighted composite of indicators x_1 - x_3 and (b) Y, which is a weighted composite of indicators y_1 - y_3 (Rönkkö & Evermann, 2013). Assume further that X and Y are positively correlated. The PLS weighting algorithm consists of two alternating steps, referred to as inner and outer estimation. The weight algorithm starts by initializing the composites, X and Y, as the unweighted sums of standardized indicators (i.e., unit-weighted) x_1 - x_3 and y_1 - y_3 , respectively. Next, during the first inner estimation step, the composites are recalculated as weighted sums of adjacent composites (i.e., connected by paths in the model). In the present example, X is the only adjacent composite of Y and vice versa. Because X and Y are positively correlated, the composite Y is recalculated as X (i.e., the sum of indicators x_1 - x_3) and vice versa. In the first outer estimation step, the indicator weights are calculated as either: (a) the correlations among composites and indicators (Mode A); or (b) the coefficients from a multiple regression analysis of the composites on the indicators (Mode B),

after which the composites are again recalculated. The two steps are repeated until there is virtually no difference in the indicator weights between two consecutive outer estimation steps³.

To more clearly illustrate the outcome of the PLS weight algorithm, consider a scenario where x_3 and y_1 are, for whatever reason, correlated more strongly with each other than are the other indicators. Therefore, in the first round of outer estimation, these two indicators are given higher weights when updating the composites, leading to an even higher correlation of x_3 and y_1 with their respective composites and increasing the weights further during subsequent outer estimation steps. The PLS algorithm thus produces weights that increase the correlation between the adjacent composites compared to the unit-weighted composites used as the starting point by using any correlations in the data (Rönkkö, 2014b; Rönkkö & Ylitalo, 2010), but this does not guarantee achievement of any global optimum (Krämer, 2006). In more complex models, the weights for a given composite will be contingent on the associations between the indicators of that composite and those adjacent to it, and therefore weights will vary across different model specifications. If a composite is intended to have theoretical meaning, it is difficult to consider the model-dependent nature of the indicator weights as anything but a disadvantage.

To empirically demonstrate the outcomes of the PLS weight algorithm in a real research context, we obtained the data from the third round of the High Performance Manufacturing (HPM) study (Schroeder & Flynn, 2001), in order to replicate the analysis of Peng and Lai (2012)⁴. We first calculated six sets of composites: One set used unit weights, another set was

³ The PLS algorithm can also be applied without the so-called inner estimation step (or alternatively, considering a composite to be only self-adjacent), producing the first principal component of each block (McDonald, 1996). This version of the algorithm is sometimes referred to as “PLS regression” in the literature (Kock, 2015). This labeling of the algorithm is confusing, because: (a) the term “PLS regression” is used also for a completely different algorithm (Hastie, Tibshirani, & Friedman, 2013, Chapter 3; Vinzi, Chin, Henseler, & Wang, 2010); and (b) it obfuscates the fact that the analysis is simply regression with principal components.

⁴ The data consisted of 266 observations, of which 190 were complete. Following Peng and Lai, we used mean substitution to replace the missing values (F. Lai, personal communication, July 15, 2015), although this

based on a replication of Peng and Lai (2012, Figure 3, p. 474), and the remaining four sets were calculated with alternative inner model configurations switching the composite serving as the mediator. The standardized regression coefficients are completely determined by correlations, which we focus on for simplicity.

The correlations in Table 2 show that PLS weights provide no advantage in reliability. If the PLS composites were in fact more reliable, we should expect: (a) all correlations between the PLS composites to be higher than the corresponding correlations between unit-weighted composites; (b) the cross-model correlations between a PLS composite (e.g., Trust with Suppliers) to be higher than the correlations between unit-weighted and PLS composites (see also Rönkkö et al., 2015); and (c) the absolute differences in the correlations between PLS and unit-weighted composites to be larger for large correlations, because the attenuation effect due to measurement error is proportional to the size of the correlation. Instead, we see that PLS weights increase some correlations at the expense of others, depending on which composites are adjacent; in particular, when a correlation is associated with a regression path during PLS weight calculation, the correlation is on average 0.039 *larger* than when the same correlation is not associated with a regression path. If we omit correlations involving the single indicator composite Market Share, which does not use weights, this difference increases to 0.051. Also, the correlations between PLS composites are usually larger than correlations between unit-weighted composites when the composites are adjacent (i.e. associated with regression paths), and smaller when not associated with regression paths (not adjacent). Furthermore, the mean

strategy is known to be suboptimal (Enders, 2010, sec. 2.6). Our replication results were similar, but not identical to the results of Peng and Lai. To ensure that this was not due to differences in PLS software, we also replicated the analysis with SmartPLS 2.0M3, which was used by Peng and Lai. We contacted Peng and Lai to resolve this discrepancy, but received no further replies. To facilitate replication, this and all other analysis scripts used in the current study are included in Appendix A (R code) and Appendix B (Stata code).

correlation between PLS composites and the corresponding unit-weighted composite was always higher than the mean correlation between the cross-model PLS composites. For example, the mean correlation between PLS and unit-weighted composites for Trust with Suppliers was 0.83, but the mean correlation between the different PLS composites using the same data was only 0.72. Finally, no clear pattern emerged regarding how differences between the techniques depend on the size of the correlation.

----- Insert Table 2 about here -----

The results also reveal that model-dependent weights create an additional problem: a composite formed of the same set of indicators can have substantially different weights in different contexts, thus leading to interpretational confounding (Burt, 1976). This issue is apparent regarding the composite for Trusts with Suppliers and the third set of PLS weights, where the correlation between this PLS composite and others calculated using the same data but having different adjacent composites only ranged between 0.24-0.43. In extreme cases, composites calculated using the same data but different PLS weights can even be negatively correlated. The effect is therefore similar to that of factor score indeterminacy discussed by Rigdon (2012).

In sum, PLS indicator weights do not generally provide a meaningful improvement in reliability, a conclusion which is also supported by decades of research on indicator weights (e.g., Bobko et al., 2007; Cohen, 1990; Cohen et al., 2003, pp. 97–98; Grice, 2001a; McDonald, 1996). Moreover, as demonstrated by the empirical example, the model-dependency of the indicator weights leads to instability of the weights, thereby increasing some correlations over others. It is difficult to consider either of these two features as advantages. We therefore find no compelling reason to recommend PLS weights over unit weights, and it is likely that the frequent

assertions regarding the purported advantages of the indicator weighting have done more harm than good in applied research. We will now explain a series of additional, specific methodological problems in PLS.

3. Methodological Problems

3.1. Inconsistent and Biased Estimation

The key problem with approximating latent variables with composites is that the resulting estimator is both inconsistent and biased (Bollen, 1989, pp. 305–306)⁵. Consistency is critically important because it guarantees that the estimates will approach the population value with increasing sample size (Wooldridge, 2009, p. 168). Although the literature about PLS acknowledges that the estimator is not consistent (Dijkstra, 1983), many introductory texts ignore this issue or seemingly dismiss it as trivial. For example, based on the results of a single study (Reinartz, Haenlein, & Henseler, 2009), Hair and colleagues (2014) conclude that: “Thus, the extensively discussed PLS-SEM bias is not relevant for most applications.” (p. 18). What do Reinartz et al.’s results actually show? First, these authors correctly observe that when all latent variables are measured with 8 indicators with loadings of .9, the bias caused by approximating them with composites is trivial (i.e., “ideal scenario”, in their Table 5). However, this is a rather unrealistic situation where the Cronbach’s alphas for the composites are 0.99 in the population, and in such a high reliability setting *any* approach for composite-based approximation will work well (McDonald, 1996). In more realistic conditions, the bias was appreciable: averaged over all conditions, estimates for strong paths (0.5) are biased by -19%, the sole medium strength path (0.3) is biased by -8%, and the estimates for the weak paths (0.15) are biased by +6%; the

⁵ There are a few important exceptions where regression with composites is a consistent estimator of latent variable models, such as model-implied instrumental variables (MIIV; Bollen, 1996), correlation-preserving factor scores (Grice, 2001b); and in special cases, Bartlett factor scores (Skrondal & Laake, 2001).

positive bias for the weaker paths occurs due to capitalization on chance, which we address in the next subsection.

3.2. Capitalization on Chance

The widely-held belief that PLS weights would reduce or eliminate the effect of measurement error rests on the idea that the indicator weights depend on indicator reliabilities. The study by Chin, Marcolin, and Newsted (2003) typifies this reasoning: “indicators with weaker relationships to related indicators and the latent construct are given lower weightings [...] resulting in higher reliability” (p. 194). The claim has some truth to it, because the indicator correlations in such models indeed depend on the reliability of the indicators; however, as mentioned before, decades of research have demonstrated that the advantages of empirically-determined weights are generally small.

A major problem in simulation studies making claims about the reliability of PLS composites is that instead of focusing on assessing the reliability of the composites, these studies focus on path coefficients, typically using setups where measurement error caused regression with unit-weighted composites to be negatively biased (e.g., Chin et al., 2003; Chin & Newsted, 1999; Goodhue, Lewis, & Thompson, 2012; Henseler & Sarstedt, 2013; Reinartz et al., 2009). Thus, the larger coefficients produced by PLS composites were interpreted as evidence of higher reliability of the composites, for example: “PLS performed well, demonstrating its ability to handle measurement error and produce consistent results” (Chin et al., 2003, p. 209). Although better reliability implies larger observed correlations, the converse is not necessarily true, given that the correlations can be larger simply because of capitalization on chance (Goodhue et al., 2015; Rönkkö, 2014b; Rönkkö, Evermann, & Aguirre-Urreta, 2016; Rönkkö et al., 2015). Moreover, there appears to be a lack of awareness in simulation studies of PLS regarding a host

of anomalies signaling capitalization on chance, such as positively biased correlations, path coefficient estimates that become larger with decreasing sample size, non-normal distributions of the estimates, and bias that depends inversely on the size of the population paths (Rönkkö, 2014). As discussed by Rönkkö (2014b), the path coefficients reported by Chin et al. were larger not because of increased reliability, but rather capitalization on chance. This effect is illustrated in Figure 1, where the PLS estimates show a clear bias away from zero, are distributed bimodally (i.e., with two peaks) when the population parameter is close to zero (Rönkkö & Evermann, 2013), or have long negative tails in other scenarios (Henseler et al., 2014).

Although capitalization on chance has been suggested as an explanation for PLS results years ago (e.g., Goodhue, Lewis, & Thompson, 2007), this concern has largely been ignored by the literature about PLS until recently (Rönkkö, 2014b). In contrast, some PLS proponents claimed that this anomaly is beneficial, because it ensures that “the expected value of the PLS-SEM estimator in small samples is closer to the true value than its probability limit (its value in extremely large samples) would predict” (Sarstedt, Ringle, & Hair, 2014, p. 133; see Rönkkö et al., 2015). However, one cannot depend on upward bias due to random error to cancel out downward bias due to measurement error, for two reasons. First, attenuation of a correlation due to measurement error is proportional to the size of the population correlation, whereas the effect of capitalization on chance decreases with increasing sample size (see Figure 1). In general, the size of a correlation is unrelated to sample size. Second, the magnitude of chance sampling variability depends on sample size, whereas measurement error attenuation does not (Rönkkö, 2014b, p. 177). This latter feature of capitalization on chance can be seen in many PLS studies, where the mean parameter estimates increase with decreasing sample size and can sometimes substantially exceed the population values (Rönkkö, 2014b). Given that the disattenuation

formula for correcting correlations for measurement error has now been available for more than 120 years (Spearman, 1904), there is no reason to risk relying on one source of bias to compensate for another.

3.3. Problems in Model Testing

Model testing refers to imposing constraints on the model parameters and then assessing the probability of the observed statistics (e.g., the sample variance-covariance matrix of the observed variables), given the imposed constraints (Lehmann & Romano, 2005, Chapter 3). The principle behind such constraint-based model testing can be illustrated with the model presented by Peng and Lai (2012, Figure 3, p. 474). The model states that the effect of Trust with Suppliers on Customer Satisfaction is fully mediated by Operational Performance. This model imposes a constraint that the correlation between Trust with Suppliers and Customer Satisfaction must equal the product of two other correlations: one between Trust with Suppliers and Operational Performance, and the other between Operational Performance and Customer Satisfaction. Testing of such theory-driven constraints is essential, because if these do not hold true in the data, there is evidence that the model is not an accurate representation of the causal mechanisms that generated the data and any estimates are likely biased. However, as originally presented by Herman Wold, PLS was not intended to impose constraints on the data, making the technique incompatible with the idea of model testing (Dijkstra, 2014; see Rönkkö et al., 2015). As even its proponents sometimes admit: “Since PLS-SEM does not have an adequate global goodness-of-model fit measure [such as chi-square], its use for theory testing and confirmation is limited” (Hair et al., 2014, pp. 17–18). Although statistical testing may not have been the original purpose of PLS, many researchers attempt to apply PLS for model testing purposes (Rönkkö & Evermann, 2013).

To clearly demonstrate that significant lack of model fit can go undetected in a real, empirical PLS study, we used maximum likelihood estimation to test the model presented by Peng and Lai (2012). The chi-square test of exact fit strongly rejected the model $\chi^2(74) = 173.718$ ($p < .001$), indicating that misspecification was present. In addition, even the values for the “approximate fit” indexes did not meet Hu and Bentler’s (1999) guidelines on cutoff criteria (CFI = 0.930, TLI = 0.901, RMSEA = 0.071, 90% CI = 0.057-0.085, SRMR = 0.096)⁶. Moreover, the modification indices suggested that the model cannot adequately account for the correlation between Trust with Suppliers and Customer Satisfaction. Therefore, following recent recommendations for studying mediation (Rungtusanatham, Miller, & Boyer, 2014), we estimated an alternative model that included a direct path between these two latent variables (std. $\beta = 0.479$, $p < .001$). The fit of the model was better but still not exact, $\chi^2(73) = 117.521$ ($p < .001$).⁷ This analysis demonstrates that Peng and Lai’s conclusions regarding full mediation were incorrect. Thus, due to the lack of tools for model testing, it is evident that PLS practitioners will be prone to incorrect causal inference.

Recent guidelines on using PLS make an attempt to assuage concerns about misspecification, suggesting that model quality should not be based on fit tests, but on the “model’s predictive capabilities” (Hair et al., 2014, Chapter 6), and that “fit” in PLS means predictive accuracy (e.g., Henseler & Sarstedt, 2013). However, this advice is problematic: First, multiple equation models are almost invariably overidentified, and if this is the case, there is no excuse for not calculating overidentification tests. Model testing (i.e., assessing fit) is important

⁶ We report these indices for descriptive purposes, noting that they are not useful for model testing (Kline, 2011, Chapter 8). Also, the failure to fit was not due to model complexity or a small sample, as indicated by the Swain correction ($\chi^2_{Swain} = 169.539$ ($p < .001$); the Satorra-Bentler correction for non-normality ($\chi^2_{SB} = 167.067$ ($p < .001$)) gave similar results.

⁷ For comparison purposes, we also estimated this revised model with PLS, resulting in a standardized coefficient of 0.370 for the additional direct path, much higher than the effect of performance (0.094).

to ensure that the specified structural constraints (e.g., zero cross-loadings and error covariances in a CFA model) are correct, because this speaks to the veracity of the underlying theory—estimates from misspecified models can be misleading. Second, predictive accuracy does not reflect model fit (McIntosh et al., 2014), as misspecified models are sometimes more predictive than correctly specified ones (Shmueli, 2010). The fact that PLS cannot test structural constraints does not warrant using measures of predictive accuracy to determine model quality.

Furthermore, Henseler, Hubona, and Ray (2016) state that: “PLS is a limited information estimator and is less affected by model misspecification in some subparts of a model (Antonakis et al., 2010)” (p. 5). Aside from ignoring the fact that Antonakis, Bendahan, Jacquart, and Lalive (2010) specifically cautioned against using PLS given its intractable problems, these statements are problematic because the indicator weights are calculated using information from all adjacent composites. Consider the Peng and Lai (2012) model: If the errors of the indicators of Trust with Suppliers were not independent of the errors of the indicators for Operational Performance, the correlations between the errors would affect the weights of both composites, thereby influencing the estimates of the path from Operational Performance to Customer Satisfaction as well. In contrast, other limited information estimators such as two stage least squares (2SLS) would be unaffected by this type of misspecification. Given the current paucity of research on the performance of PLS with misspecified models and inconsistency of the technique, if a limited information estimator is needed, researchers should instead consider 2SLS, a more established, consistent technique that can be applied to latent variable models (Bollen, 1996).

3.4. Problems in Assessing Measurement Quality

The commonly-used guidelines on applying PLS (e.g., Gefen et al., 2011; Hair et al., 2014; Peng & Lai, 2012) typically suggest that the measurement model be evaluated by

comparing the composite reliability (CR) and average variance extracted (AVE) indices against certain rule-of-thumb cutoffs. Apart from not being statistical tests, the main problem with the CR and AVE indices in a PLS analysis stems from the practice of calculating these statistics based on correlations between the indicators and the composites that they form (as opposed to using factor analysis results), which creates a strong positive bias (Aguirre-Urreta, Marakas, & Ellis, 2013; Evermann & Tate, 2010; Rönkkö & Evermann, 2013). Indeed, simulation studies have demonstrated that these commonly-used statistics cannot detect even severe model misspecifications (Evermann & Tate, 2010; Rönkkö & Evermann, 2013).

Another drawback with using CR and AVE to evaluate measurement models is that neither of these indices can assess the unidimensionality of the indicators, that is, whether they measure the same construct, which renders the resulting composite conceptually ambiguous (Edwards, 2011; Gerbing & Anderson, 1988; Hattie, 1985) as well as makes reliability indices uninterpretable (Cho & Kim, 2015). Introductory PLS texts take three approaches to this problem: (a) ignore the issue altogether (e.g., Hair et al., 2014; Peng & Lai, 2012); (2) state that unidimensionality cannot be assessed based on PLS results, but must be “assumed to be there *a priori*” (Gefen & Straub, 2005, p. 92); or (3) argue incorrectly that the AVE index (e.g., Henseler, Ringle, & Sinkovics, 2009) or CR and Cronbach’s alpha actually test unidimensionality (e.g., Esposito Vinzi, Trinchera, & Amato, 2010, p. 50). Although factor analysis can be used to both assess unidimensionality (see Cho & Kim, 2015 for a review of modern techniques) and produce unbiased loading estimates for calculating the CR and AVE statistics, we have not seen this technique used in conjunction with PLS. In contrast, many researchers incorrectly claim that PLS performs factor analysis (Rönkkö & Evermann, 2013; see e.g., Peng & Lai, 2012; Adomavicius, Curley, Gupta, & Sanyal, 2013; Venkatesh, Chan, & Thong, 2012).

3.5. Use of the One Sample t test Without Observing its Assumptions

After “assessing” the measurement model with the CR and AVE indices, a PLS analysis continues by applying null hypothesis significance testing to the structural path coefficients. Peng and Lai (2012) explain the procedure as follows: “Because PLS does not assume a multivariate normal distribution, traditional parametric-based techniques for significance tests are inappropriate” (p. 472). They then go on to discuss how resampling methods (i.e., bootstrapping), rather than analytical approaches, are needed to estimate the standard errors; yet, the significance of the parameter estimates is assessed by the one-sample t test, which is, ironically, a parametric test and assumes normality of the parameter estimates⁸.

Furthermore, although the parameter estimates in a PLS analysis are simply OLS regression coefficients, the degrees of freedom for the significance tests presented in the introductory texts on PLS do not match those described in the methodological literature on OLS. The significance of OLS regression coefficients is tested by the one-sample t test with $n - k - 1$ degrees of freedom, where n is the number of observations and k is the number of independent variables (Wooldridge, 2009, sec. 4.2). However, introductory texts on PLS argue that the degrees of freedom should be $n - 1$ (Hair et al., 2014, p. 134) or $n + m - 2$,⁹ where m is always 1 and n is the number of bootstrap samples (Henseler et al., 2009, p. 305). Unfortunately, these texts do not explain how the degrees of freedom for the test were derived. When used with large samples, the differences between the results obtained using these three formulas are nearly identical, but with small samples and complex models, the formulas yield differences that can be

⁸ The fact that parametric significance tests may not be appropriate with PLS has been discussed at least since the early 2000’s in the context of multi-group analysis (cf., Sarstedt, Henseler, & Ringle, 2011, p. 199), but this concern is not discussed in the broader PLS literature.

⁹ In the more general statistical literature, this formula is used as the degrees of freedom for the two-sample t test, where n and m are the respective sample sizes (e.g., Efron & Tibshirani, 1993, p. 222).

substantial, where using a reference distribution with larger degrees of freedom leads to smaller p values (i.e., larger Type I error rates).

An additional problem is that neither of the two cited texts, or any other texts on PLS that we have read, explain why the computed test statistic should follow the t distribution under the null hypothesis of no effect. Rönkkö and Evermann (2013; see also McIntosh et al., 2014) recently showed that the PLS estimates are non-normally distributed under the null hypothesis of no effect. As such, the ratio of the estimate and its standard error cannot follow the t distribution, making comparisons against this distribution erroneous. Although Rönkkö and Evermann did not provide evidence on the distribution of the actual test statistic, Rönkkö, McIntosh, et al. (2015) demonstrated that the statistic did not follow the t distribution ($df = n - k - 1$), and comparisons against this reference distribution lead to inflated false positive rates. Although these studies can be criticized for using simplified population models¹⁰, these criticisms miss a crucial point: even though it may be possible to generate simulation scenarios where the parametric one-sample t test happens to work well, it does not work well in all scenarios; thus the test has not been proven to be a general test with known properties in the PLS context. Because a researcher applying the test in empirical work has no way of knowing whether it works properly in her particular situation, the results from the test cannot be trusted.

There are also misconceptions about the properties and justification of using the bootstrap. Many introductory texts incorrectly argue that PLS is a non-parametric technique and

¹⁰ For example, Henseler et al. (2014) argued that calculating the indicator weights based on a single path does not represent how PLS is typically used. However, a recent review by Goodhue et al (2015) indicates that this type of model is common in PLS applications. Indeed, in the Peng and Lai (2012) model, three out of four composites have just one path. (Market Share is not a composite but a single indicator variable used directly as such.)

therefore bootstrapping is required (e.g., Hair, Ringle, & Sarstedt, 2011)¹¹, ignoring the fact that bootstrapping itself has certain assumptions. Most importantly, although some articles use small datasets for demonstration (Efron & Gong, 1983), bootstrapping is generally a large-sample technique (e.g., Efron & Tibshirani, 1993; Davison & Hinkley, 1997; Yung & Bentler, 1996). It is therefore unclear how well this procedure works with PLS when applied to the sample sizes typically used in empirical research (Rönkkö & Evermann, 2013). Although bootstrapping is commonly viewed as being particularly applicable to small samples, this notion is a methodological myth that goes beyond PLS (Koopman, Howe, & Hollenbeck, 2014; Koopman, Howe, Hollenbeck, & Sin, 2015).

A further complication is the use of the so-called sign-change corrections in conjunction with the bootstrap, a procedure implemented in popular PLS software and recommended in some guidelines (Hair, Sarstedt, Ringle, & Mena, 2012; Henseler et al., 2009). These corrections are unsupported by either formal proofs or simulation evidence. Moreover, recent work showed that the individual sign-change corrections result in a 100% false positive rate when used with empirical confidence intervals (Rönkkö et al., 2015). Fortunately, some recent works on PLS are more cautionary toward these corrections (Hair et al., 2014), coupled with admissions that they should never be used (Henseler et al., 2016).

4. The Reasons for Choosing PLS: How Valid Are They?

PLS is typically described by its proponents as the preferred statistical method for evaluating theoretical models when the assumptions of latent variable structural equation

¹¹ A technique is parametric if sample statistics (e.g., the mean vector and covariance matrix) determine the parameter estimates (e.g., Davison & Hinkley, 1997, p. 11), and these statistics completely determine the point estimates of a PLS analysis (cf., Rönkkö, 2014b). The reason why the standard errors of PLS are bootstrapped is because there are no analytical formulas for deriving the standard errors, which is partly because the finite-sample properties of PLS have *still* not been formally analyzed (McDonald, 1996).

modeling (SEM), are unmet (e.g., Hair et al., 2014; Peng & Lai, 2012). The essential claim is that because the ML estimator typically used in SEM has been proven to be optimal only for large samples and multivariate normal data (Kline, 2011, Chapter 7), then PLS should be used in cases where conditions are not met. This rationale for justifying PLS is problematic, for three reasons. First, the fact that an estimator has been proven to be optimal in certain conditions means neither that it *requires* those conditions nor that it would be suboptimal in other scenarios. Second, the inappropriateness of one method does not automatically imply the appropriateness of an alternative method (cf., Rönkkö & Evermann, 2013). Third, such recommendations present a false dichotomy, because PLS is just one of a large number of approaches for calculating scale scores to be used in regression analysis, rather than being the only composite-based alternative to latent variable SEM. Although claims of the advantages of PLS are widespread, these claims are rarely subjected to scrutiny, and formal analyses of the statistical properties of PLS are still lacking (Westland, 2015, Chapter 3). We now address the validity of five claims that have been used to justify the use of PLS over SEM.

4.1. Small Sample Size

PLS is often argued to work well in small samples, but even editors of journals that publish PLS applications express difficulty locating an authoritative source for the claim: In an MISQ editorial, Marcoulides and Saunders (2006, p. iii) stated that they were “frustrated by sweeping claims” about the apparent utility of PLS in small samples, and pointed out that several earlier works on PLS, including some by Wold, emphasized that large sample sizes were clearly preferable to smaller ones. Yet for example, Gefen, Rigdon, and Straub (2011) claim that “PLS path modeling [has] the ability to obtain parameter estimates at relatively lower sample sizes (Chin et al. 2008; Fornell and Bookstein et al. 1982).” (p. vii). This quote represents a fairly

common confusion in the PLS literature; the fact that some estimates can be calculated from the data does not automatically imply that the estimates are useful. As aptly stated by Westland (2015): “Responsible design of software would stop calculation when the information in the data is insufficient to generate meaningful results, thus limiting the potential for publication of false conclusions. Unfortunately, much of the methodological literature associated with PLS software has conflated its ability to generate coefficients without abnormally terminating as equivalent to extracting information.” (p. 42). Indeed, The idea that PLS would perform better in small samples can be traced back to a conference presentation by Wold (1980, partially republished in 1985a), where he presented an empirical application of PLS with 27 variables and 10 observations (Rönkkö & Evermann, 2013). However, the presentation itself made no methodological claims about the small-sample performance of the estimator.

Although ML SEM can be biased in small samples, resorting to PLS is inappropriate, because doing so would replace a potentially biased estimator with an estimator that is both biased and inconsistent (Dijkstra, 1983; see also Rönkkö et al., 2015). In fact, studies comparing PLS and ML SEM with small samples generally demonstrate that ML SEM has less bias (Chumney, 2013; Goodhue et al., 2012; Reinartz et al., 2009)¹². Considering that capitalization on chance is more likely to occur in smaller samples, and that the bootstrap procedures generally work well only in large samples, it difficult to recommend PLS when sample sizes are small. In such situations, a viable alternative is two-stage least squares, which has good small-sample

¹² Recently, Henseler et al. (2014) argued that instead of focusing solely on bias, researchers should evaluate techniques by their ability to converge to admissible solutions, using an example where ML SEM failed to converge to admissible solutions due to severe model misspecification. In this particular scenario, inadmissibility of the estimates tells very little of the small sample behavior of the estimator because the ML estimates calculated from population data were inadmissible as well. Moreover, as explained by McIntosh et al. (2014), the fact that ML SEM did not produce admissible solutions is actually a plus, because it provides a strong indication of model misspecification, which went undetected by PLS.

properties and is non-iterative (i.e., has a closed-form solution), therefore sidestepping potential convergence problems (Bollen, 1996).

4.2. Non-normal Data

PLS has been recommended for handling non-normal data. However, because PLS uses OLS regression analysis for parameter estimation, it inherits the OLS assumptions that errors are normally distributed and homoscedastic (Wooldridge, 2009, Chapter 3). Perhaps the assumption that PLS is appropriate for non-normal data is rooted in Wold's (e.g., 1982, p. 34) claim that using OLS to calibrate a model for prediction makes no distributional assumptions. Although normality is not required for consistency, unbiasedness, or efficiency of the OLS estimator (Wooldridge, 2009, Chapter 3), normality is assumed when using inferential statistical tests. Furthermore, an estimator cannot at the same time have fewer distributional assumptions and work better with smaller samples, because this notion violates the basics of information theory (Rönkkö et al., 2015). To be sure, some recent work on PLS urges researchers to “drop the ‘normality’ or ‘distribution-free’ argument in arguing about the relative merits of PLS and [SEM]” (Dijkstra, 2015, p. 26, see also 2010; Gefen et al., 2011, p. vii; Henseler et al., 2016, Table 2), an assertion that is supported by recent simulations (e.g., Goodhue et al., 2012).

4.3. Prediction vs. Explanation

The original papers on PLS state that its purpose is prediction of the *indicators* (Jöreskog & Wold, 1982), but are vague about the details on how this done (Dijkstra, 2010; McDonald, 1996). Nevertheless, prediction is often highlighted in introductory PLS texts (Peng & Lai, 2012; see also Shah & Goldstein, 2006) and the terms “prediction” or “predictive” are commonly used when justifying the use of PLS in empirical studies (e.g., Johnston, McCutcheon, Stuart, & Kerwood, 2004; Oh, Teo, & Sambamurthy, 2012).

Prediction can have many meanings in OM research. Prediction can take place on the level of a theory because good theories can also be used to predict outcomes (Bacharach, 1989; Wacker, 1998). Prediction on level of a theory can also be useful for understanding of a relationship between two constructs before we can explain why the constructs are related (Singleton & Straits, 2009, Chapter 2), and hypotheses can be considered as predictions of empirical relationships derived from theory (Sutton & Staw, 1995). Statistical prediction, on the other hand, is concerned with constructing models that are used for calculating useful predictions, without any reference to a theory about the relationships among the variables. This approach often relies on “models that are strictly predictive, and are not intended to be explanatory - personnel selection models, expert systems that diagnose or predict problems in complex machines, and quantitative forecasting models would be examples of such models” (Amundson, 1998, p. 342).

Helm, Alaeddini, Stauffer, Bretthauer, and Skolarus’s (2016) recent study is useful for illustrating the difference between statistical prediction and explanation. In this study, demographic data, health history, and hospital stay data available at time of discharge were used to predict the probability of readmission and expected time to readmission for the patient. Calculation of readmission risk on a patient level is useful for hospitals, because it enables targeting post-discharge procedures more efficiently and effectively; it also helps patients, because better post-discharge monitoring allows complications to be identified and treated earlier, thereby avoiding expensive and unnecessary readmissions. Prediction on the observation level is therefore about calculating or forecasting values for individual data points that are yet to be observed and the quality of predictions is judged based on comparing past predictions to their eventual realizations. Forecasting differs from explanation, which is about understanding causal

processes and thus focuses on interpreting data about events that have already occurred. Going back to our example, predicting whether and when specific patients are be readmitted differs from explaining why certain patients are more likely to be readmitted than others; the latter is more within the purview of medical rather than OM research, requires that the model is correctly causally specified, and has important implications for interventions, policy, and theory (because the $X \rightarrow Y$ relationship is a causal one).

Therefore, forecasting outcomes, or testing causal relationships among constructs are two distinct issues that often require different kinds of data, research designs, and analysis tools (Shmueli, 2010). For example, neural network techniques have been shown to produce models that predict well, yet the prediction equations are substantively uninterpretable. The distinction between the design and techniques can be clearly seen in the Helm et al. (2016) article, which contains no mention of hypotheses derived from theory, p values, confidence intervals, or any other type of inferential statistical tools typically seen in theory-testing research, but instead relies on applied machine learning (i.e., automated) techniques. Most notably, one does not require a theory to make relevant predictions.

Although guidelines on using PLS state that it is suitable for prediction, the methodological support for these claims is close to non-existent, and there is not even agreement on what exactly is meant by prediction in the PLS literature (Evermann & Tate, 2016; Shmueli, Ray, Velasquez Estrada, & Chatla, 2016). In a prediction-focused study, it is important to begin by addressing what is being predicted, with what, whether this refers to within-sample or out-of sample predictions, how the predictions were cross-validated, and to present statistics on the estimated prediction errors and intervals (cf., Shmueli et al., 2016). The latter is particularly problematic, because thus far, the PLS literature has only proposed how prediction errors and

intervals could be calculated (Shmueli et al., 2016), and evidence is still lacking on the usefulness of these approaches. Indeed, we have not seen a single comparison of PLS against more modern prediction techniques in a real-world prediction setting.

Although PLS is a potentially useful technique for calculating predictions in scenarios where the structure of the model that generated the data is known, it is not clear when this would be the case in real-world prediction problems. Nonetheless, arguments that PLS is optimal for prediction should demonstrate why this 50-year-old algorithm is superior to modern prediction and machine learning techniques such as neural networks (cf., Bishop, 2006; Hastie, Tibshirani, & Friedman, 2013)¹³. Indeed, although there are a few OM studies that focus on prediction rather than explanation (e.g., Helm et al., 2016; Juran & Schruben, 2004; Karmarkar, 1994; Mazhar, Kara, & Kaebernick, 2007; Vig & Dooley, 1993), none of these studies used PLS, or seem to have considered its use. This is understandable because, except for the R program, *matrixpls* (Rönkkö, 2016b), none of the currently available PLS software can calculate observation-level predictions based on the PLS estimates (Temme, Kreis, & Hildebrandt, 2010).

4.4. Exploratory vs. Confirmatory Modeling

PLS is often recommended for exploratory studies (e.g., Peng & Lai, 2012; Roberts, Thatcher, & Grover, 2010). Unfortunately, many of the guidelines on PLS and studies applying the method do not clearly explain what they mean by exploratory research or how PLS can be used for exploration (Rönkkö & Evermann, 2013). In their review, Rönkkö and Evermann (2013) identified that the term was used for two different purposes in the PLS literature: (a)

¹³ PLS *regression* (a technique similar to principal component regression, where one observed dependent variable is predicted using composites) may still be a useful technique (Hastie et al., 2013, Chapter 3), but it is different from the PLS path modeling approach that we address in this article. That both of these techniques are referred to as PLS in their respective literatures is an unfortunate oversight that continues to confuse applied researchers (Westland, 2015, Chapter 3).

characterizing theory-focused research in its early stages; and (b) extracting patterns or models from the data. Claims that PLS can serve the first purpose rest on the idea that PLS works well when the accuracy of the model is in question, samples small, and measures may be poor (e.g., Roberts et al., 2010, p. 4338). However, considering that heuristics used to evaluate PLS models cannot detect model misspecification and that both small samples and poor measurement increase capitalization on chance, recommending PLS for this situation is unjustifiable (Rönkkö & Evermann, 2013).

As to whether PLS can discover patterns in data, it is useful to initially categorize this type of exploratory modeling into two classes: (a) Whether the purpose of the analysis is searching for models that may have generated the data and could have meaningful theoretical implications; and (b) to simply discover patterns in the data without attributing any theoretical meanings to these patterns. The first class can be further divided into two subclasses, depending on whether a researcher sets to discover the model from the data or starts with an initial theory-based model, which is then modified based on the empirical results. The second, largely descriptive approach can mean “suggesting potential relations among blocks without necessarily making any assumptions regarding which LV model generated the data” (Chin, 2010, p. 663). Such exploratory modeling is amenable to data mining and unsupervised machine learning techniques (Hastie et al., 2013) — techniques that may not be relevant for OM researchers who want to pursue more than merely descriptive summaries of their data.

Modern SEM techniques provide an array of tools for building models from data: Specification searches have been available in SEM software for decades¹⁴, and significant strides have been made in the area of causal discovery algorithms in the SEM literature during recent

¹⁴ We note that these techniques should be used judiciously (Green & Thompson, 2010).

years (see Marcoulides & Ing, 2012, for a review; McIntosh et al., 2014). In contrast, it is difficult to find explanations and evidence on how model discovery can be facilitated with PLS, beyond generic assertions that PLS is useful for this kind of research. In fact, two features of the PLS algorithm speak directly against this kind of use. First, in order to analyze a model in PLS software, practitioners require *a priori* knowledge of both the measurement and structural models. In this sense, therefore, PLS is in fact no different from classical, confirmation-oriented SEM (McIntosh et al., 2014; Rönkkö & Evermann, 2013). Second, although exploratory analysis pertain to situations where the “overall nomological network has not been well understood” (Peng & Lai, 2012, p. 469), PLS uses information from the said nomological network (i.e., the hypothesized pathways among the composites) for calibrating the indicator weights (Dijkstra & Henseler, 2015b; Henseler et al., 2014, 2016). The notion that PLS is oriented toward exploratory modeling thus inevitably leads to a classic Catch-22: Using PLS in an exploratory manner means the nomological network is unknown, but no model estimates are possible in PLS unless one first specifies a model based on the nomological network (Rönkkö & Evermann, 2013).

Ultimately, the question of whether PLS is a useful technique for exploratory analyses may be moot. Although PLS has been advocated for exploratory analyses, it is only rarely used as such. In their review of PLS-based studies in four leading management journals, Rönkkö and Evermann (2013) found that in all cases, results were presented as confirmatory tests of *a priori* hypotheses, precisely as we have found in an informal review of PLS-based articles in JOM.

4.5. Formative Measurement

Although the application of formative measurement models is often used to justify PLS (Peng & Lai, 2012), such an approach to measurement requires flawed assumptions (Edwards,

2011). One such assumption is that indicators have causal potency. However, indicators are simply measures, collected by the researcher, and exist as data used for analysis. These measures might reflect factors that cause latent variables, but the measures do not cause anything (Edwards, 2011; Rigdon, 2013). Equating measures with their corresponding constructs amounts to operationalism, a view of measurement that was discarded decades ago (Campbell, 1960; Messick, 1975). Indeed, formative measurement is actually not measurement at all – in the sense that it describes any real relationships between constructs and their indicators – but rather an unfortunate misnomer for a data aggregation procedure (i.e. indexing; Markus & Borsboom, 2013, p. 172; Rhemtulla, Bork, & Borsboom, 2015).

Aggregating multiple dimensions into a composite is also problematic because it leads to variables that are often uninterpretable (Cadogan & Lee, 2013; Edwards, 2011). All variables, whether they are single observed variables, composites, or latent variables, are fundamentally unidimensional statistical entities. Therefore, to represent multiple dimensions, we need multiple variables (Bollen, 1989, p. 180; Bainter & Bollen, 2015, p. 66). By aggregating several dimensions, one implies that the dimensions that form the aggregate construct always have the same effects on outcomes, which is probably rare (Edwards, 2001). According to Peng and Lai (2012), “with respect to nomological network, one cannot expect that different operational performance items will [...] lead to the same set of consequences. [...] Similarly, the effect of various operational performance dimensions on outcome variables such as business performance can vary considerably (White, 1996)” (pp. 470-471) clearly indicating that aggregating the four dimensions as one variable is not appropriate. Therefore, aggregation only serves obscure the nature of a given concept’s associations with its antecedents and outcomes in the model, thus preventing researchers from achieving clear understanding of causal effects (Cadogan & Lee,

2013; Edwards, 2001). Indeed, firms do not measure any general notion of "operational performance", but rather its multiple dimensions separately, such as cost and quality.

Aside from conceptual problems in formative models, it is not clear why PLS would be the method of choice for estimating such models. The use of PLS is often justified by stating that formative models have certain requirements to be identified (Bollen & Davis, 2009), and by claiming that PLS would either not require identification or could somehow solve the issue of identification (e.g., Peng & Lai, 2012; Roberts et al., 2010). These claims confuse the concepts of statistical model (a set of equations with free parameters) and estimator (a principle or algorithm for calculating values for the free parameters from the data) (Aguirre-Urreta & Marakas, 2014). Both PLS and ML SEM estimates are completely determined by the sample covariance matrix. If two alternative sets of parameter values for a model produce the same population covariance matrix, is impossible to empirically determine from which of the two models a particular sample originated – this is the definition of model identification. Although model non-identification does not prevent estimation, estimates from non-identified models generally cannot be trusted (Henseler et al., 2016; Rigdon, 1995; Rönkkö, Evermann, et al., 2016). Therefore, the fact that a given statistical software package produces estimates without generating a warning does not mean that identification was addressed, but simply that model identification checks had not been programmed into the software.

The formative model states that indicators cause but do not completely determine the latent variable (Chin, 1998; Henseler et al., 2009). However, PLS involves replacing latent variables with composites, resulting in estimation of a misspecified model (Bollen & Diamantopoulos, 2015; Diamantopoulos, 2011). Indeed, Rönkkö, Evermann, and Aguirre-Urreta

(2016) showed that current empirical evidence resoundingly supports ML SEM over PLS for estimating formative models, and their own simulation study reached the same conclusion.

It seems that the link between formative models and Mode B estimation may have happened accidentally, with no mathematical justification. The statistical model proposed by Wold for use with the PLS algorithm was identical to the original LISREL model (Jöreskog & Wold, 1982), which disallows regression paths originating from measured variables (Bollen, 1989, Chapter 9)¹⁵. Rather, the idea appears to originate from an article by Fornell and Bookstein (1982), where the authors confuse the Mode A and Mode B weighting routines with the statistical model being estimated (Rönkkö, Evermann, et al., 2016). However, the model presented by Fornell and Bookstein was not a full formative model, because the error term of the model remained fixed at zero (1982, e.g., p. 446). The first time that a full formative model with an error term for the formative latent variable appeared in the PLS literature seems to be in a book chapter by Fornell and Cha (1994), where the authors also argue that the formative relations and the composite weights are conceptually different, thus contradicting the work by Fornell and Bookstein that presented these as the same. After this book chapter, presenting the formative model and the weights as distinct concepts became a standard practice in the PLS literature (e.g., Chin, 1998, eq. 1, 7, 9; Henseler et al., 2009, eq. 1, 3, 5). Indeed, more recent work has appeared to have rediscovered Wold's original intent: Rather than representing different measurement models, Mode A and Mode B are simply different ways to weight indicators (Rigdon, 2012; Rönkkö, Evermann, et al., 2016). According to Wold (1982), the choice is practical: "*Choose Mode B for the exogenous LV's and Mode A for the endogenous LV's*" (p. 11), which is also

¹⁵ Bollen (1989, pp. 311–312) suggest a workaround in which formative indicators are included as latent variables that are constrained to be equal to their indicators. An advantage of this approach is that it makes the assumption about error-free measurement explicit.

consistent with using the method for predicting the indicators of endogenous composites from the indicators of exogenous composites (Evermann & Tate, 2016).

5. Can the Problems with PLS be Fixed?

In recent years, several attempts have been made to rectify some of the key shortcomings of PLS. However, these purported remedies do not address all problems with the estimator, as the root causes of many of these are that: (1) the sampling distribution of the PLS weights remains unknown; and (2) the weights have been shown to capitalize on chance. Therefore the logical remedy would be to abandon the PLS weights in favor of better understood and robust alternatives such as unit weights (Rönkkö, McIntosh, & Aguirre-Urreta, 2016) thus also obviating the need for specialized software packages. However, this solution does not appear to be acceptable in the pro-PLS literature. In this section, we critically review some of the recent enhancements to PLS and introduce some links to prior literature that are missing from the articles promoting these techniques.

Based on the issues discussed thus far, it is clear that “like a hammer is a suboptimal tool to fix screws, PLS is a suboptimal tool to estimate common factor models” (Henseler, 2014) and that “if the common factor model is correct, the choice between factor-based SEM and PLS path modeling is really no choice at all” (Rigdon, 2012, p. 344). Given that factor-based models are central to theory-focused research¹⁶, it follows that PLS is not useful for this type of application. Regarding this concern, two new streams of work have recently emerged. On one hand, there are efforts to fix the problems of PLS as an estimator for latent variable models; on the other hand, some researchers have started to advocate alternatives to factor models, for which PLS would presumably be suitable. In what follows, we address these two streams of research.

¹⁶ The majority of theory-focused studies in OM use scales with multiple parallel items and assess their reliabilities with Cronbach’s alpha or the CR index, which both assume that the common factor model holds.

5.1. Composite Factor Model

The “composite factor model” (CFM) was introduced into the PLS literature by Henseler et al. (2014) to counter Rönkkö and Evermann (2013)’s criticism toward PLS. The CFM was presented as a generalization of the common factor model where all error correlations between the indicators of a construct are freely estimated. However, the CFM entails serious conceptual sleights of hand, as well as basic statistical shortcomings. For example, Henseler et al.’s (2014) assertion that “PLS is clearly [a] SEM method that is designed to estimate composite factor models” (p. 187) is fundamentally inconsistent with much of the PLS literature, which has traditionally focused on common factor models (McIntosh et al., 2014). Moreover, using the label “composite” is actually a misnomer because the CFM does not actually contain *any composite variables* (McIntosh et al., 2014), and will likely only confuse applied researchers further. Furthermore, the CFM is not identified, and “research resources devoted to estimating and testing an underidentified model may ultimately be wasted” (Rigdon, 1995, p. 359). Thus, we conclude that the CFM is a meaningless construction that serves no purpose in either methodological or applied research.

5.2. Composites as Proxies for Constructs

Another perspective that is currently gaining traction in the PLS literature is to reject the common factor model altogether and instead treat composites as proxies (approximations) of theoretical constructs (conceptual variables). In contrast to the formative measurement perspective, this perspective does not make any explicit claims about the relationships between the indicators and the construct; rather, it merely states that the composite is a potentially valid means of representing the theoretical construct. The use of composites is rationalized by arguing that factor models are problematic because of factor score indeterminacy and because factor

models do not always fit the data well. However, the argument that factors are simply proxies for, rather than isomorphic with, theoretical constructs is a specious critique. It is well-known that at best, factors are simply approximations of constructs and their interrelationships (MacCallum, Browne, & Cai, 2007). Furthermore, that factor models do not always fit the data is true, but responding by shifting to composite-based modeling (Henseler et al., 2016) is misguided, as researchers should instead diagnose the model to discover the sources of misfit, which are themselves informative (Kline, 2011, Chapter 8). Misfit can signify that the indicators are not unidimensional or that the measurement errors are not independent, which are both problems that should be addressed rather than ignored.

Adopting a composite-based approach to modeling implies that measurement models and hence, measurement theory can be disregarded. Ignoring measurement models leaves indicators without theoretical meaning, thus disallowing abstraction beyond the observed data (Bollen & Diamantopoulos, 2015). Conversely, retaining measurement theory, but rejecting just the idea that these theories can be modeled statistically, is at loggerheads with the position that statistical modeling of substantive theory would be useful, essentially leaving a large methodological hole in terms of theory (dis)confirmation. As an alternative to traditional measurement model validation, Rigdon (2012, p. 354) suggests using the composite proxies as “forecasts”, in order to determine “the magnitude of the gap between proxy and concept.” However, given that it is only possible to measure the proxy (and not the concept), it is unclear how one could ever assess the adequacy of the predictions derived from such a framework, or more fundamentally, evaluate the discrepancy between the proxy and the concept (Dijkstra, 2014).

Finally, if composites are merely intended to summarize indicators, then the need for calculating sample-specific weights is dubious, and could even lead to substantively meaningless

results. For example, in the Peng and Lai (2012) study, it is unclear why the observed indicator “On time delivery performance” should receive a weight four times as large as the weight for “Unit cost of manufacturing” (Table 4). There is no apparent theoretical or managerial justification for these weights. A better alternative would be to combine the items based on *a priori* defined index weights (Howell, Breivik, & Wilcox, 2013; Lee & Cadogan, 2013), which facilitates clear interpretation of the composite that does not vary across samples or models.

5.3. Consistent PLS

Another approach to improving PLS has been to make the estimator suitable to factor models. Pioneered by Dijkstra and his colleagues (Dijkstra & Henseler, 2015a, 2015b; Dijkstra & Schermelleh-Engel, 2014) and showcased in recent special issues (Sarstedt et al., 2014), *consistent* PLS (PLSc) removes the inconsistency of PLS estimates by correcting for measurement error. PLSc starts by estimating factor loadings using a modified version of the minimum residuals (MINRES) estimator, an early factor analysis technique (Harman & Jones, 1966) that is now largely obsolete (Bartholomew, Knott, & Moustaki, 2011, sec. 3.9). PLSc augments the MINRES estimator by constraining the factor loadings to be proportional to PLS weights. Next, the correlation matrix of the composites is corrected for attenuation using reliability estimates based on the estimated loadings, and the adjusted correlation matrix is used in OLS estimation¹⁷. This procedure produces a consistent estimator for correctly specified factor models, but it is not without problems.

¹⁷ Dijkstra and Henseler (2015a) note that two stage least squares (2SLS) can be applied instead of OLS regression. More generally, the PLS composites can of course be used with any statistical technique that works with raw data and similarly the disattenuated correlation matrix can be used with any analysis that can be calculated from correlation matrix. The use of OLS (and now 2SLS as well) is mostly a limitation of the currently available PLS software, which can of course be avoided by exporting the composites (or their disattenuated correlations) to a general statistical package, which provides a much broader range of analysis options.

First, why constraining loadings to be proportional to PLS weights produces better estimates than MINRES or other factor analysis techniques is unclear. Indeed, both Huang (2013) and Rönkkö, McIntosh, and Aguirre-Urreta (2016) showed that PLSc is less precise, more biased and has twice the variance compared to ML or MINRES. Second, because the weights are proportional to estimated loadings, indicators with positive loadings are over-weighted, leading to positive bias in the reliability estimates (Rönkkö, McIntosh, et al., 2016). Third, because correlations between the PLS composites are inflated by capitalization on chance, the expected value of these correlations depends on the sample size, making the sample composite correlations biased estimators of their population counterparts. Fourth, the use of parametric significance tests in PLSc (e.g., *t* test) remains problematic for the same reasons described in Section 3.5. To be sure, the negative bias caused by under-correcting for attenuation (due to positively biased reliability estimates) could cancel out the effects of capitalization on chance, resulting in negligible overall bias (Dijkstra & Henseler, 2015a, 2015b). However, this scenario is unlikely in real data, and studies show that the PLSc estimator can be substantially more biased and less efficient than ML SEM (Huang, 2013; Rönkkö, McIntosh, et al., 2016).¹⁸

The problems with PLSc can be eliminated by simply constructing composites with unit-weights and estimating their reliabilities based on traditional factor analysis techniques (Rönkkö, McIntosh, et al., 2016). Although these techniques have been combined with the correction for

¹⁸ Huang (Bentler & Huang, 2014; Huang, 2013) proposed two extensions to PLSc, labeled PLSe1 and PLSe2. These methods are very different from PLS and PLSc, and are in fact conventional SEM techniques that estimate the model by minimizing the discrepancy between the model implied covariance matrix and the data covariance matrix. PLSe1 uses PLSc estimates as starting values for one iteration of normal ML estimation, and PLSe2 is generalized least squares (GLS) estimation, in which the model-implied covariance matrix from PLSc replaces the sample covariance matrix as the weight matrix (Huang, 2013). Although these approaches bring with them the full suite of model testing capabilities offered in SEM software, the advantages these techniques provide over modern starting value techniques, traditional GLS, or iteratively reweighted least squares estimators remain unclear. In any case, labeling these techniques as “PLS” is misleading because parameters are estimated by fitting the model to a covariance matrix instead of calculating composite approximations for latent variables.

attenuation for decades, the correction for attenuation is no longer widely-used because it tends to produce inadmissible correlation estimates and is less efficient than modern SEM techniques (Cohen et al., 2003, pp. 55–57; Muchinsky, 1996); therefore, it is considered obsolete by some researchers (e.g., Dimitruk, Schermelleh-Engel, Kelava, & Moosbrugger, 2007). Moreover, modern SEM techniques are more flexible and efficient (McIntosh et al., 2014; see also Henseler et al., 2016).

5.4. Model Testing and Inference

In addition to improving the properties of PLS as an estimator, recent research has also proposed a number of new approaches for statistical inference and model evaluation. As part of the development of PLSc, Dijkstra and Henseler (2015a) have constructed global fit statistics (like the SEM χ^2 statistic) for PLS to test the discrepancies between the observed and model-implied covariances for the indicator variables, and recent guidelines advocate these new global test statistics (Henseler et al., 2016). Because the sampling distribution of the PLS weights is unknown (Dijkstra, 1983), the test statistics cannot be expected to follow the χ^2 or any other known distribution. Thus, Dijkstra and Henseler (2015a) proposed simulating the reference distribution via bootstrapping. They showed that the technique did not result in excessive Type I error rates; however, there is currently no evidence demonstrating whether this approach can actually detect misspecification.

After recognizing that PLS measurement model assessment indices (e.g., AVE, CR) are unsuitable for examining misspecification, there have been some efforts to develop alternative approaches for assessing measurement models. In particular, Henseler, Ringle, and Sarstedt (2015) developed a discriminant validity statistic that can be computed from the indicator correlation matrix without using PLS output. This statistic was applied by Voorhees, Brady,

Calantone, and Ramirez (2015), who demonstrated its usefulness¹⁹. Although this statistic avoids the use of PLS, the information it provides is less comprehensive than that yielded by confirmatory factor analysis (Cho & Kim, 2015; Shaffer, DeGeest, & Li, 2015).

Lastly, to deal with bimodal coefficient distributions on significance testing, Henseler et al. (2014) suggested using empirical confidence intervals (i.e., lower and upper limits chosen from the bootstrap distribution). Because this method does not require a theoretical probability distribution (the t distribution), one can ignore the normality assumption (Davison & Hinkley, 1997, Chapter 5). However, the scenarios under which empirical CIs have been studied with PLS are limited. Particularly, even though the BCa intervals that correct for skewness in the estimates are promising (Henseler et al., 2014), it is not clear how the approach performs with the bimodal distributions that plague PLS estimates.

6. “PLS-SEM”: A Misleading Label

What do consumers expect in a label? Truth in advertising—a label should be a clear, honest, and informative signal. In the case of PLS, we have recently been witnessing a feat of prestidigitation, namely the explicit advertising of PLS as a SEM technique. A rebranding effort has culminated in relabeling what was known as *Partial Least Squares Path Modeling* (PLS-PM) as *Partial Least Squares Structural Equation Modeling* (PLS-SEM) (e.g., Hair et al., 2014, 2011). This marketing effort has been extremely successful, with the use of PLS increasing about three-fold in the last five years (cf., Rönkkö, 2014b). Concurrently, however, many researchers,

¹⁹ Note that Voorhees, Brady, Calantone, and Ramirez (2015) use a factor correlation of 0.9 as their lack of discriminant validity condition, but tested this with a CFA model where the correlation was constrained at 1. Because the model did not fit the data, they concluded that the CFA model did not perform well in detecting lack of discriminant validity. Because modern SEM software allows using inequality constraints, it would have been more appropriate to constrain the correlation to be greater than 0.9 instead of exactly 1 when testing the model.

including some writing about the method (e.g., Lu, Kwan, Thomas, & Cedzynski, 2011), appear to have forgotten that PLS is simply regression with weighted sums of the indicators.

The question of whether PLS is a SEM technique is not useful, because the discussion reduces to a debate over definitions. PLS is as much SEM as is any other form of regression with scale scores, regardless of indicator weightings (McIntosh et al., 2014; Rönkkö & Evermann, 2013; Rönkkö et al., 2015). However, two questions relevant for research practice are whether: (a) PLS possesses the capabilities that one expects from an SEM technique; and (b) the labeling of PLS as a SEM technique clarifies or obscures what the technique does. The most common capabilities that researchers usually want in SEM techniques are the ability to model latent variables and measurement errors, as well as formally test the fit between the model and the observed data (Chin, 1998, p. 297; Kline, 2011, Chapter 1; Shah & Goldstein, 2006). Based on our assessment thus far, it is clear that PLS, in its original form and as it is currently used, does not provide these capabilities. Moreover, the labeling obscures that PLS is simply a series of OLS regressions with scale scores that does not account for measurement error, as exemplified by the following quote from Oh et al. (2012): “Although a series of multiple regression analyses could be used to test the model, PLS is a better approach because it can simultaneously account for measurement errors for unobserved constructs and examine the significance of structural paths.” (p. 374). Therefore, simply using the “PLS-SEM” label to place the technique in the same league as well-established SEM techniques can undeservedly enhance its reputation and perceived capabilities merely by association.

Marketing PLS as SEM not only obscures what the method actually does and implies capabilities it does not have, but also leads to omission of important analytical steps and even erroneous analyses that could be avoided if the method were simply presented as regression with

scale scores. For example, summarizing multiple indicators as one scale score must be preceded by a factor analysis to analyze the dimensionality of the data and justify aggregating the measures (e.g., Cho & Kim, 2015). Also, the application of OLS regression should always be followed by residual diagnostics (e.g., Fox, 1997), neither of which are followed in empirical applications of PLS or guidelines explaining the technique. However, adopting these techniques would of course not solve the problems inherent in the use of PLS weights for calculating the scale scores.

As an example of an erroneous analysis, consider the study by Liang, Saraf, and Hu (2007), in which PLS was incorrectly used as the vehicle for an SEM-based procedure, namely the “unmeasured latent method construct” (ULMC) model for diagnosing and controlling for common method variance²⁰. This SEM strategy is based on modeling the indicator variance as trait, method, and random error components by regressing each indicator on both a latent variable representing the construct that the indicator measures and a method factor shared by all indicators. Liang et al. implemented this technique by combining the indicators into composites representing the constructs as well as a global “method” composite consisting of all indicators, on which all indicators were also regressed. This modeling strategy cannot remove any method variance from the composites, because no matter how the indicators are weighted, the composites will always be contaminated with the same sources of error as the indicators. Yet, incorrectly believing that PLS would have the capabilities of modern SEM techniques has misled a large number of researchers into believing that the ULMC method as presented by Liang et al. would work (Chin, Thatcher, & Wright, 2012), including a few authors in JOM (Kortmann,

²⁰ The ULMC technique is not without its problems (Richardson, Simmering, & Sturman, 2009), but these weaknesses are not relevant for the purposes of this example.

Gelhard, Zimmermann, & Piller, 2014; Oh et al., 2012; Perols, Zimmermann, & Kortmann, 2013; Rosenzweig, 2009).

7. Some Concluding Remarks

Some OM researchers who are heavily invested in applying PLS might have been disturbed by the recent JOM editorial (Guide & Ketokivi, 2015), which announced that articles using the technique are likely to be desk-rejected. However, JOM readers should carefully consider the core message of the editorial, which states that to justify the PLS method, “rhetorical appeals must be replaced with methodological justification” (Guide & Ketokivi, 2015, p. vii). Merely saying that a method is good for a particular purpose does not make it so. Rather, applied researchers should keep in mind the possibility that the person making the statement is simply: (a) making an honest mistake; or (b) intentionally misrepresenting the merits and weaknesses of the technique. Thus, neither the number of authors advocating PLS nor the number of articles that apply PLS are relevant to the inherent problems with this method. These issues are not matters of opinion; there are fundamental methodological problems that speak for themselves. As such, authors who might consider using PLS should not rely on the weight of opinion or who expressed these opinions, but instead should focus on the substance of the matters involved seeking out and verifying the methodological foundations for any claims. We believe that a critical and unbiased evaluation of these claims will lead to the conclusion that PLS is inferior to other methods for constructing scales scores, particularly the even simpler approach of using unit-weights (Bobko et al., 2007; Cohen, 1990).

Given the host of methodological problems with the PLS technique, a natural query to pose is: How did we get here and how should we move forward? Whereas the increasing popularity of PLS in various applied disciplines is a fairly recent phenomenon, the PLS

algorithm itself was presented already 50 years ago (Wold, 1966) as an iterative procedure for calculating principal components and canonical correlations, and was extended to path models with composites during the 70's (Jöreskog & Wold, 1982). However, the technique never gained traction in the methodological literature because Wold never demonstrated the statistical properties of PLS, and the rationale for some parts of the algorithm remained unclear (Dijkstra, 1983). Instead the method re-emerged through the marketing and information systems disciplines (Hair, Ringle, & Sarstedt, 2012; Rönkkö, 2014a, pp. 18–22), where its popularity can be attributed to a number of guidelines written by users (e.g., Peng & Lai, 2012). At the same time, the PLS method remained almost non-existent in the mainstream research methods journals and textbooks (Rönkkö & Evermann, 2013), presumably because the use of the PLS weights have no firm basis in statistical theory and are methodologically unjustified. One notable exception is an article by McDonald (1996) in *Multivariate Behavioral Research*, about which he later recollected: “PLS (soft modeling) seemed to have acquired something of a mystique among European investigators, yet it was an algorithm without either a model or any clear optimal properties. And graduate students occasionally asked whether they should follow fashion and use it! The mild irritation induced by these facts forced me to an investigation. The resulting [article] discovered global properties for the algorithm and gave half a dozen alternative methods, most of which were easier to apply and clearly better than PLS. I think I wrote the paper too politely. A stronger response to woolly thinking was really needed.” (McDonald, Wainer, & Robinson, 2007, p. 327)

As our review shows, the majority of claims made in the PLS literature should be categorized as statistical and methodological myths and urban legends (SMMULs; Vandenberg, 2006, 2011). SMMULs are common beliefs that typically have some basis in methodological

research, but over time, the original results have been forgotten and replaced with a socially-constructed, accepted truth that has been institutionalized into researcher training and the manuscript review process. As Vandenberg (2006) vividly puts the matter: “Doctoral students may be taught or told something to do within the research process as if it were an absolute truth when in reality it is not, and yet, being who they are, they accept that presumed fact as the ‘truth.’” Similarly, authors may accept something from an editor or a reviewer who in turn was told that ‘this’ is the way it must be as well. [...] The overall end result, however, is a degradation of the whole research process” (pp. 195-196). Indeed, it appears that a large number of researchers have simply accepted, at face value, the methodologically unjustified claims presented in the PLS literature.

To realize the methodological revolution called for by Vandenberg (2011), several parties must play crucial and complementary roles. First, the growing popularity of PLS can in part be attributed to shortcomings in research methods teaching. Most mainstream methodology journals and textbooks ignore PLS altogether (Rönkkö & Evermann, 2013), which in itself can be taken as an indication of the limitation of the technique. However, the virtual absence of PLS from such venues has contributed to an imbalance in the literature toward more positive expositions on the method. Although our article is a step toward correcting this imbalance, we encourage textbook authors to dedicate more space toward educating and cautioning practitioners about the shortcomings of PLS. Second, as per the recommendations of Vandenberg (2011) on tackling SMMULs, journal editors and reviewers require a set of clear and accurate decision rules. To foster more critical thinking about PLS as well as more rigorous review of manuscripts applying the technique, Table 3 presents a list of common claims about PLS, counterarguments to these claims, as well as corresponding questions that PLS adopters should answer. Reviewers and

editors can also use this list in order to hold authors accountable for their choices. Third, authors have a personal responsibility to stay abreast of developments in statistical modeling – particularly the ongoing innovations for estimation and testing in the SEM area – and more carefully reflect on the methodological undergirding of PLS, in light of the present and other critiques. To facilitate this self-educational process, we have included both the R and Stata analysis code for all results presented in this article in Appendix A and B, so researchers can experiment with PLS weights. We would conjecture that increased practitioner awareness will contribute to a reduction in the use of problematic and unjustified methods such as PLS.

The change we call for is not likely to happen overnight, particularly because PLS proponents, whose writings form much of the methodological “canon” for applied researchers in the information systems and marketing domains, have been largely unreceptive to the mounting critiques. The possibility that PLS capitalizes on chance was suggested eight years ago (Goodhue et al., 2007), and the initial explanation of the mechanics underlying this phenomenon dates back five years (Rönkkö & Ylitalo, 2010), as does the observation that the model quality indices commonly used with PLS cannot detect model misspecifications (Evermann & Tate, 2010). In the normal course of research, such serious issues would be expected to quickly attract the attention of the leading scholars and trigger efforts to solve the problems. Instead, we have witnessed a dramatic increase in the promotion of PLS method in certain disciplines largely continuing the tradition of either downplaying or completely ignoring the serious shortcomings of PLS and dismissing works critical of PLS: For example, PLS supporters claim that critical papers contain “prejudiced boycott calls” (Hair, Ringle, et al., 2012, p. 313), suffer from “lack of academic neutrality”, and are “non-constructive and misguided” (Sarstedt et al., 2014, p. 133). One empirical study on several limitations of PLS (i.e., Rönkkö & Evermann, 2013) was referred

to as “nothing short of a polemic” (Sarstedt et al., 2014, p. 134; see also Henseler et al., 2014). At the same time, two new commercial PLS programs, ADANCO (Henseler & Dijkstra, 2014) and SmartPLS 3 (Ringle, Wende, & Becker, 2014) have entered the software market. Whereas conflicts of interests are not unethical per se (and are often unavoidable), nor do they necessarily make research untrustworthy, reviewers and readers have the right to know of their potential existence in order to fully weigh the merits of the arguments presented in a given paper.

Ironically, the primary methodological shortcomings of PLS may actually contribute to its staying power; that is, given the pervasive phenomenon of publication bias (Franco, Malhotra, & Simonovits, 2014), both the bias away from zero (owing to capitalization on chance) and the absence of evidence against the model (due to a lack of model testing procedures) potentially makes PLS appealing to some practitioners (see also Rönkkö & Evermann, 2013). Therefore we emphasize that we are strongly supportive of journal editors — and the recent JOM editorial (Guide & Ketokivi, 2015) — with the dual goals of educating researchers and enforcing a high standard of evidence for supporting methodological choices. Unfortunately for PLS users, the current foundation of methodological support for the technique is too flimsy. Particularly, the PLS literature needs to directly and fully address two issues that are most central to the usefulness of the method, and on which PLS proponents have thus far been silent: (a) capitalization on chance, going beyond the spin-doctoring where it is described as a blessing in disguise (Sarstedt et al., 2014); and (b) clearly articulate the purpose and usefulness of the PLS weights in comparison to more established indicator weighting systems. Until such time as the PLS literature can furnish applied researchers with a solid undergirding of analytical proofs, logical justifications, and strong simulation results regarding the merits of the technique, we would conclude that the only logical and reasonable action stemming from objective

consideration of these issues is to discontinue the use of PLS and instead pursue superior alternatives, namely the ongoing stream of methodological innovations in latent variable-based SEM.

8. References

- Adomavicius, G., Curley, S. P., Gupta, A., & Sanyal, P. (2013). User acceptance of complex electronic market mechanisms: Role of information feedback. *Journal of Operations Management*, *31*(6), 489–503. <http://doi.org/10.1016/j.jom.2013.07.015>
- Aguirre-Urreta, M. I., & Marakas, G. M. (2014). A rejoinder to Rigdon et al. (2014). *Information Systems Research*, *25*(4), 785–788. <http://doi.org/10.1287/isre.2014.0545>
- Aguirre-Urreta, M. I., Marakas, G. M., & Ellis, M. E. (2013). Measurement of composite reliability in research using partial least squares: Some issues and an alternative approach. *The DATA BASE for Advances in Information Systems*, *44*(4), 11–43. <http://doi.org/10.1145/2544415.2544417>
- Amundson, S. D. (1998). Relationships between theory-driven empirical research in operations management and other disciplines. *Journal of Operations Management*, *16*(4), 341–359. [http://doi.org/10.1016/S0272-6963\(98\)00018-7](http://doi.org/10.1016/S0272-6963(98)00018-7)
- Antonakis, J., Bendahan, S., Jacquart, P., & Lalive, R. (2010). On making causal claims: A review and recommendations. *The Leadership Quarterly*, *21*(6), 1086–1120. <http://doi.org/10.1016/j.leaqua.2010.10.010>
- Bacharach, S. B. (1989). Organizational theories: Some criteria for evaluation. *The Academy of Management Review*, *14*(4), 496–515.

- Bainter, S. A., & Bollen, K. A. (2015). Moving forward in the debate on causal indicators: Rejoinder to comments. *Measurement: Interdisciplinary Research and Perspectives*, *13*(1), 63–74. <http://doi.org/10.1080/15366367.2015.1016349>
- Bartholomew, D. J., Knott, M., & Moustaki, I. (2011). *Latent variable models and factor analysis: A unified approach* (3rd ed.). Chichester: Wiley.
- Bentler, P. M., & Huang, W. (2014). On components, latent variables, PLS and simple methods: Reactions to Rigdon's rethinking of PLS. *Long Range Planning*, *47*(3), 138–145. <http://doi.org/10.1016/j.lrp.2014.02.005>
- Bishop, C. M. (2006). *Pattern recognition and machine learning*. New York, NY: Springer.
- Bobko, P., Roth, P. L., & Buster, M. A. (2007). The usefulness of unit weights in creating composite scores: A literature review, application to content validity, and meta-analysis. *Organizational Research Methods*, *10*(4), 689–709. <http://doi.org/10.1177/1094428106294734>
- Bollen, K. A. (1989). *Structural equations with latent variables*. New York, NY: John Wiley & Son Inc.
- Bollen, K. A. (1996). An alternative two stage least squares (2SLS) estimator for latent variable equations. *Psychometrika*, *61*(1), 109–121. <http://doi.org/10.1007/BF02296961>
- Bollen, K. A., & Davis, W. R. (2009). Causal indicator models: Identification, estimation, and testing. *Structural Equation Modeling: A Multidisciplinary Journal*, *16*(3), 498–522. <http://doi.org/10.1080/10705510903008253>
- Bollen, K. A., & Diamantopoulos, A. (2015). In defense of causal-formative indicators: A minority report. *Psychological Methods*. <http://doi.org/10.1037/met0000056>

- Burt, R. S. (1976). Interpretational confounding of unobserved variables in structural equation models. *Sociological Methods & Research*, 5(1), 3–52.
<http://doi.org/10.1177/004912417600500101>
- Cadogan, J. W., & Lee, N. (2013). Improper use of endogenous formative variables. *Journal of Business Research*, 66(2), 233–241. <http://doi.org/10.1016/j.jbusres.2012.08.006>
- Campbell, D. T. (1960). Recommendations for APA test standards regarding construct, trait, or discriminant validity. *American Psychologist*, 15(8), 546–553.
- Chin, W. W. (1998). The partial least squares approach to structural equation modeling. In G. A. Marcoulides (Ed.), *Modern methods for business research* (pp. 295–336). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.
- Chin, W. W. (2010). How to write up and report PLS analyses. In V. Esposito Vinzi, W. W. Chin, J. Henseler, & H. Wang (Eds.), *Handbook of partial least squares* (pp. 655–690). Berlin Heidelberg: Springer.
- Chin, W. W., Marcolin, B. L., & Newsted, P. R. (2003). A partial least squares latent variable modeling approach for measuring interaction effects: Results from a monte carlo simulation study and an electronic-mail emotion/adoption study. *Information Systems Research*, 14(2), 189–217.
- Chin, W. W., & Newsted, P. R. (1999). Structural equation modeling analysis with small samples using partial least squares. In R. H. Hoyle (Ed.), *Statistical strategies for small sample research* (pp. 307–342). Thousand Oaks, CA: Sage Publications.
- Chin, W. W., Thatcher, J. B., & Wright, R. T. (2012). Assessing common method bias: Problems with the ULMC technique. *MIS Quarterly*, 36(3), 1003–1019.

Cho, E., & Kim, S. (2015). Cronbach's coefficient alpha well known but poorly understood.

Organizational Research Methods, 18(2), 207–230.

<http://doi.org/10.1177/1094428114555994>

Chumney, F. (2013). *Structural equation models with small samples: A comparative study of four approaches* (Doctoral dissertation). University of Nebraska-Lincoln, Lincoln, NE.

Retrieved from <http://digitalcommons.unl.edu/cehsdiss/189>

Cohen, J. (1990). Things I have learned (so far). *American Psychologist*, 45(12), 1304–1312.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences*. London: Lawrence Erlbaum Associates.

Dana, J., & Dawes, R. M. (2004). The superiority of simple alternatives to regression for social science predictions. *Journal of Educational and Behavioral Statistics*, 29(3), 317–331.

<http://doi.org/10.3102/10769986029003317>

Davison, A. C., & Hinkley, D. V. (1997). *Bootstrap methods and their application*. Cambridge, NY: Cambridge University Press.

Diamantopoulos, A. (2011). Incorporating formative measures into covariance-based structural equation models. *MIS Quarterly*, 35(2), 335–358.

Dijkstra, T. K. (1983). Some comments on maximum likelihood and partial least squares methods. *Journal of Econometrics*, 22(1-2), 67–90.

Dijkstra, T. K. (2010). Latent variables and indices: Herman Wold's basic design and partial least squares. In V. Esposito Vinzi, W. W. Chin, J. Henseler, & H. Wang (Eds.), *Handbook of partial least squares* (pp. 23–46). Berlin Heidelberg: Springer.

- Dijkstra, T. K. (2014). PLS' Janus face – Response to professor Rigdon's "Rethinking partial least squares modeling: In praise of simple methods." *Long Range Planning*, 47(3), 146–153. <http://doi.org/10.1016/j.lrp.2014.02.004>
- Dijkstra, T. K. (2015, June). *PLS & CB SEM, a weary and a fresh look at presumed antagonists (keynote address)*. Presented at the 2nd International Symposium on PLS Path Modeling, Sevilla, Spain. Retrieved from [http://www.researchgate.net/publication/277816598_PLS__CB_SEM_a_weary_and_a_fresh_look_at_presumed_antagonists_\(keynote_address\)](http://www.researchgate.net/publication/277816598_PLS__CB_SEM_a_weary_and_a_fresh_look_at_presumed_antagonists_(keynote_address))
- Dijkstra, T. K., & Henseler, J. (2015a). Consistent and asymptotically normal PLS estimators for linear structural equations. *Computational Statistics & Data Analysis*, 81, 10–23. <http://doi.org/10.1016/j.csda.2014.07.008>
- Dijkstra, T. K., & Henseler, J. (2015b). Consistent partial least squares path modeling. *MIS Quarterly*, 39(2), 297–316.
- Dijkstra, T. K., & Schermelleh-Engel, K. (2014). Consistent partial least squares for nonlinear structural equation models. *Psychometrika*, 79(4), 585–604. <http://doi.org/10.1007/s11336-013-9370-0>
- Dimitruk, P., Schermelleh-Engel, K., Kelava, A., & Moosbrugger, H. (2007). Challenges in nonlinear structural equation modeling. *Methodology: European Journal of Research Methods for the Behavioral & Social Sciences*, 3(3), 100–114. <http://doi.org/10.1027/1614-2241.3.3.100>
- Edwards, J. R. (2001). Multidimensional constructs in organizational behavior research: An integrative analytical framework. *Organizational Research Methods*, 4(2), 144–192.

- Edwards, J. R. (2011). The fallacy of formative measurement. *Organizational Research Methods*, *14*(2), 370–388. <http://doi.org/10.1177/1094428110378369>
- Efron, B., & Gong, G. (1983). A leisurely look at the bootstrap, the jackknife, and cross-validation. *The American Statistician*, *37*(1), 36–48.
- Efron, B., & Tibshirani, R. J. (1993). *An introduction to the bootstrap*. New York, NY: Chapman and Hall/CRC.
- Enders, C. K. (2010). *Applied missing data analysis*. New York, NY: Guilford Publications.
- Esposito Vinzi, V., Trinchera, L., & Amato, S. (2010). PLS path modeling: From foundations to recent developments and open issues for model assessment and improvement. In V. Esposito Vinzi, W. W. Chin, J. Henseler, & H. Wang (Eds.), *Handbook of partial least squares* (pp. 47–82). Berlin Heidelberg: Springer.
- Evermann, J., & Tate, M. (2010). Testing models or fitting models? Identifying model misspecification in PLS. In *ICIS 2010 Proceedings*. St. Louis, MO. Retrieved from http://aisel.aisnet.org/icis2010_submissions/21
- Evermann, J., & Tate, M. (2016). Assessing the predictive performance of structural equation model estimators. *Journal of Business Research*. <http://doi.org/10.1016/j.jbusres.2016.03.050>
- Fornell, C., & Bookstein, F. L. (1982). Two structural equation models: LISREL and PLS applied to consumer exit-voice theory. *Journal of Marketing Research*, *19*(4), 440–452.
- Fornell, C., & Cha, J. (1994). Partial least squares. In R. P. Bagozzi (Ed.), *Advanced methods of marketing research* (pp. 52–78). Cambridge, MA: Blackwell.
- Fox, J. (1997). *Applied regression analysis, linear models, and related methods*. Thousand Oaks, CA: SAGE Publications.

- Franco, A., Malhotra, N., & Simonovits, G. (2014). Publication bias in the social sciences: Unlocking the file drawer. *Science*, *345*(6203), 1502–1505.
<http://doi.org/10.1126/science.1255484>
- Gefen, D., Rigdon, E. E., & Straub, D. W. (2011). An update and extension to SEM guidelines for administrative and social science research. *MIS Quarterly*, *35*(2), iii–xiv.
- Gefen, D., & Straub, D. W. (2005). A practical guide to factorial validity using PLS-Graph: Tutorial and annotated example. *Communications of the Association for Information Systems*, *16*(1), 91–109.
- Gerbing, D. W., & Anderson, J. C. (1988). An updated paradigm for scale development incorporating unidimensionality and its assessment. *Journal of Marketing Research*, *25*(2), 186–192.
- Goodhue, D. L., Lewis, W., & Thompson, R. (2007). Statistical power in analyzing interaction effects: questioning the advantage of PLS with product indicators. *Information Systems Research*, *18*(2), 211–227.
- Goodhue, D. L., Lewis, W., & Thompson, R. (2012). Does PLS have advantages for small sample size or non-normal data. *MIS Quarterly*, *36*(3), 981–1001.
- Goodhue, D. L., Lewis, W., & Thompson, R. (2015). PLS pluses and minuses in path estimation accuracy. In *AMCIS 2015 Proceedings*. San Juan, Puerto Rico. Retrieved from <http://aisel.aisnet.org/amcis2015/ISPhil/GeneralPresentations/3>
- Green, S., & Thompson, M. (2010). Can specification searches be useful for hypothesis generation? *Journal of Modern Applied Statistical Methods*, *9*(1), Article 16.
- Grice, J. W. (2001a). A comparison of factor scores under conditions of factor obliquity. *Psychological Methods*, *6*(1), 67–83. <http://doi.org/10.1037//1082-989X.6.1.67>

- Grice, J. W. (2001b). Computing and evaluating factor scores. *Psychological Methods*, 6(4), 430–450.
- Guide, D., & Ketokivi, M. (2015). Notes from the editors: Redefining some methodological criteria for the journal. *Journal of Operations Management*, 37, v–viii.
[http://doi.org/10.1016/S0272-6963\(15\)00056-X](http://doi.org/10.1016/S0272-6963(15)00056-X)
- Hair, J. F., Hult, G. T. M., Ringle, C. M., & Sarstedt, M. (2014). *A primer on partial least squares structural equations modeling (PLS-SEM)*. Los Angeles, CA: SAGE Publications.
- Hair, J. F., Ringle, C. M., & Sarstedt, M. (2011). PLS-SEM: Indeed a silver bullet. *Journal of Marketing Theory & Practice*, 19(2), 139–152. <http://doi.org/10.2753/MTP1069-6679190202>
- Hair, J. F., Ringle, C. M., & Sarstedt, M. (2012). Partial least squares: The better approach to structural equation modeling? *Long Range Planning*, 45(5–6), 312–319.
<http://doi.org/10.1016/j.lrp.2012.09.011>
- Hair, J. F., Sarstedt, M., Ringle, C. M., & Mena, J. A. (2012). An assessment of the use of partial least squares structural equation modeling in marketing research. *Journal of the Academy of Marketing Science*, 40(3), 414–433. <http://doi.org/10.1007/s11747-011-0261-6>
- Harman, H. H., & Jones, W. H. (1966). Factor analysis by minimizing residuals (minres). *Psychometrika*, 31(3), 351–368. <http://doi.org/10.1007/BF02289468>
- Hastie, T., Tibshirani, R., & Friedman, J. H. (2013). *The elements of statistical learning: Data mining, inference, and prediction* (2nd ed. 2009. Corr. 10th printing 2013 edition). New York, NY: Springer. Retrieved from
http://statweb.stanford.edu/~tibs/ElemStatLearn/printings/ESLII_print10.pdf

- Hattie, J. (1985). Methodology review: Assessing unidimensionality of tests and items. *Applied Psychological Measurement*, 9(2), 139–164. <http://doi.org/10.1177/014662168500900204>
- Helm, J. E., Alaeddini, A., Stauffer, J. M., Bretthauer, K. M., & Skolarus, T. A. (2016). Reducing hospital readmissions by integrating empirical prediction with resource optimization. *Production and Operations Management*, 25(2), 233–257. <http://doi.org/10.1111/poms.12377>
- Henseler, J. (2014). Common beliefs and reality about PLS. Retrieved from <http://managementink.wordpress.com/2014/05/23/common-beliefs-and-reality-about-pls/>
- Henseler, J., & Dijkstra, T. K. (2014). ADANCO (Version 1.0). Kleve, Germany: Composite Modeling. Retrieved from <http://www.composite-modeling.com/>
- Henseler, J., Dijkstra, T. K., Sarstedt, M., Ringle, C. M., Diamantopoulos, A., Straub, D. W., ... Calantone, R. J. (2014). Common beliefs and reality about PLS: Comments on Rönkkö and Evermann (2013). *Organizational Research Methods*, 17(2), 182–209. <http://doi.org/10.1177/1094428114526928>
- Henseler, J., Hubona, G., & Ray, P. A. (2016). Using PLS path modeling in new technology research: Updated guidelines. *Industrial Management & Data Systems*, 116(1), 2–20. <http://doi.org/10.1108/IMDS-09-2015-0382>
- Henseler, J., Ringle, C. M., & Sarstedt, M. (2015). A new criterion for assessing discriminant validity in variance-based structural equation modeling. *Journal of the Academy of Marketing Science*, 43(1), 115–135.
- Henseler, J., Ringle, C. M., & Sinkovics, R. R. (2009). The use of partial least squares path modeling in international marketing. *Advances in International Marketing*, 20, 277–319.

- Henseler, J., & Sarstedt, M. (2013). Goodness-of-fit indices for partial least squares path modeling. *Computational Statistics*, 28(2), 565–580. <http://doi.org/10.1007/s00180-012-0317-1>
- Howell, R. D., Breivik, E., & Wilcox, J. B. (2013). Formative measurement: A critical perspective. *The DATA BASE for Advances in Information Systems*, 44(4), 44–55.
- Huang, W. (2013). *PLSe: Efficient estimators and tests for partial least squares* (Doctoral dissertation). University of California, Los Angeles, CA. Retrieved from <http://escholarship.org/uc/item/2cs2g2b0>
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1–55.
- Hwang, H., & Takane, Y. (2004). Generalized structured component analysis. *Psychometrika*, 69(1), 81–99. <http://doi.org/10.1007/BF02295841>
- Johnston, D. A., McCutcheon, D. M., Stuart, F. I., & Kerwood, H. (2004). Effects of supplier trust on performance of cooperative supplier relationships. *Journal of Operations Management*, 22(1), 23–38. <http://doi.org/10.1016/j.jom.2003.12.001>
- Jöreskog, K. G., & Wold, H. (1982). The ML and PLS techniques for modeling with latent variables. In K. G. Jöreskog & H. Wold (Eds.), *Systems under indirect observation: Causality, structure, prediction*. Amsterdam: North-Holland.
- Juran, D. C., & Schruben, L. W. (2004). Using worker personality and demographic information to improve system performance prediction. *Journal of Operations Management*, 22(4), 355–367. <http://doi.org/10.1016/j.jom.2004.05.003>

- Karmarkar, U. S. (1994). A robust forecasting technique for inventory and leadtime management. *Journal of Operations Management*, 12(1), 45–54. [http://doi.org/10.1016/0272-6963\(94\)90005-1](http://doi.org/10.1016/0272-6963(94)90005-1)
- Kettenring, J. R. (1971). Canonical analysis of several sets of variables. *Biometrika*, 58(3), 433–451.
- Kline, R. B. (2011). *Principles and practice of structural equation modeling* (3rd ed.). New York, NY: Guilford Press.
- Kock, N. (2015). PLS-based SEM algorithms: The good neighbor assumption, collinearity, and nonlinearity. *Information Management and Business Review*, 7(2), 113–130.
- Koopman, J., Howe, M., & Hollenbeck, J. R. (2014). Pulling the Sobel test up by its bootstraps. In C. E. Lance & R. J. Vandenberg (Eds.), *More statistical and methodological myths and urban legends* (pp. 224–244). New York: Routledge.
- Koopman, J., Howe, M., Hollenbeck, J. R., & Sin, H.-P. (2015). Small sample mediation testing: Misplaced confidence in bootstrapped confidence intervals. *Journal of Applied Psychology*, 100(1), 194–202.
- Kortmann, S., Gelhard, C., Zimmermann, C., & Piller, F. T. (2014). Linking strategic flexibility and operational efficiency: The mediating role of ambidextrous operational capabilities. *Journal of Operations Management*, 32(7–8), 475–490. <http://doi.org/10.1016/j.jom.2014.09.007>
- Krämer, N. (2006). *Analysis of high-dimensional data with partial least squares and boosting* (Doctoral dissertation). Technische Universität Berlin, Berlin. Retrieved from <https://opus4.kobv.de/opus4-tuberlin/frontdoor/index/index/docId/1446>

- Lee, N., & Cadogan, J. W. (2013). Problems with formative and higher-order reflective variables. *Journal of Business Research*, *66*(2), 242–247.
<http://doi.org/10.1016/j.jbusres.2012.08.004>
- Lehmann, E. L., & Romano, J. P. (2005). *Testing statistical hypotheses* (3rd ed.). New York, NY: Springer.
- Liang, H., Saraf, N., Hu, Q., & Xue, Y. (2007). Assimilation of enterprise systems: The effect of institutional pressures and the mediating role of top management. *MIS Quarterly*, *31*(1), 59–87.
- Lu, I. R. R., Kwan, E., Thomas, D. R., & Cedzynski, M. (2011). Two new methods for estimating structural equation models: An illustration and a comparison with two established methods. *International Journal of Research in Marketing*, *28*(3), 258–268.
<http://doi.org/10.1016/j.ijresmar.2011.03.006>
- MacCallum, R. C., Browne, M. W., & Cai, L. (2007). Factor analysis models as approximations. In R. Cudeck & R. C. MacCallum (Eds.), *Factor analysis at 100: Historical developments and future directions* (pp. 153–175). Mahwah, NJ: Lawrence Erlbaum Associates.
- Marcoulides, G. A., & Ing, M. (2012). Automated structural equation modeling strategies. In R. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 690–704). New York, NY: Guilford Press.
- Marcoulides, G. A., & Saunders, C. (2006). PLS: A silver bullet? *MIS Quarterly*, *30*(2), iii–ix.
- Markus, K. A., & Borsboom, D. (2013). *Frontiers in test validity theory: Measurement, causation and meaning*. New York, NY: Psychology Press.

- Mazhar, M. I., Kara, S., & Kaebernick, H. (2007). Remaining life estimation of used components in consumer products: Life cycle data analysis by Weibull and artificial neural networks. *Journal of Operations Management*, 25(6), 1184–1193.
<http://doi.org/10.1016/j.jom.2007.01.021>
- McDonald, R. P. (1996). Path analysis with composite variables. *Multivariate Behavioral Research*, 31(2), 239–270.
- McDonald, R. P., Wainer, H., & Robinson, D. H. (2007). Interview with Roderick P. McDonald. *Journal of Educational and Behavioral Statistics*, 32(3), 315–332.
<http://doi.org/10.3102/1076998607305835>
- McIntosh, C. N., Edwards, J. R., & Antonakis, J. (2014). Reflections on partial least squares path modeling. *Organizational Research Methods*, 17(2), 210–251.
<http://doi.org/10.1177/1094428114529165>
- Messick, S. (1975). The standard problem: Meaning and values in measurement and evaluation. *American Psychologist*, 30(10), 955–966. <http://doi.org/10.1037/0003-066X.30.10.955>
- Muchinsky, P. M. (1996). The correction for attenuation. *Educational and Psychological Measurement*, 56(1), 63–75. <http://doi.org/10.1177/0013164496056001004>
- Oh, L.-B., Teo, H.-H., & Sambamurthy, V. (2012). The effects of retail channel integration through the use of information technologies on firm performance. *Journal of Operations Management*, 30(5), 368–381. <http://doi.org/10.1016/j.jom.2012.03.001>
- Peng, D. X., & Lai, F. (2012). Using partial least squares in operations management research: A practical guideline and summary of past research. *Journal of Operations Management*, 30(6), 467–480. <http://doi.org/10.1016/j.jom.2012.06.002>

- Perols, J., Zimmermann, C., & Kortmann, S. (2013). On the relationship between supplier integration and time-to-market. *Journal of Operations Management*, 31(3), 153–167.
<http://doi.org/10.1016/j.jom.2012.11.002>
- Reinartz, W. J., Haenlein, M., & Henseler, J. (2009). An empirical comparison of the efficacy of covariance-based and variance-based SEM. *International Journal of Research in Marketing*, 26(4), 332–344. <http://doi.org/10.1016/j.ijresmar.2009.08.001>
- Reise, S. P., & Revicki, D. A. (2014). *Handbook of item response theory modeling*. New York, NY: Taylor & Francis.
- Rhemtulla, M., Bork, R. van, & Borsboom, D. (2015). Calling models with causal indicators “measurement models” implies more than they can deliver. *Measurement: Interdisciplinary Research and Perspectives*, 13(1), 59–62.
<http://doi.org/10.1080/15366367.2015.1016343>
- Richardson, H. A., Simmering, M. J., & Sturman, M. C. (2009). A tale of three perspectives: Examining post hoc statistical techniques for detection and correction of common method variance. *Organizational Research Methods*, 12(4), 762–800.
<http://doi.org/10.1177/1094428109332834>
- Rigdon, E. E. (1995). A necessary and sufficient identification rule for structural models estimated in practice. *Multivariate Behavioral Research*, 30(3), 359–383.
http://doi.org/10.1207/s15327906mbr3003_4
- Rigdon, E. E. (2012). Rethinking partial least squares path modeling: In praise of simple methods. *Long Range Planning*, 45(5–6), 341–358.
<http://doi.org/10.1016/j.lrp.2012.09.010>

- Rigdon, E. E. (2013). Lee, Cadogan, and Chamberlain: An excellent point . . . But what about that iceberg? *AMS Review*, 3(1), 24–29. <http://doi.org/10.1007/s13162-013-0034-0>
- Ringle, C. M., Wende, S., & Becker, J.-M. (2014). *SmartPLS 3*. Hamburg, Germany: SmartPLS. Retrieved from <http://www.smartpls.com>
- Roberts, N., Thatcher, J. B., & Grover, V. (2010). Advancing operations management theory using exploratory structural equation modelling techniques. *International Journal of Production Research*, 48(15), 4329–4353. <http://doi.org/10.1080/00207540902991682>
- Rönkkö, M. (2014a). *Methodological myths in management research: Essays on partial least squares and formative measurement* (Doctoral dissertation). Aalto University, Espoo. Retrieved from <http://urn.fi/URN:ISBN:978-952-60-5733-0>
- Rönkkö, M. (2014b). The effects of chance correlations on partial least squares path modeling. *Organizational Research Methods*, 17(2), 164–181. <http://doi.org/10.1177/1094428114525667>.
- Rönkkö, M. (2016a). An overview of the matrixpls package. Retrieved from <https://cran.r-project.org/web/packages/matrixpls/vignettes/matrixpls-intro.pdf>
- Rönkkö, M. (2016b). matrixpls: Matrix-based partial least squares estimation (Version 1.0.0). Retrieved from <https://cran.r-project.org/web/packages/matrixpls/index.html>
- Rönkkö, M., & Evermann, J. (2013). A critical examination of common beliefs about partial least squares path modeling. *Organizational Research Methods*, 16(3), 425–448. <http://doi.org/10.1177/1094428112474693>
- Rönkkö, M., Evermann, J., & Aguirre-Urreta, M. I. (2016). Estimating formative measurement models in IS research: Analysis of the past and recommendations for the future.

- Unpublished Working Paper*. Retrieved from <http://urn.fi/URN:NBN:fi:aalto-201605031907>
- Rönkkö, M., McIntosh, C. N., & Aguirre-Urreta, M. I. (2016). Improvements to PLSc: Remaining problems and simple solutions. *Unpublished Working Paper*. Retrieved from <http://urn.fi/URN:NBN:fi:aalto-201603051463>
- Rönkkö, M., McIntosh, C. N., & Antonakis, J. (2015). On the adoption of partial least squares in psychological research: Caveat emptor. *Personality and Individual Differences*, (87), 76–84. <http://doi.org/10.1016/j.paid.2015.07.019>
- Rönkkö, M., & Ylitalo, J. (2010). Construct validity in partial least squares path modeling. In *ICIS 2010 Proceedings*. St. Louis, MO. Retrieved from http://aisel.aisnet.org/icis2010_submissions/155
- Rosenzweig, E. D. (2009). A contingent view of e-collaboration and performance in manufacturing. *Journal of Operations Management*, 27(6), 462–478. <http://doi.org/10.1016/j.jom.2009.03.001>
- Rungtusanatham, M., Miller, J. W., & Boyer, K. K. (2014). Theorizing, testing, and concluding for mediation in SCM research: Tutorial and procedural recommendations. *Journal of Operations Management*, 32(3), 99–113. <http://doi.org/10.1016/j.jom.2014.01.002>
- Sarstedt, M., Henseler, J., & Ringle, C. M. (2011). Multigroup analysis in partial least squares (PLS) path modeling: Alternative methods and empirical results. *Advances in International Marketing*, 22, 195–218. [http://doi.org/10.1108/S1474-7979\(2011\)0000022012](http://doi.org/10.1108/S1474-7979(2011)0000022012)
- Sarstedt, M., Ringle, C. M., & Hair, J. F. (2014). PLS-SEM: Looking back and moving forward. *Long Range Planning*, 47(3), 132–137. <http://doi.org/10.1016/j.lrp.2014.02.008>

- Schroeder, R. G., & Flynn, B. B. (2001). *High performance manufacturing: Global perspectives*. New York, NY: Wiley.
- Shaffer, J. A., DeGeest, D., & Li, A. (2015). Tackling the problem of construct proliferation: A guide to assessing the discriminant validity of conceptually related constructs. *Organizational Research Methods*. <http://doi.org/10.1177/1094428115598239>
- Shah, R., & Goldstein, S. M. (2006). Use of structural equation modeling in operations management research: Looking back and forward. *Journal of Operations Management*, 24(2), 148–169. <http://doi.org/10.1016/j.jom.2005.05.001>
- Shmueli, G. (2010). To explain or to predict? *Statistical Science*, 25(3), 289–310. <http://doi.org/10.1214/10-STS330>
- Shmueli, G., & Koppius, O. (2011). Predictive analytics in information systems research. *MIS Quarterly*, 35(3), 553–572.
- Shmueli, G., Ray, S., Velasquez Estrada, J. M., & Chatla, S. B. (2016). The elephant in the room: Predictive performance of PLS models. *Journal of Business Research*. <http://doi.org/10.1016/j.jbusres.2016.03.049>
- Singleton, R. A. J., & Straits, B. C. (2009). *Approaches to social research* (5th ed.). New York, NY: Oxford University Press.
- Skrondal, A., & Laake, P. (2001). Regression among factor scores. *Psychometrika*, 66(4), 563–575. <http://doi.org/10.1007/BF02296196>
- Sosik, J. J., Kahai, S. S., & Piovoso, M. J. (2009). Silver bullet or voodoo statistics?: A primer for using the partial least squares data analytic technique in group and organization research. *Group Organization Management*, 34(1), 5–36. <http://doi.org/10.1177/1059601108329198>

- Spearman, C. (1904). The proof and measurement of association between two things. *American Journal of Psychology*, *15*(1), 72–101.
- Sutton, R. I., & Staw, B. M. (1995). What theory is not. *Administrative Science Quarterly*, *40*(3), 371–384.
- Temme, D., Kreis, H., & Hildebrandt, L. (2010). A comparison of current PLS path modeling software: Features, ease-of-use, and performance. In V. Esposito Vinzi, J. Henseler, W. W. Chin, & H. Wang (Eds.), *Handbook of partial least squares* (pp. 737–756). Berlin Heidelberg: Springer.
- Thomson, G. H. (1938). Methods of estimating mental factors. *Nature*, *141*, 246.
- Vandenberg, R. J. (2006). Introduction: Statistical and methodological myths and urban legends. *Organizational Research Methods*, *9*(2), 194–201.
<http://doi.org/10.1177/1094428105285506>
- Vandenberg, R. J. (2011). The revolution with a solution: All is not quiet on the statistical and methodological myths and urban legends front. In D. D. Bergh & D. J. Ketchen (Eds.), *Building methodological bridges* (Vol. 6, pp. 237–257). Bingley: Emerald Group Publishing.
- Venkatesh, V., Chan, F. K. Y., & Thong, J. Y. L. (2012). Designing e-government services: Key service attributes and citizens' preference structures. *Journal of Operations Management*, *30*(1–2), 116–133. <http://doi.org/10.1016/j.jom.2011.10.001>
- Vig, M. M., & Dooley, K. J. (1993). Mixing static and dynamic flowtime estimates for due-date assignment. *Journal of Operations Management*, *11*(1), 67–79.
[http://doi.org/10.1016/0272-6963\(93\)90034-M](http://doi.org/10.1016/0272-6963(93)90034-M)

- Vinzi, V. E., Chin, W. W., Henseler, J., & Wang, H. (2010). Editorial: Perspectives on partial least squares. In V. E. Vinzi, W. W. Chin, J. Henseler, & H. Wang (Eds.), *Handbook of partial least squares*. Berlin Heidelberg: Springer.
- Voorhees, C. M., Brady, M. K., Calantone, R., & Ramirez, E. (2015). Discriminant validity testing in marketing: An analysis, causes for concern, and proposed remedies. *Journal of the Academy of Marketing Science*, 44(1), 119–134. <http://doi.org/10.1007/s11747-015-0455-4>
- Wacker, J. G. (1998). A definition of theory: research guidelines for different theory-building research methods in operations management. *Journal of Operations Management*, 16(4), 361–385. [http://doi.org/10.1016/S0272-6963\(98\)00019-9](http://doi.org/10.1016/S0272-6963(98)00019-9)
- Westland, J. C. (2015). *Structural Equation Models* (Vol. 22). Cham: Springer International Publishing. Retrieved from <http://link.springer.com/10.1007/978-3-319-16507-3>
- Widaman, K. F. (2007). Common factors versus components: Principals and principles, errors and misconceptions. In R. Cudeck & R. C. MacCallum (Eds.), *Factor analysis at 100: Historical developments and future directions* (pp. 177–203). Mahwah, N.J: Lawrence Erlbaum Associates.
- Wold, H. (1966). Nonlinear estimation by iterative least squares procedures. In F. N. David (Ed.), *Research papers in statistics: Festschrift for J. Neyman* (pp. 411–444). London: Wiley.
- Wold, H. (1980). Factors influencing the outcome of economic sanctions: an application of soft modeling. Presented at the Fourth World Congress of Econometric Society, Aix-en-Provence, France.

- Wold, H. (1982). Soft modeling: The basic design and some extensions. In K. G. Jöreskog & S. Wold (Eds.), *Systems under indirect observation: Causality, structure, prediction* (pp. 1–54). Amsterdam: North-Holland.
- Wold, H. (1985a). Factors influencing the outcome of economic sanctions. *Trabajos de Estadística Y de Investigación Operativa*, 36(3), 325–338.
<http://doi.org/10.1007/BF02888567>
- Wold, H. (1985b). Systems analysis by partial least squares. In P. Nijkamp, L. Leitner, & N. Wrigley (Eds.), *Measuring the unmeasurable* (pp. 221–252). Dordrecht, Germany: Marinus Nijhoff.
- Wooldridge, J. M. (2009). *Introductory econometrics: A modern approach* (4th ed). Mason, OH: South Western, Cengage Learning.
- Yung, Y.-F., & Bentler, P. M. (1996). Bootstrapping techniques in analysis of mean and covariance structures. In G. A. Marcoulides & R. E. Schumacker (Eds.), *Advanced structural equation modeling: Issues and techniques* (pp. 195–226). Hoboken, NJ: Psychology Press.

9. Tables and figures

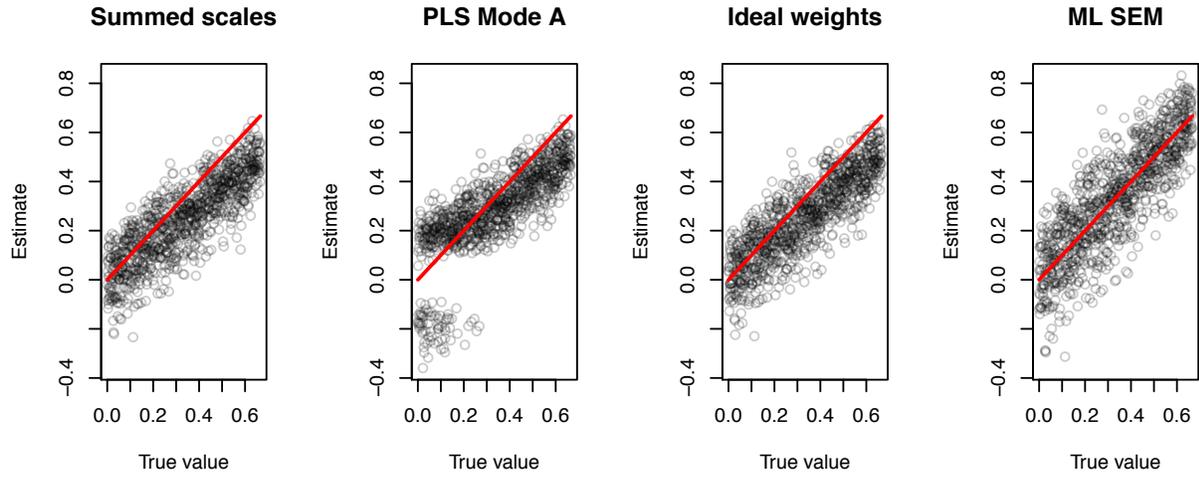


Figure 1 Comparison of regression with summed scales and PLS Mode A composites with latent variable SEM

Table 1: Inconsistency on the purpose of PLS weights across different sources

Purpose of weights	Source and illustrative quotes	Our explanation and interpretation
Minimize the residual variances of regressions between inner approximation composites and observed variables.	<p>“all residual variances in the weight relations are minimized jointly” (Wold, 1982, p. 25)</p> <p>“the PLS algorithm does provide an overall minimum of the residual variances of the weight relations.” (Wold, 1985b, p. 237)</p>	The weight relations refer to the inner approximation composites, which are different from the final composites from outer approximation.
Maximize the variance of the indicators explained by the composites	<p>“mode A minimizes the trace of the residual variances in the "outer" (measurement)” (Fornell & Bookstein, 1982, p. 442)</p> <p>“[when modeling reflective indicators] the weights are meant to form the single best score to maximally predict its own measures (i.e., the first principal component)” (Chin, 2010, p. 664)</p>	Maximizing the explained variance in the indicators is the calculation criterion for principal components.
To maximize predictive power.	<p>“Residual variances are minimized to enhance optimal predictive power.” (Fornell & Bookstein, 1982, p. 443)</p> <p>“PLS optimally weights and combines items of multi-item variables to increase the predictive power of a model.” (Sosik, Kahai, & Piovosio, 2009, p. 16)</p>	Predictive power should not be confused with explanatory power, often quantified with the R^2 statistic (Shmueli & Koppius, 2011). None of the articles specify what they mean by predictive power.
To minimize the sum of residual variances in all model equations (regressions of composites on composites and regressions of indicators on composites)	<p>“PLS aims to minimize the trace (sum of the diagonal elements) of [residual covariances of regressions of composites on composites] and, with reflective specification, also [residual covariances of regressions of indicators on composites]” (Fornell & Bookstein, 1982, p. 443)</p> <p>“weights [...] that creates a score that jointly accounts for as much variance as possible in [indicators], and [endogenous composites].” (Chin, 1998, pp. 301–302)</p>	This is the optimization criterion of generalized structured component analysis (Hwang & Takane, 2004) when all blocks are specified as “reflective”.
To create composites that are maximally correlated with adjacent composites	<p>“the objective is to obtain weights that create the best variate or construct score such that it maximally correlates with the neighboring constructs.” (Chin, 2010)</p>	This is a special case of the optimization criterion of generalized canonical correlation analysis (Kettenring, 1971).
Minimizing the random measurement error in the composites, i.e. maximize composite reliability.	<p>“Optimization of [the] weights aims to maximize the explained variance of dependent variables. [...] Since random measurement error is by definition not predictable, maximizing explained variance will also tend to minimize the presence of random measurement error in these latent variable proxies” (Gefen et al., 2011, p. v)</p> <p>“indicators with weaker relationships to related indicators and the latent construct are given lower weightings resulting in higher reliability for the construct estimate” (Chin et al., 2003, p. 194).</p>	This is the purpose of the regression method for calculating factor scores.
Maximize the explained variance in the endogenous composites	<p>“PLS-SEM uses available data to estimate the path relationships in the model with the objective of minimizing the error terms (i.e., the residual variance) of the endogenous constructs.” (Hair et al., 2014, p. 16)</p> <p>“Minimizes the amount of unexplained variance (i.e., maximizes the R^2 values)” (Hair et al., 2014, p. 16).</p>	This is the optimization criterion of generalized structured component analysis (Hwang & Takane, 2004) when all blocks are specified as “formative”.

Table 2 Correlations between unit weighted composites and five sets of PLS composites

	Unit weights					PLS, original					PLS, alternative 1					PLS, alternative 2					PLS, alternative 3					PLS, alternative 4											
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29								
1 TRUST																																					
2 CROSSFUN	.23																																				
3 MARKET	.01	-.10																																			
4 SATISFACTION	.35	.27	.03																																		
5 PERFORMANCE	.14	.30	.18	.17																																	
6 TRUST	1.00	.23	.01	.37	.15																																
7 CROSSFUN	.23	1.00	-.10	.27	.30	.23																															
8 MARKET	.01	-.10	1.00	.03	.18	.01	-.10																														
9 SATISFACTION	.36	.26	.03	.99	.19	.38	.26	.03																													
10 PERFORMANCE	.15	.29	.19	.19	.98	.16	.30	.19	.20																												
11 TRUST	.99	.24	.02	.37	.15	1.00	.24	.02	.38	.16																											
12 CROSSFUN	.24	1.00	-.10	.27	.30	.24	1.00	-.10	.27	.29	.25																										
13 MARKET	.01	-.10	1.00	.03	.18	.01	-.10	1.00	.03	.19	.02	-.10																									
14 SATISFACTION	.36	.26	.04	1.00	.18	.38	.26	.04	1.00	.20	.39	.27	.04																								
15 PERFORMANCE	.15	.28	.19	.19	.93	.16	.29	.19	.20	.98	.16	.28	.19	.19																							
16 TRUST	.97	.26	.03	.34	.14	.98	.26	.03	.36	.15	.99	.27	.03	.36	.16																						
17 CROSSFUN	.23	1.00	-.10	.27	.30	.23	1.00	-.10	.26	.30	.24	1.00	-.10	.26	.28	.26																					
18 MARKET	.01	-.10	1.00	.03	.18	.01	-.10	1.00	.03	.19	.02	-.10	1.00	.04	.19	.03	-.10																				
19 SATISFACTION	.35	.27	.03	1.00	.17	.36	.27	.03	.99	.19	.37	.27	.03	1.00	.19	.34	.27	.03																			
20 PERFORMANCE	.14	.31	.15	.14	.95	.14	.32	.15	.16	.94	.14	.31	.15	.15	.90	.14	.31	.15	.14																		
21 TRUST	.21	.20	.12	.11	.02	.24	.20	.12	.12	.04	.31	.21	.12	.12	.06	.43	.20	.12	.11	.05																	
22 CROSSFUN	.23	1.00	-.10	.27	.30	.23	1.00	-.10	.26	.29	.24	1.00	-.10	.26	.28	.26	1.00	-.10	.27	.31	.20																
23 MARKET	.01	-.10	1.00	.03	.18	.01	-.10	1.00	.03	.19	.02	-.10	1.00	.04	.19	.03	-.10	1.00	.03	.15	.12	-.10															
24 SATISFACTION	.40	.25	.05	.97	.19	.42	.25	.05	.98	.21	.42	.25	.05	.98	.21	.39	.25	.05	.96	.17	.12	.25	.05														
25 PERFORMANCE	.14	.23	.21	.18	.88	.14	.23	.21	.19	.90	.14	.23	.21	.18	.90	.14	.23	.21	.18	.74	.05	.23	.21	.19													
26 TRUST	.99	.23	.01	.38	.15	1.00	.23	.01	.39	.16	1.00	.24	.01	.40	.16	.97	.23	.01	.38	.15	.25	.23	.01	.43	.15												
27 CROSSFUN	.23	1.00	-.10	.27	.30	.24	1.00	-.10	.26	.29	.25	1.00	-.10	.27	.28	.26	1.00	-.10	.27	.31	.20	1.00	-.10	.25	.23	.23											
28 MARKET	.01	-.10	1.00	.03	.18	.01	-.10	1.00	.03	.19	.02	-.10	1.00	.04	.19	.03	-.10	1.00	.03	.15	.12	-.10	1.00	.05	.21	.01	-.10										
29 SATISFACTION	.36	.27	.03	1.00	.18	.38	.27	.03	1.00	.19	.38	.27	.03	1.00	.19	.35	.27	.03	1.00	.15	.11	.26	.03	.98	.18	.39	.27	.03									
30 PERFORMANCE	.14	.20	.16	.23	.77	.14	.21	.16	.23	.85	.13	.21	.16	.23	.84	.13	.21	.16	.23	.65	.00	.20	.16	.23	.78	.14	.21	.16	.23								

Cells in bold are correlations that correspond to a regression paths between composites.

Table 3 Claims, counter arguments, and questions to consider when evaluating PLS

Claim	Counter argument	Questions to consider
PLS produces optimal weights.	It is not clear for what specific purpose the weights would be optimal; there are no proofs or simulation evidence supporting the optimality claim.	What is the purpose of the indicator weights and why do you think that PLS is the optimal computational algorithm for that purpose? (particularly compared to unit-weights, principal components, well-known factor score methods, and direct optimization of the weights with respect to particular statistic).
PLS reduces or corrects for the effects of measurement error.	Weighting the indicators cannot produce a meaningful improvement over unit weights. Many published simulation studies have misinterpreted coefficients that are larger due to capitalization on chance as evidence of increased reliability of the composites.	Can you provide any simulation evidence that shows a meaningful improvement in reliability?
Although capitalization on chance may occur in some models and with small samples, its impact is exaggerated.	Evidence of capitalization on chance can be seen in virtually all simulation studies of PLS (Rönkkö, 2014b). Given that even in theory the potential advantages of indicator weights are small, capitalization on chance is the main difference between PLS and unit weights.	How can you demonstrate that your analysis is not subject to capitalization on chance? How stable are the indicator weights across alternative model configurations?
PLS can test a model or produces model quality diagnostics.	The model quality heuristics commonly used in a PLS analysis cannot reliably detect model misspecification. The standard of evidence for regression with composite variables does not depend on which particular set of weights is used: an analysis should include a proper factor analysis and regression residual diagnostics.	How do you know that your model is correctly specified?
PLS can be used for null hypothesis significance testing (i.e. p values).	The sampling distribution of PLS weights is unknown, which means that there is no known reference distribution that the test statistic (estimate/standard error) can be compared against. The literature provides conflicting recommendations on which reference distribution to use, and simulation evidence shows that the test statistic does not necessarily approximate the t distribution.	Which reference distribution is used to calculate p values? How do you know that the test statistics follow the said distribution when the null hypothesis holds, and under which conditions?
Bootstrapping and empirical confidence intervals can be used to overcome the problems with NHST.	Bootstrapping is generally a large sample technique, and assumes that the distribution of the bootstrap estimates follow the sampling distribution of the original statistics. It is not clear when this would be the case with PLS, because this has not been addressed in the literature (Rönkkö & Evermann, 2013). Moreover, even though the BCa intervals correct for skewness, it is not clear how these intervals perform when the distribution is bimodal.	If using bootstrapping, how do you know that you have a sufficiently large sample size, given that bootstrapping is a large sample technique? How do you know that your confidence intervals are valid (i.e., have appropriate coverage and balance)?
The concerns about PLS	The composite factor model is not identified	If you reject the measurement

<p>apply to common factor models, but not component factor models, formative measurement models, or when composites are used as proxies for constructs without using any explicit measurement model.</p>	<p>and therefore cannot be meaningfully estimated. Formative measurement builds on assumptions that are difficult to defend, and it is also not clear why one would expect PLS to be a good estimator for this type of model. Approximating constructs with composites without reference to any measurement model is problematic, because it implies rejecting measurement theory.</p>	<p>theories commonly used with factor models, what exactly is your measurement theory and how do you assess measurement? How do you ensure that the model is identified?</p>
<p>PLS is the most appropriate tool because the purpose of the analysis is prediction.</p>	<p>Prediction and explanation are not the same. It is not clear why PLS, developed more than 50 years ago, should be expected to outperform modern prediction techniques, such as neural networks. The current evidence of the predictive capabilities of PLS is very limited and we have no tested tools for evaluating the accuracy of the predictions</p>	<p>What specifically is predicted and with what (cf., Shmueli et al., 2016)? How is the accuracy of the predictions assessed? How were the predictions validated and how did they compare against predictions calculated with more modern techniques?</p>
<p>PLS is the most appropriate tool because the study is explorative.</p>	<p>What exactly do you mean with the term “explorative” (cf., Rönkkö & Evermann, 2013). The literature about PLS does not provide any evidence that PLS would be useful for discovering explanatory models from data.</p>	<p>How do you know that your model is correctly specified?</p>
<p>The recent methodological developments (e.g. PLSc) have completely addressed the problems of PLS.</p>	<p>The PLS weights provide no advantage over more established approaches such as unit weights and the issue of capitalization on chance remains completely unaddressed.</p>	<p>How exactly have the more recent developments addressed the fundamental weakness that PLS weights capitalize on chance?</p>
<p>Although PLS is not without problems, it is nevertheless a convenient technique that is recommended and used by many researchers.</p>	<p>That a technique is simple to use, and having a simple user interface, does not mean that it produces useful results and could contribute to creating “push-button” statisticians who do not understand the analytic foundations of what they are doing (cf., Antonakis et al., 2010).</p>	<p>Do you understand what the PLS algorithm is actually doing? Do you believe that your results are actually theoretically and practically meaningful, in light of the various methodological limitations of PLS?</p>