Author(s): Evermann, Jöerg; Rönkkö, Mikko

Title: Recent Developments in PLS

Year: 2023

Version: Published version

Communications of the Association for Information Systems

Volume 52 Paper 29

6-19-2023

Recent Developments in PLS

Jöerg Evermann
Memorial University of Newfoundland

Mikko Rönkkö
University of Jyväskylä

Recommended Citation

Evermann, J., & Rönkkö, M. (2023). Recent Developments in PLS. Communications of the Association for Information Systems, 52, 663-667. https://doi.org/10.17705/1CAIS.05229

Communications of the Association for Information Systems

Debate Article DOI: 10.17705/1CAIS.05229 ISSN: 1529-3181

Recent Developments in PLS

Jöerg Evermann
Memorial University of Newfoundland
St. John's, NL, Canada

Mikko Rönkkö
University of Jyväskylä
Jyväskylä, Finland

Abstract:

Partial least squares (PLS) path modeling is a widely used method in the information systems (IS) discipline for estimating linear structural equation models. At the same time, researchers have debated its relative merits compared to simple summed scores or to covariance-based estimation of structural equation models. In this paper, we comment on recent developments in PLS to ensure that IS researchers have up-to-date methodological knowledge and best practices if they decide to use PLS. In particular, we briefly review its mechanisms, its well-known properties, and its usage history in IS research. We briefly revisit a high-impact critique and debate a few years ago to identify the critical arguments around current PLS practices and use. That critique proved to be the driver for many advances in the PLS method and its applications which we discuss extensively and use to make 14 recommendations for how and when to use PLS or alternatives.

Keywords: Partial Least Squares, PLS, Structural Equation Modeling, Statistics, Research Methods.

This manuscript underwent editorial review. It was received in 2021. John Venable served as Associate Editor.

1 Introduction

Partial least squares (PLS) path modeling is a statistical method used to estimate linear structural equation models. It originated with Herman Wold (Wold, 1982) as an alternative to the then nascent covariance estimation that Karl Jöreskog developed (Jöreskog & Wold, 1982). IS researchers were among the earliest PLS adopters in the social sciences (Chin & Gopal, 1995; Chin et al., 1996, 2003; Chin & Newsted, 1999). PLS has been increasingly popular not only in the IS discipline (Ringle et al., 2012) but in many other disciplines, such as marketing (Henseler et al., 2009; Hair et al., 2012a), tourism (Do Valle & Assaker, 2016), and hospitality (Ali et al., 2018). Researchers have also introduced the technique to other business disciplines such as operations management (Peng & Lai, 2012) and strategic management (Hulland, 1999; Hair et al., 2012b) but it gained less traction there.

In 2012, realizing that many claims that PLS proponents made lacked sufficient substantiation and the growing voices that cautioned researchers on using PLS (Goodhue et al., 2007, 2012; Aguirre-Urreta & Marakas, 2008; Rouse & Corbitt, 2008; Reinartz et al., 2009; Evermann & Tate, 2010; Rönkkö & Ylitalo, 2010), we independently began to empirically investigate how PLS behaves. After we became aware of each other's research at the 2010 International Conference on Information Systems (Evermann & Tate, 2010; Rönkkö & Ylitalo, 2010), we joined forces to critically assess PLS using a basic structural equation model in a journal paper published in 2013 (Rönkkö & Evermann, 2013). Motivated by our critique, researchers have made many positive developments around PLS in recent years both to further investigate its performance and, more importantly, to improve the method itself and its use. In sum, we know a lot more about how PLS behaves and we know better how and when to use it.

While acknowledging our critical perspective in this paper, we do not intend to argue for or against using PLS. Instead, we highlight recent developments to ensure that IS researchers have up-to-date methodological knowledge about PLS if they decide to use it. As we present in this paper, we have seen a veritable flood of new developments in the past five to 10 years. In this paper, we also help to set minimum standards for methodological competence when using PLS. In essence, *if* researchers choose to use PLS, they must keep the latest advances for the method in mind and use best practices.

We first clarify what we mean by "minimum standards". Consider a study that uses a sample size of 50 to estimate a complex common factor model with PLS mode A, uses t-tests to perform significance tests on weak path estimates, justifies its use of PLS by appealing to its predictive capabilities while reporting only R^2 and focusing on properties of the factor model. This is clearly inadequate and inspires little confidence in its validity, for reasons discussed below. On the other hand, a paper that uses consistent PLS with a large sample size based on an a priori simulation power analysis and that reports fit statistics to indicate a well-fitting model while citing methodological research that supports using these procedures inspires significantly more confidence in its validity. However, researchers will often find themselves between these two ideals. In that situation, they need to recognize the pros and cons and the trade-offs inherent in every statistical technique and argue persuasively based on substance (i.e., what we know based on methodological research) and not based on an appeal to authority or prior studies.

Again, we do not intend to argue for or against using PLS compared to other methods; we leave it to readers to come to their own conclusions. Moreover, in reviewing recent developments, we do not engage with the many "review and guideline" papers (Sarstedt et al., 2020; Ringle et al., 2020; Ali et al., 2018; Do Valle & Assaker, 2016; Hulland 1999; Hair, Sarstedt et al., 2012; Hair et al., 2017; Henseler et al., 2009; Hair et al., 2012; Peng & Lai, 2012; Ringle et al., 2012) and polemic "advocacy" pieces (Hair et al., 2011, 2013; Petter, 2018; Hair et al., 2019) in the PLS literature that seem, to be broadly scattered across many research domains and yet contribute no new substantive knowledge. Instead, we focus on studies that present either algebraic deductions or convincing evidence from simulation studies to make their point.

This paper proceeds as follows: In Section 2, we review linear structural equation models, PLS, and the history of PLS use in IS research. In Section 3, we revisit our 2013 paper in which we critique PLS as a starting point for this commentary. In Section 4, we present recent PLS methodological developments and make recommendations for researchers based on the presented arguments and evidence. In Section 5, we discuss alternatives to PLS. Finally, in Section 6, we conclude the paper.

2 A Brief Introduction to PLS and its Use in IS Research

Structural equation modeling refers to methods to estimate the parameters of a system of (typically linear) equations that include observed (manifest) and unobserved (latent) variables. PLS is one method that is commonly used to estimate the parameters of such a system of linear regressions.

2.1 Structural Equation Models

Structural equation models differ in type. While one can classify models in different ways, for our purposes, we distinguish them based on the way in which they connect observed and unobserved variables.

First, consider the common factor model that we show in Figure 1. The variables ξ and η are latent variables, while the x_i are manifest variables. For every dependent variable (or every regression equation), an error term (ϵ_i, ζ) exists. The figure shows the parameters of the linear regression relationships on the arrows. Additionally, every independent variable (ϵ_i, ξ, ζ) has a variance; one does not usually show the variances. Covariances exist between all exogenous variables (which double-headed arrows show if they are part of the estimated model). Researchers typically assume (as a model constraint) that most of these covariances equal zero (e.g., those between the ϵ_i and ξ or between ξ and ζ). However, they may relax this assumption for a subset of the indicators as long as the model remains statistically identified.

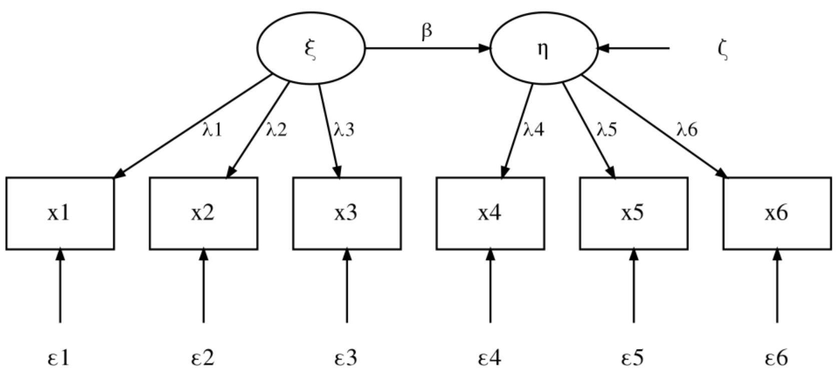

Figure 1. Common Factor Model: Two Latent Variables Each with Three Indicators and Measurement Errors

Another model type is the mixed formative-reflective model such as that in Figure 2. ξ is a formative latent variable (i.e., a function of some manifest variables), and some of the x_i are independent variables in this model. In this case, one need not estimate their covariances since one observes them, and they are not parameters in the model. Note also that ξ is a dependent variable in this model and, therefore, has an associated error term ζ_1.

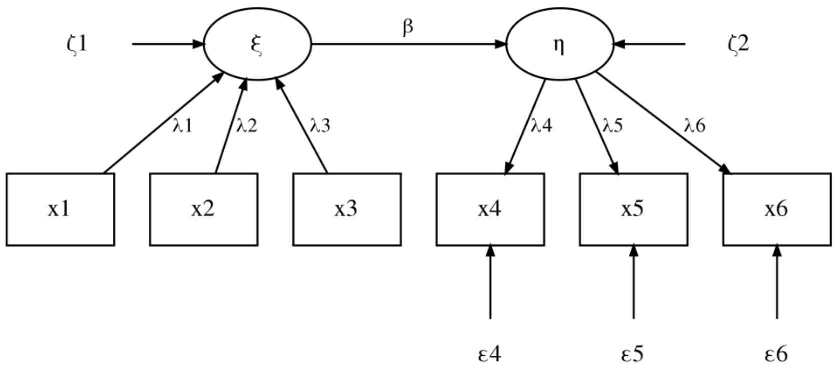

Figure 2. Mixed Formative-reflective Model (We Omit Correlations among the Exogenous x Indicators for Clarity)

The literature contains considerable debate about the usefulness and appropriateness of different model types for various research applications (Cadogan & Lee, 2013; Edwards, 2011; Howell et al., 2007a, 2007b; Bagozzi 2007; Wilcox et al., 2008; Hardin, 2017; Markus, 2018; Guyon, 2018). This debate has occurred in the IS discipline as well (Marakas et al., 2007, 2008; Hardin et al., 2008). However, we do not address that debate in this paper; instead, we assume that a researcher has identified a suitable statistical model and seeks a method to estimate its parameters.

2.2 Principles of PLS

Partial least squares (PLS) path modeling is a method for estimating structural equation models. With that said, we acknowledge that researchers have debated whether PLS constitutes a good technique for doing so. Nonetheless, IS researchers widely use PLS to estimate all of these model types (Ringle et al., 2012). In this section, we describe classic PLS. We describe consistent PLS (”PLSc”)—a significant extension to it— in Section 4.1.

PLS is a two-phase method. In the first phase, it calculates a weighted sum (composite) from each block of manifest variables (i.e., from each set of manifest variables linked to the same latent variable). In other words, PLS estimates the kind of model that we show in Figure 3, which uses composites to approximate the latent variables. In our example model, the two variables ξ and η are weighted sums (“aggregates”, “composites”) of the observed component variables x_i. Note that the model lacks an error term for ξ, while the error term ζ on η is supposed to represent the error term of the structural regression only (see Section 4.8 for further discussion). We can see that the composites can only be imperfect approximations (”proxies”) of the latent variables in the common factor and the formative model: Figure 1 shows that the latent variables in the common factor model are not functions of the manifest ones as is the case for the PLS composite but rather vice versa. Figure 2 shows that formative latent variables have an associated error term, which is not part of the PLS composite.

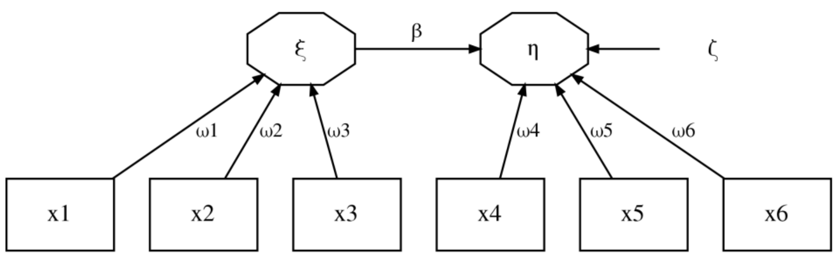

Figure 3. Composite Model (We Omit Correlations among the x Indicators for Clarity)

PLS forms composites by weighting the manifest variables using the following iterative algorithm. It begins with equal weights w and iterates until the change in weights is below a threshold between successive iterations. PLS begins with an “outer approximation” step to determine “outer weights”. To do so, it uses either mode A, which regresses indicators on the composite ($\hat{\eta} = \sum_{j \in k_\eta} w_j\, x_j + d_j$, similarly for ξ), or mode B, which regresses the composite on the indicators ($x_j = w_j \hat{\eta} + e_j \forall j \in k_\eta$, similarly for ξ). Using the weights, the outer approximation of the latent variable is computed ($\hat{\eta} = f_\eta \sum_{j \in k_\eta} w_j\, x_j$, where f_η is used to standardize $\hat{\eta}$, similarly for ξ). The subsequent “inner approximation” determines “inner weights. These two steps are repeated until stable weights are reached.

In the second phase, PLS uses ordinary least squares regression (OLS) to estimate the regression relationships between the composite proxies and between the composite proxies and indicators.

The indicator weights constitute the only difference between PLS and regression with simple (unweighted) summed scales. Yet, there is still significant uncertainty in the PLS research community about the precise purpose and properties of the PLS weights (Rönkkö et al. 2016; Rönkkö, McIntosh, and Antonakis 2015). In the IS discipline, there is a commonly held belief that the weighting process is intended to produce better composite proxies by weighting more reliable indicators, i.e. those with less measurement error, more strongly than less reliable ones (Rönkkö and Ylitalo 2010). Indeed, in a recent *MIS Quarterly* editorial, Gefen et al. (2011) noted as much in stating that:

> *Optimization of these weights aims to maximize the explained variance of dependent variables [which] will also tend to minimize the presence of random measurement error in these latent variable proxies, especially as compared to using simple unit-weighted composites* (p. v).

However, as no set of perfectly reliable indicators exists, no weighting system can completely exclude measurement error (Henseler et al., 2014). We address to what extent PLS weights effectively reduce the effects of measurement error in Section 3.1.

2.3 Known Limitations

Including measurement error in the independent variables in a regression attenuates (biases, reduces) the estimate of regression coefficients, a well-known fact that Spearman (1904) first discussed more than 100 years ago. Hence, one should expect biased estimates when composites are used as proxies to estimate the regression coefficients between latent variables, as the measurement errors inherent in the imperfectly reliable manifest variables are summed into the composite proxies.

In fact, researchers have known PLS estimates to exhibit bias even before they began applying it in the IS discipline. Indeed, Herman Wold, who originally developed PLS, acknowledged this bias as early as 1982 (Hui & Wold, 1982). Specifically, the bias decreases and the estimates approach the true value as the sample size increases and the number of indicators for each latent variable increases. Comparing PLS and SEM, Jöreskog and Wold (1982 p. 266) wrote:

> *Under general conditions of regularity, ML [maximum likelihood] estimates of the unknowns are known to be optimal in the large-sample sense (asymptotic minimum of the confidence intervals), while the accuracy of the PLS estimates is less sharp—they are asymptotically correct in the joint sense of consistency (large number of cases) and consistency at large (large number of indicators for each latent variable).*

Wold (1982, p. 28) wrote: “Sacrificing optimality, PLS rests content with consistency, albeit in the qualified sense of consistency at large”. In a separate paper three years later (Wold, 1985, p. 213), he added:

> *The PLS estimates of parameters and of LV case values are inconsistent in as much as they do not tend to the true values when the number of observed cases (N) increases indefinitely. However, the PLS estimates are consistent at large in the sense that they tend to the true values when there is indefinite increase not only in the number of observed cases (N), but also in the number of indicators for each LV.*

Dijkstra (1983) proved these properties algebraically by stating that:

> *PLS will tend to underestimate the correlations between the latent variables; the discrepancy between true values and probability limits depends in a simple way on the quality of the proxies as measured by [the reliability of the composite]* $R^2(q_i, f_i)$. (p. 81).

In his seminal work on PLS, Lohmöller (1989, p. 207) demonstrated that “the bias factor depends primarily on the value of s (the higher the true loadings, the smaller the bias can be) and on K, i.e. the number of MVs”.

Even the earliest applications to IS research demonstrated this bias as Chin (1998, p. 330) noted: “Two points become clear from these formulas: the bias decreases as the loadings become more reliable, and the bias decreases as the number of indicators increases”. More than 10 years later, in an extensive and systematic simulation study, Reinartz et al. (2009) found that, with respect to the quality of the estimates:

> *ML-based CBSEM emerges as the more precise estimation method, as the mean parameter estimates are much closer to their theoretical values for CBSEM than for PLS (absolute difference 0.00–1.03% for CBSEM, 6.10–19.99% for PLS). Therefore, if consistency matters, ML-based CBSEM should be preferred over PLS.* (p. 338)

Given that the bias of the estimator was well known (one might say it was designed to be biased) and published since the early 1980s, one has to wonder why the IS discipline thought it a good idea to adopt a designed-to-be-biased estimator for the common factor model. Of course, we ask this question rhetorically and leave it as an exercise for the reader to ponder or perhaps research historians to investigate.

An early understanding of consistency at large and the role of large samples should also have pre-empted the debate around minimum sample size of PLS that began with a hypothetical statement that Chin (1998, p. 311) made:

> *If one were to use a regression heuristic of 10 cases per predictor, the sample size requirement would be 10 times either (a) or (b), whichever is the greater.... Under this condition, it may be possible to obtain stable estimates for the weights and loadings of each component independent of the final estimates for the structural model.*

In a subsequent chapter on the small sample argument, Chin and Newsted (1999) provided simulation evidence to support their claims about the small sample advantages of PLS. It took more than 10 years for an exhaustive simulation study (Goodhue et al., 2012) to emerge and conclude, as theoretically expected, that:

> *The belief among MIS researchers that PLS has special powers at small sample size or with non-normal distributions is strongly and widely held in the MIS research community. Our study, however, found no advantage of PLS over the other techniques for non-normal data or for small sample size (other than the universally stated concern that with smaller samples, LISREL may not converge).* (p. 998).

In the following year, we conducted a study (Rönkkö & Evermann, 2013) in which we noted how Chin and Newsted (1999) had misinterpreted their evidence, which Rönkkö (2014) further confirmed.

Today, researchers commonly agree that PLS often provides parameter estimates for very small samples, whereas other methods may not converge. For example, Henseler et al. (2014, p. 199) conclude that "PLS can be applied in many instances of small samples when other methods fail". But should it? The more important question concerns not whether it estimates parameters but whether one can trust the estimates. Rigdon (2016, p. 600) succinctly stated as much in noting: "Yes, PLS path modeling will produce parameter estimates even when sample size is very small, but reviewers and editors can be expected to question the value of those estimates, beyond simple data description". As theoretically expected, multiple studies have shown that alternative methods have less bias than PLS (Chumney, 2013; Goodhue et al., 2012; Reinartz et al., 2009). In fact, Henseler et al. (2016, p. 14) have unequivocally stated that "PLS does not need fewer observations than other techniques when it comes to inference statistics."

Given that even Wold noted the asymptotic consistency of the estimator for large samples in his original works, one has to wonder why the IS discipline thought that it was a good idea to use PLS, particularly for small sample sizes. Again, we leave this rhetorical question as an exercise for the reader (and future historians) to ponder.

2.4 Premature Adoption and Damage to the IS Discipline

It seemingly only took IS researchers 25 years, much debate, and many simulation studies to realize what Wold pointed out early on. How much damage has occurred in these 25 years? IS researchers were one of the earliest PLS adopters (Ringle et al., 2012), and many main IS theories build on empirical research that uses PLS. In Table 1, we list early, prominent IS theories whose foundational study builds on PLS analyses. Sample sizes in these studies ranged from 52 to 929, and they all relied on common factor models using t-tests for significance testing. These studies did not use consistent PLS (Section 4.1) as it had not yet been developed, did not use confidence intervals (Section 4.2) as the inappropriateness of the t-statistic was not known then, used the AVE and CR criteria for model fit assessment despite a lack of evidence supporting their use; they have since been shown to be completely inadequate for this task and better techniques are now available (Sections 4.4, 4.6). Of course, hindsight is always perfect, and those studies' authors acted on the best available knowledge and advice at the time. While these seminal studies have undergone rigorous peer review, our insufficient early understanding also limited that review process.

In retrospect, it becomes apparent that the IS discipline prematurely adopted PLS, which now casts significant doubts on seminal studies' validity. Of course, researchers have since expanded, extended, adopted, and/or replicated these theories, sometimes with other statistical methods, which may lend more credibility to them. But the fact remains that we, as a discipline, have staked our collective reputation on PLS and "bet the farm" with very little understanding. However, while this realization may cast serious doubt on those 25 years of IS research, it also finally provides the impetus to not only debate and argue back and forth about the merits of PLS but also to make significant progress in improving it and its use in applied research.

Table 1. Prominent Early IS Theories with Foundational Studies Based on PLS (sorted by citation count)

Theory	Citation	Approx. no. of citations
User acceptance	Venkatesh et al. (2003)	28000
Computer self-efficacy	Compeau & Higgins (1995)	7000
Technology acceptance	Venkatesh (2000)	6100
Cognitive absorption and IT use	Agarwal and Karahanna (2000)	5000
IT adoption	Karahanna et al. (1999)	4300
User satisfaction and technology acceptance	Wixom & Todd (2005)	3000
Interorganizational IT adoption	Teo et al. (2003)	1800
IT usage	Bhattacherjee & Premkumar (2004)	1800
Computer self-efficacy	Thatcher & Perrewé (2002)	800
IT-business linkage	Bassellier & Benbasat (2004)	700

3 Our 2013 Critique

In 2013, we presented a widely noted paper in which we critically assessed PLS (Rönkkö & Evermann, 2013) after which appeared an even more widely noted response by a group of PLS researchers (Henseler et al., 2014) and a rejoinder by independent methodology experts to take stock and provide additional comments and context (McIntosh et al., 2014). Our critique focused on six beliefs that we called "statistical and methodological myths and urban legends" in the spirit of Vandenberg (2006). At the time, researchers widely accepted these beliefs but had not substantiated them well; in some cases, it was difficult to find any evidence to support a belief and, in other cases, a belief had its roots in misinterpreted evidence. Specifically, we asked the following questions:

1) Does PLS have an advantage over traditional techniques as a SEM estimation method?
2) Does PLS reduce the effect of measurement error?
3) Can one use PLS to validate measurement models?
4) Can one use PLS to test null hypotheses about path coefficients?
5) Does PLS have minimal requirements on sample size?
6) Is PLS most appropriate for exploratory research?

We presented evidence and arguments to answer each question in the negative. In our discussion, we focused largely on two issues: 1) estimation bias due to including measurement error, and 2) capitalization on chance due to sample error correlations (Rönkkö, 2014). In this section, we do not revisit the arguments in their entirety but briefly present our main findings and the main arguments that emerged from the subsequent exchange with Henseler et al. (2014) and McIntosh et al. (2014).

In our critique, we noted that many studies referred to PLS as a SEM estimator and claimed that it would, therefore, have an advantage over other methods. For example, Gefen, Rigdon, and Straub (2011 pg. iv) state that "SEM has potential advantages over linear regression models that make SEM a priori the methods of choice in analyzing path diagrams when these involve latent variables with multiple indicators." Our critique has been misinterpreted as denying that PLS is a SEM technique. However, we make no such claim: What is or is not a SEM technique is a matter of definitions and semantics (Rönkkö et al., 2015). Instead, we claim that this labeling has misled and continues to mislead researchers to project other SEM methods' abilities onto PLS (Rönkkö et al., 2016). Rather than debating whether PLS is a SEM technique, we focus here on its capabilities.

3.1 Reliability

In our critique, we used a simple two factor model to show that PLS composites are less reliable than simple summed scales. This finding counters claims that the weighting mechanism favors more reliable indicators and, thus, that it improves the reliability and, consequently, addresses (at least partially) the attenuation of estimates due to measurement error. While the idea of weighting indicators by their reliability sounds reasonable, in most cases, the reliability advantages lack sufficient size to be relevant even if one knew the ideal weights (Rönkkö et al., 2015; Rönkkö et al., 2016c). Moreover, even if indicator weighting can increase reliability, techniques other than PLS, such as maximal reliability composites (Aguirre-Urreta et al., 2019), are specifically designed for this purpose and can be expected to perform better.

We noted that, in computing mode A weights, the error correlations between indicators of related composites play a substantial part in the weighting mechanism. In particular, in the absence of a true structural effect ($\beta = 0$), the error correlation is the only term that enters the weighting mechanism, which we noted as "problematic, because as a result of sampling variation, the errors are never exactly uncorrelated" (Rönkkö & Evermann, 2013, p. 434). For mode B weights, we noted that highly correlated indicators, as one typically finds in the common factor model, can lead to unstable regression results. We demonstrated these issues empirically by showing that mode B weights generally lead to far less reliable composites than mode A weights, which, in turn, are slightly worse than simple summed scales.

The PLS method propagates the indicator weighting problems to the estimation of structural path coefficients from the composites. We showed analytically and demonstrated empirically that error correlations form a part of the structural path estimate. Especially in the absence of a true structural effect ($\beta = 0$), this leads to parameter estimates that are biased bimodally away from zero. As error correlations due to sampling fluctuations are larger with smaller sample sizes, we termed this behavior capitalization on chance correlations. Rönkkö (2014) further expanded on this finding.

Our study was not the first to suggest this idea. In fact, Goodhue et al. (2007) first suspected capitalization on chance in PLS in the interaction analysis context. In particular, they concluded that PLS capitalizes on chance through its weighing of interaction indicators based on their correlation with the dependent variable Y, but they hid away this important insight in their online appendix. Rönkkö (2014) examined capitalization on chance in a more extensive simulation study and showed that the effect persists with larger models, larger and smaller sample sizes, and different indicator loadings.

In their response, Henseler et al. (2014) extended our own simulation studies to larger samples and larger structural effects (β). They found that the reliability of composites depends strongly on sample size, the size of the true parameter, and the loadings of the indicators. While summed scales also dominate mode B composites also in their study mode A composites appear to have a marginal advantage over simple summed scales for large samples ($n = 500$), strong effects ($\beta = 0.5$), or higher loadings (or some combination of these). One would expect the sample size effect as chance error correlations diminish with larger sample size. The higher loading of the most reliable indicator in Henseler et al.'s (2014) extension consequently led to a more reliable resulting composite due to its higher weighting. Nevertheless, in these ideal scenarios, the techniques differed only to a small degree (i.e., in the third decimal) (McIntosh et al., 2014).

In line with Henseler et al.'s (2014) conclusions about the requirement of a "broader nomological net" to reduce the effects of capitalization on chance (p. 190), Rigdon (2016, p. 602) points out that we specified a model "that violated the known conditions under which the PLS path modeling estimation algorithm works. This algorithm requires that every composite proxy must be correlated with at least one other composite" and that our "simulation showed what happens when you 'break' a statistical method, asking it to work outside of its boundary conditions". Given the iterative weight computation mechanism that we describe in Section 2, we see this point as a valid one. However, we see it as unfortunate that PLS "breaks" in precisely that condition in which researchers typically use it (i.e., when they do not know if an effect exists or its size may be small). Indeed, in their review, Goodhue et al. (2015) concluded that such models appear commonly in IS research and, thus, that this issue may affect much PLS-based IS research.

3.2 Model Testing

In our 2013 critique, we also highlighted an issue with model testing. We investigated a set of traditionally recommended model quality metrics such as composite reliability (CR), average variance extracted (AVE), goodness of fit (GoF) and relative GoF, and the standardized root mean squared residual (SRMR) to identify how well they performed in identifying misspecified models. In line with earlier work that Evermann and Tate (2010) conducted, we found that none of these metrics can reliably identify misspecified models and accept correctly specified models. As McIntosh et al. (2014) explained, "Measures that reflect the magnitudes of model parameters, such as the CR, AVE, GoF, and relative GoF, are incapable of reflecting how well the model reproduces the data" (p. 266) and "all four of these measures are determined by the magnitudes of the parameter estimates for a model and are insensitive to how well the model fits the sample data (p. 225).

McIntosh et al. (2014)further cautioned researchers against "interpret[ing] measures based on parameter magnitude as indicating model fit, which is the province of the chi-square test and SRMR" (p. 226). Indeed, in their extended simulations, Henseler et al. (2014) concurred with our results on traditional model quality metrics: "R&E's critique on the efficacy of conventional measurement model assessment criteria...is justified" (p. 195) Additionally, they demonstrated that SRMR and what they termed the "test of exact fit" can indeed distinguish between misspecified and correctly specified models (which we discuss in more detail in Section 4.4). Ironically, IS researchers tend to favor the CR and AVE statistics and have largely ignored these tests despite their demonstrated incapability to differentiate good models from bad ones.

3.3 Parameter Testing

As we note above, we found that, with no or very weak effects, estimates of structural parameters follow a distinctly non-normal distribution, which does not satisfy a key requirement for using the t-test for significance testing. We concluded that the t statistic and the t-test will be biased and that researchers should not use it. We also noted potential issues with bootstrapping the parameter estimates by pointing out that the bootstrapped estimates do not seem to follow the original distribution and, thus, violate a key bootstrapping assumption. Admittedly, we did not investigate empirically whether any statistics calculated from bootstrap replications might be robust to these violations.

In their extended simulations, Henseler et al. (2014) showed that normal or bias-corrected and accelerated (BCa)—but not basic—bootstrapping provides acceptable inference. They also showed it to be reasonably robust even when there is no true effect. Bootstrap performance improves for larger sample sizes and a more extended latent variable model. They stated that, "with regard to type II errors, the BCa confidence intervals turn out to show the lowest power. However, BCa confidence intervals are the only ones that under all conditions maintained the alpha protection level of 5%" (Henseler et al., 2014, p. 198). McIntosh et al. (2014) agreed that "the routine use of the t distribution to test parameters is questionable" (p. 229) and cautioned that "significance tests are meaningful only if the estimates can be trusted to be consistent" (p.. 229). Of course, we can question the latter, especially in situations with no effect ($\beta = 0$). Indeed, in a recent simulation, Aguirre-Urreta and Rönkkö (2018) demonstrated that, while bootstrap-based confidence intervals performed well when testing an existing path, they could not always trust them in the no-effect scenarios.

3.4 Exploratory Research

In our 2013 critique, we noted that, given its shortcomings, PLS might not be appropriate for what we understood as exploratory modeling. We argued that, in situations with tentative theory, identifying model fit and testing parameter significance become important for further model improvements. We also lamented the fact that PLS provides few diagnostics tools that could indicate to researchers specific changes for improving their models.

In response, Henseler et al. (2014, p. 200) argued that "a blind faith in covariance-based SEM's exploratory capabilities appear both unwise and unjustified". We concur with this assertion in that one should never apply blind faith toward any scientific method or theory. As McIntosh et al. (2014, p. 233) stated, "conventional SEM modification indices (MIs) show suboptimal performance". However, they also noted that:

> *There are numerous automated search algorithms that can determine the optimal number of latent variables and system of relations between them.... These techniques will almost invariably return several models that provide a good fit to the data.* (p. 234)

Importantly, they noted that "subject matter expertise can then be used to help select the most plausible of the discovered models, which can be subjected to cross-validation using independent data". (p. 234). In fact, we should replace the word "can" in the prior quote with the word "must".

We concluded our critique by suggesting that "PLS is decidedly not a "silver bullet", and it is very difficult to justify its use for theory testing" (p. 443). At the same time, we agree with Henseler et al.'s (2014, p. 202) conclusions that:

> *There are many more important questions about the behaviour of PLS that deserve scholarly attention, such as these: 'How well can PLS predict?' and 'How well does PLS perform if the composite factor model is indeed correct?'*

With respect to that last point, McIntosh et al. (2014, p. 235) suggested that "much of the controversy surrounding the viability of PLS-PM as a statistical method can be attributed to its original development and ongoing application as a technique that attempts to imitate factor-based SEM". Henseler et al. (2014) also emphasized this latter point in their response, and it has led to important recent developments that we discuss in Section 4.

4 Recent Developments in PLS

Our 2013 critique spurred interesting and important developments of the PLS method and its use. Henseler et al. (2014) already foreshadowed some developments in their response or researchers had already begun working on them, but they have seen further advancement since then.

4.1.1 Consistent PLS

Dijkstra and Henseler (2015b, 2015a) developed consistent PLS (PLSc) to correct for the bias of structural regression parameter estimates due to measurement error in the composites when used to estimate the common factor model. Based on PLS indicator weights for mode A weighting and reflective indicators, they developed a consistent way to estimate a composite variable proxy's composite reliability:

$$\rho_A = (\hat{w}^T \hat{w})^2 \frac{\hat{w}^T (S - \operatorname{diag}(S))\hat{w}}{\hat{w}^T (\hat{w}\hat{w}^T - \operatorname{diag}(\hat{w}\hat{w}^T))\hat{w}} \tag{1}$$

Here, $\hat{w}$ is the weight vector for a latent variable proxy and S is the sample covariance matrix for that proxy's indicators.

In the first step, one conducts the traditional PLS algorithm, which computes the scores for the latent variable proxies ($\hat{\xi}$) and the weight vectors ($\hat{w}$) for each latent variable proxy. The biased estimate of the latent variable correlation between latent variables i and j is then:

$$r_{ij}^* = \operatorname{cor}(\hat{\xi}_i, \hat{\xi}_j) \tag{2}$$

Next, one computes the reliability ρ_A for each latent variable proxy. One then estimates the corrected latent variable correlations using the well-known correction for attenuation that Spearman (1904) introduced:

$$r_{ij} = \frac{r_{ij}^*}{\sqrt{\rho_A(\hat{\xi}_i)\rho_A(\hat{\xi}_j)}} \tag{3}$$

Finally, one can derive consistent estimates of the loadings using ρ_A and $\hat{w}$.

In their initial simulation studies, Dijkstra and Henseler (2015b) showed that PLSc provides estimates close to what the maximum likelihood estimator provides with little bias and comparable precision for the structural parameters, but less precision for the item loadings. In a later comparison using a real data set, Henseler (2017) confirmed that PLSc provided results close to unweighted least squares (ULS)-based covariance estimation.

However, not all studies have found similarly encouraging effects. Both Huang (2013) and Rönkkö et al. (2016b) found less precise and more biased PLSc loading estimates compared to traditional ML estimates. They showed that PLSc tends to overestimate small correlations and underestimate large correlations. The positive bias may be due to capitalization on chance (Rönkkö et al., 2016b) and the negative bias due to overestimating reliability (Aguirre-Urreta et al., 2019). PLSc also produces more non-convergent or inadmissible results. Most recently, Yuan et al. (2020, p. 334) argued that PLSc provides no clear advantage over disattenuating unweighted scales with coefficient (Cronbach's) alpha. Finally, Dijkstra and Henseler (2015a, p. 309) noted that "among the consistent techniques, PLSc typically had the lowest statistical power". Other studies have drawn the same conclusion.

Dijkstra and Schermelleh-Engel (2014) and Becker et al. (2018) have also examined PLSc to identify its performance in estimating models with interaction terms. Using a simulation study with sample size

$n = 500$, Becker et al. (2018) investigated different methods for modeling and estimating the interaction term and also compared PLS to PLSc. They found evidence for PLSc's bias correction also in the interaction terms. We would expect as much since PLSc estimates the interaction term using an OLS regression based on the product of the bias-corrected main effects composite scores (Henseler & Fassott, 2010). Their findings extend to both common factor models and to mixed formative-reflective models.

In further extending PLSc to address the bias from ignoring correlated measurement errors, Rademaker et al. (2019) demonstrated that their extension showed a small improvement over the original PLSc method, although they concluded that "original PLSc is comparatively robust with respect to misspecification" (p. 459). Rademaker et al. (2019) suggested that handling correlated measurement errors may yield better results particularly with the product-indicator approach to interaction modelling in PLS (Henseler & Fassott, 2010). However, noting Becker et al.'s (2018) results, the PLSc-corrected two-step method might be preferable. On the other hand, Becker et al. (2018) only examined PLSc without Rademaker et al.'s (2019) extension. Hence, identifying the best way to estimate interaction models in PLS remains an open issue.

While PLSc represents a significant improvement over traditional PLS when applied to the common factor model, further simulation research is required to clearly identify the limits of its applicability and its robustness to various different conditions. In particular, given Rönkkö et al.'s (2016b) and Yuan et al.'s (2020) recent findings, its advantage over the simpler approach of using indicator means and disattenuating the correlations with coefficient alpha or CR (Devlieger & Rosseel, 2017; Devlieger et al., 2015; Rosseel, 2020) requires further investigation. As a result, we recommend (R):

R1: If researchers use PLS for factor models (reflective), they should use consistent PLS with measurement error correction.

4.2 T-tests are Inappropriate

In our earlier critique (Rönkkö & Evermann, 2013), we already showed that t-tests for null-hypothesis significance testing of the parameter estimates are inappropriate in PLS and suggested that researchers should use bootstrap confidence intervals instead. Aguirre-Urreta and Rönkkö (2018) have examined different bootstrapping methods for PLSc. They found strong evidence that one should use percentile, bias-corrected (BC), and bias-corrected and accelerated (BCa) bootstrapping over normal distribution-based bootstrapping. We do not find this result surprising as Rönkkö and Evermann (2013) noted that the underlying parameter estimates do not follow a normal distribution. While Aguirre-Urreta and Rönkkö (2018) found BC and BCa bootstrapped confidence intervals appear more efficient in that they converge faster to their theoretical values with increasing sample size, they are optimistic in that they do not reach the nominal coverage level, generally yielding an interval that is too narrow. On the other hand, Aguirre-Urreta and Rönkkö (2018) found percentile intervals to be less efficient but more conservative and, thus, the preferred option. Aguirre-Urreta and Rönkkö (2018, p. 26) concluded that "the proposed approach [percentile bootstrap] is not only a valid alternative for statistical inference with PLSc, but also a substantial improvement over current practice". Their results imply that researchers should use conservative *percentile* bootstrapping of PLSc estimates rather than BC/BCa-based bootstrapping; Henseler et al. (2014) recommended the latter, and many PLS software tools use it as the default.

R2: If researchers use PLS, they should avoid t-tests and instead use percentile bootstrapping for statistical inference.

Aguirre-Urreta and Rönkkö (2018) also showed that all bootstrap methods struggle when the latent variables correlate weakly or do not correlate at all, especially for models with a small number of sparsely connected latent variables. Thus, bootstrap methods do not entirely address weakly connected latent variables, which leads to a bimodal parameter estimate distribution and an inappropriate t-test. Because parameter significance tests are particularly important when researchers are uncertain about the existence of a relationship, PLS, even with bootstrapping, does not always perform well in critical situations. Aguirre-Urreta and Rönkkö (2018) also recommended diagnostics such as inspecting the histogram of bootstrap replications, which most contemporary PLS software programs provide, for a severely non-normal shape. As a remedy for potential problems, they suggest using equal weights to calculate the problematic composites.

R3: If researchers use bootstrapping for significance testing, they should examine the bootstrap estimates for a severely non-normal distribution.

4.3 Power Estimation Using Simulation

Aguirre-Urreta and Rönkkö (2015) recently discussed how to select an appropriate sample size for a PLS study—a particularly important aspect of using PLS since it affects the estimator's consistency, its convergence behavior, and the confidence intervals for parameter significance testing. Examining the latter, Aguirre-Urreta and Rönkkö (2015) developed a guide to using simulation studies for power calculations.

In general, one should not use traditional rules of thumb such as ten times the largest number of indicators on a construct (Chin, 1998) or using regression based power calculations (e.g., the popular GPower software). Aguirre-Urreta and Rönkkö (2015) discussed the questionable origin of the "ten times" rule at length and noted that regression-based power computations assume perfectly reliable regressors and regressands. As such, researchers should not use them with multiple-indicator models that one assumes to contain measurement error.

Aguirre-Urreta and Rönkkö (2015) provided a step-by-step guide using a simple script for the R statistical system. However, the method as reported still uses uncorrected mode A weights by default and the t-test for significance testing in the power calculations, which one should avoid and replace with percentile confidence intervals as we note in Section 4.2. Recent versions of the matrixpls package for the R system, on which Aguirre-Urreta and Rönkkö (2015) built their guide, allow one to choose percentile confidence intervals (using the citype = "perc" parameter for the matrixpls.sim function) and use PLSc (using the disattenuate = TRUE and parametersReflective = estimator.plscLoadings parameters for the matrixpls.sim function). The precise way to use matrixpls has changed since Aguirre-Urreta and Rönkkö (2015) appeared, but Rönkkö's (2020) user manual contains an update to their example.

Given that researchers can easily perform simulation studies for PLS with contemporary PLS software tools, they should explicitly hypothesize about likely effect sizes and data distribution assumptions, set them up for a simulation study, and identify the statistical power for various sample sizes *before* collecting data.

R4: If researchers use PLS, they should use a simulation study to determine an appropriate sample size and study power a priori.

4.4 "Goodness of Fit" does not Measure Goodness of Fit

Traditionally, model evaluation in PLS rested on model "quality criteria" such as the composite reliability (CR) and the average variance extracted (AVE). However, these criteria do not test whether the model fits the data but evaluate the factor structure and the relationships between observed and latent variables. Even the goodness-of-fit (GoF) metrics developed by Tenenhaus, Amato, and Esposito Vinzi (2004) are misnamed, as they also examine the factor structure of the model, rather than the fit of the model with the observed data. These metrics rely on the parameter estimates but the estimates will be biased if one misspecifies the model is. Needless to say, simulation studies confirmed that these indices cannot detect misspecified (i.e., non-fitting) models (Evermann & Tate, 2010; Rönkkö & Evermann, 2013; Henseler & Sarstedt, 2013; Henseler et al., 2014).

Researchers can demonstrate model testing adequacy using empirical datasets. An adequate model test or model quality heuristic should reject obviously incorrect models. While one can never know if a model is correct in an empirical application, one can easily come up with surely misspecified models (e.g., by combining two factors into one, assigning indicators to obviously incorrect factors, or adding more latent variables to the model). If the model quality indices that one uses do not reject a surely incorrect model, it would provide strong evidence against these indices. Rönkkö et al. (2012) demonstrated such an analysis in evaluating various incorrectly specified models based on various models published in leading IS journals. They did not make an encouraging conclusion: "Based on the analysis of the goodness of fit indices we can conclude that...any model fits approximately as well as the original model" (p. 9).

R5: If researchers use PLS-based indices for model testing or fit evaluation, they should test or evaluate also models that they assume to be incorrect for the data to ensure that the model quality indices would actually lead them to reject incorrect models.

Early in PLS's development, Lohmöller (1989) proposed fit indices and reliability indices based in part on the residual covariances and on the maximum-likelihood χ^2 fit function. However, the early PLS literature never adopted them. Only later did Henseler et al. (2014) and Dijkstra and Henseler (2015b) develop these ideas further in proposing three ways to evaluate overall model fit by comparing to the observed covariance matrix, a technique adopted from covariance-based SEM. They defined three distance metrics as follows:

$$SRMR = \sqrt{\frac{2}{p(p+1)}\sum_{i=1}^{p}\sum_{j=1}^{i}\frac{(s_{ij}-\hat{\sigma}_{ij})^2}{s_{ii}s_{jj}}} \tag{4}$$

$$d_{LS} = \frac{1}{2}trace(S-\hat{\Sigma})^2 \tag{5}$$

$$d_G = \frac{1}{2}\sum_{k=1}^{p}(\log(\varphi_k))^2 \tag{6}$$

Here, S refers to the sample (observed) covariance matrix, $\hat{\Sigma}$ to the estimated population covariance matrix, and ψ_k to the k-th eigenvalue of $S^{-1}\hat{\Sigma}$. One can do the test for exact fit using either the d_{LS} (least squares) or the d_G (geodesic) distance metric.

As McIntosh et al. (2014) have pointed out, standardized root mean squared residual (SRMR) is a fit index, not a statistical test, so the acceptable value is subjective and does not reflect pre-determined error rates. One interprets SRMR the same way as covariance-based SEM in that one can assume values < 0.08 indicate acceptable model fit (Hu & Bentler, 1999), although debate continues around the appropriateness and usefulness of fit indices in general and simple cutoff rules of thumb in particular (Kline, 2011).

Researchers use the metrics d_{LS} and d_G by bootstrapping them and comparing the observed sample statistic to the 95th percentile of the bootstrap. Initial simulation studies suggest that model rejection rates based on these metrics are too low for small sample sizes (n = 300, 600) and approach the theoretical values for larger samples (n = 1200) (Dijkstra & Henseler, 2015b). Note that researchers usually consider what Dijkstra and Henseler (2015b) characterize as “small samples” in their simulation as quite large samples in PLS applications.

R6: If researchers use PLS, they should establish adequate model fit using $\mathbf{d_G}$ prior to interpreting path estimates or assessing factorial structure.

Model selection involves identifying misspecified models and assessing model fit. One uses it to identify the best model among a set of equally well-fitting competing models. One should never select ill-fitting over well-fitting models. In covariance-based SEM, the χ^2 difference provides a strong and well understood statistical test for model selection. With non-nested models, researchers often rely on information theoretical metrics, such as the Akaike information criterion (AIC), the Bayesian information criterion (BIC) or the Geweke-Meese criterion (GM), all of which researchers have recently also applied to PLS estimated models (Sharma et al., 2019a, 2019b; Danks et al., 2020). These criteria penalize models for complexity to achieve a balance between fit and parsimony. When applied to PLS:

> *Closed-form formulas for the different model selection criteria do not exist. However, when the error distribution is normal with constant variance, the maximum likelihood-based formulas can be written as a function of the model residuals, and specifically the sum of squared residuals.* (Sharma et al., 2019b, p. 351).

The assumption of normally distributed errors is a significant departure of the usual characterization of PLS as a distribution-agnostic or assumption-free method. In a simulation study that they conducted, Sharma et al. (2019b) found that AIC, BIC, and GM can successfully identify the correct model significantly better than previous ”quality criteria” such as GoF, R^2 and Q^2 metrics. These findings confirm earlier results that we note above: “Overall, none of the PLS criteria performed satisfactorily” (p. 356). Sharma et al. (2019b) found that BIC and GM provide the best success rates in identifying the correct model. Their findings also held for relatively small samples (i.e., 100) and remained robust across sample sizes with performance improving with an increasing sample size. Similarly, their findings also held for small effect sizes and performance improved with an increasing effect size. They recommended that BIC or GW should replace the traditional model “quality criteria”. While Sharma et al. (2019b) used the lowest information criterion value to select a model, Danks et al. (2020) used AIC, BIC, and GW weights as selection criterion. As these weights are essentially logit transformations, they unsurprisingly arrive at similar conclusions: “weights derived from BIC and GM are well suited for separating incorrectly specified from correctly specified models” (p. 13).

Model selection is comparative in nature while model misfit identification relies on a single model and the observed data. Thus, while Sharma et al. (2019b) and Danks et al. (2020) have used selection methods to identify and reject misfitting models, their set of models included one or more well-fitting ones. In general, comparative model selection cannot substitute for identifying model misfit.

> **R7:** If researchers use PLS to compare multiple, well-fitting models, they may select models using information-theoretic criteria. They must establish and demonstrate good model fit before any comparison.

4.5 R^2 does not Measure Predictive Power

Researchers have frequently argued that PLS has particular usefulness in predicting rather than explaining, though few researchers appear to have followed through with this idea (Evermann & Tate, 2016; Rönkkö et al., 2016c). Shmueli et al. (2016) pointed out two key differences between explanation and prediction. At the data level, prediction deals with predicting *new* values for specific observations (”out-of-sample prediction”), whereas explanation deals with analyzing patterns in existing values. At the application level, prediction deals with making *individual-level* decisions, while explanation deals with average- or population-level decisions. Metrics such as R^2 or Q^2 do not measure predictive performance: R^2 is a population-level, in-sample metric rather than an out-of-sample metric, while Q^2 is also an aggregate rather than an individual-level metric.

Shmueli et al. (2016) presented a method to produce out-of-sample, individual-level predictions from PLS. Adopting the “operative prediction” concept from Lohmöller (1989), Shmueli et al. (2016) described prediction from manifest variables to manifest variables that can “fully incorporate every aspect of the theoretical model in gauging predictive validity. Furthermore, operative prediction only requires manifest data items as predictors and outcomes” (p. 4554).

Evermann and Tate (2016) compared the predictive performance of PLS, covariance-based estimation, and a theoretic prediction methods (EM and linear regression that operated only on observed variables with no structural model or mediating latent variables). Purely reflective models, such as the one in Figure 1, have no exogenous manifest predictor variables, while purely formative models have no endogenous manifest predictand variables. Consequently, in their first study on purely reflective models, they necessarily used the Q^2 statistics with blindfolding to evaluate prediction performance. They also examined PLSc but found that traditional PLS with mode A performed better than PLSc. In their second study, they examined mixed formative-reflective models of the type that Figure 2 shows. These models contain observed exogenous predictor variables and observed endogenous predicted variables. Evermann and Tate (2016) used root mean squared error (RMSE) with cross-validation to measure out-of-sample, individual-level prediction (“operative prediction” Lohmöller (1989)). In both cases, prediction from PLS estimated models has an advantage over prediction from covariance-estimated models.

In recent work, Liengaard et al. (2020) built on the cross-validated out-of-sample prediction notion that Shmueli et al. (2016) and Evermann and Tate (2016) used to suggest a prediction-oriented model selection criterion. Their CVPAT method defines summary statistics of prediction-error differences between alternative models and uses a t-test to identify significant predictive differences between models. In an initial simulation study, they found promising results in terms of error rates and test power.

In summary, we do not mean to argue for or against PLS for predictive studies over any other technique (and many others exist, including atheoretical ones with their roots in machine learning such as support vector machines and deep neural networks) but rather focus on reminding researchers that simply stating that a study is predictive in nature does not in fact make it predictive. Evermann and Tate (2016, p. 4581) suggest that, for predictive studies:

> *The theoretical motivation, parameter significance testing, and the assessment of validity and reliability of the measures should take a limited and sub-ordinate role to the presentation of appropriate blindfolding or cross-validation procedures and metrics.*

Simply put, most IS researchers do not study research questions where predictive modeling would be applicable but focus on testing theory that requires explanatory models. If predictive modeling pertains to their study, researchers should clearly explain the predictions' purpose and usefulness and use operative prediction (i.e., mixed formative-reflective) to assess their models' out-of-sample prediction at the individual level as Shmueli et al. (2016) and Evermann and Tate (2016) recommend.

> **R8:** If researchers use PLS for prediction, they should adequately justify and reflect this motivation and assess out-of-sample prediction at the individual level.

4.6 AVE and CR do not assess Discriminant Validity

Rönkkö and Evermann (2013) noted that the typically used criteria AVE and CR cannot identify model misspecification. Their argument focused on cross-loadings, which affect discriminant validity metrics. Henseler et al. (2014) confirmed their results, which served as the impetus behind efforts to develop a new technique for assessing discriminant validity. Taking the multitrait-multimethod matrix (MTMM) as a starting point, Henseler et al. (2015) derived the hetero-trait-mono-trait (HTMT) criterion for assessing discriminant validity. They proposed two hard cutoff criteria but also recommended bootstrapping to test whether the HTMT is less than one. Henseler et al. (2015) showed that the HTMT relates to coefficient alpha, and, in a recent, study, Rönkkö and Cho (2020) proved that HTMT actually constitutes a correlation between unweighted composites that is disattenuated with standardized alpha.

In a simulation study, Henseler et al. (2015) replicated our poor results for the AVE and CR criteria and showed that the HTMT method can detect low discriminant validity. However, in a more comprehensive simulation study, Rönkkö and Cho (2020) found that these results reflect the AVE criterion's problems more than any HTMT capability. They further showed that the HTMT method makes more assumptions than traditional confirmatory factor analysis. It provides little advantage when these assumptions hold and perform less well in other conditions (Rönkkö & Cho, 2020).

Because HTMT is a disattenuated correlation, using it with PLSc presents two challenges. First, PLSc calculates composites with PLS weights, whereas HTMT is based on unweighted composites. Second, disattenuation in HTMT is based on standardized alpha rather than the reliability ρ_A developed for PLSc. Two different sets of composites and disattenuation correlations will likely confuse researchers and readers alike; for most researchers, simply estimating a traditional factor analysis might be an easier and more effective alternative.

> **R9:** Researchers should not use AVE- and CR-based criteria to establish discriminant validity. They can use the HTMT if they can justify its assumptions. Otherwise, they should use traditional factor analysis techniques.

4.7 PLS is intended for Composite Models

Given the increasingly critical PLS evaluations (Chin, 1998; Chumney, 2013; Goodhue et al., 2012; Reinartz et al., 2009), PLS proponents could no longer simply ignore or sweep away the issue of biased PLS estimates for the common factor model. Prior to PLSc's development, Rigdon (2012) argued the need to refocus PLS on composite models. He recommended that researchers should "recognize and evaluate PLS path modeling as a method strictly in terms of weighted composites...and build the future of PLS path modeling in entirely factor-free terms" (p. 342). Rigdon (2016) suggested that researchers misunderstand and misapply PLS to estimate the common factor model. Sarstedt et al. (2016) have made a similar point and argued for a focus on composite models. Rigdon (2016) and Sarstedt et al. (2016) both argue that PLS should not be used to estimate models for which the data generating process was the common factor model and that evaluating the performance of PLS on such models is therefore misguided. While this argument may be true, applied IS researchers still use PLS most frequently to estimate the common factor model (Ringle et al., 2012) and the common factor models underpins prominent IS theories (see Table 1).

Rigdon's (2012) paper attracted various responses, many of which rejected his proposition to turn away from the factor model. Dijkstra (2014) responded with an argument for consistent PLS and briefly sketched its development, one of the earliest publications on PLSc. Bentler and Huang (2014) presented PLSe1 and PLSe2. While misleadingly termed "PLS", the estimation is a covariance-based estimation rather than a composite approximation mechanism such as PLS. PLSe1 uses the PLS estimates as starting values in a conventional CBSEM estimation, -PLSe2 uses PLS estimates to build a weight matrix for GLS estimation. While it remains unclear what advantage parameterizing the GLS estimator in such a way would provide, recent literature has nevertheless promoted these techniques (Ghasemy et al., 2020).

To further complicate matters, in their response to our 2013 critique, Henseler et al. (2014) introduced what they termed a "composite factor model". After McIntosh et al. (2014) pointed out that the model constituted neither a composite nor a statistically identified model, Schuberth et al. (2018) later refined it. They showed that one can rewrite the composite model in a mathematical form similar to a common factor model under the assumption that the loadings are proportional to the weights. This proportionality constraint statistically identifies the model and Schuberth et al. (2018) proposed a model fit test using the SRMR, d_{LS} and d_G metrics (see Section 3.2). In their simulation study, they found that d_G in particular can correctly identify composite model misspecifications with an error rate restricted to the theoretical expectation at the 0.05 and 0.1 significance levels. They also noted that one requires sample sizes that exceed 350 to consistently detect misspecifications. However, Schuberth et al. (2018) did not estimate their model using PLS but used generalized canonical correlation analysis (GCCA). Whether PLS weights work as well remains to be demonstrated.

The literature contains two arguments used to support composite models. First, a misspecified latent variable model can produce severely biased estimates, and, though biased, estimates from composite models are less bad (Rhemtulla et al., 2020). While true, as a better course of action, one should avoid misspecifications in the first place by engaging in proper model diagnostics. Second, PLS represents an ideal technique when the population follows a composite model rather than a common factor model. Unfortunately, the literature has yet to provide a compelling argument for why and when one should expect a phenomenon of interest to follow a composite model. While we need to examine this argument's merits, we leave it to another paper to do so and refrain from making a recommendation on using PLS at this time. Nevertheless, adopting such a perspective would require one to develop an entirely different measurement theory and abandoning Cronbach's alpha, CR, and AVE as reliability measures.

4.8 PLS and Dependent Formative Models

Aguirre-Urreta and Marakas (2014) recently examined the way researchers use PLS to estimate models with a dependent formative construct. They showed algebraically and in simulation that the model that some PLS users might expect from a graphical representation such as that in Figure 4 differs from what PLS actually estimates. They argued that the statistical model that one should infer from the path diagram in Figure 4 based on SEM path diagram rules is:

$$y_i = \lambda_i \xi + \epsilon_i$$
$$\eta_1 = \sum_i w_i x_i + \beta \xi + \zeta \quad ,$$

where β is the structural parameter of interest. Aguirre-Urreta and Marakas (2014) found that PLS estimates of β reflect only the indicator covariances between the y_i and the x_i even when the y_i correlate with the z_i. In the case of uncorrelated y_i, x_i, the estimates for β are zero. They noted that, in earlier work (Aguirre-Urreta & Marakas, 2008), they mistakenly reported this phenomenon as estimation bias.

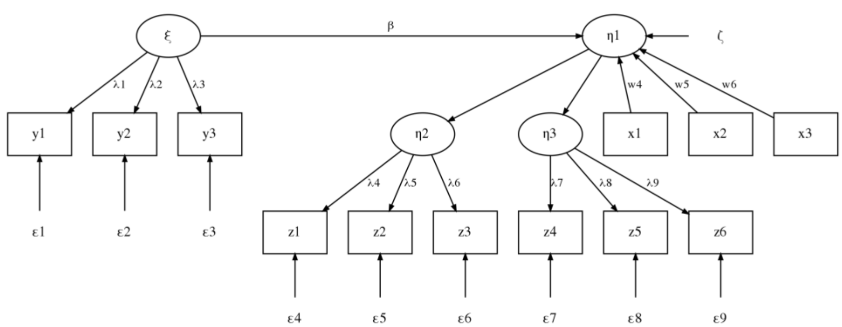

Figure 4. Formative Dependent Latent Variable (η_1) (We Omit Correlations among the x Indicators for Clarity)

In response, Rigdon et al. (2014) noted that the graphical notation may be misleading. They pointed out that PLS forms the composites prior to and separately from the structural model estimation and that the statistical model to be read from Figure 4 is:

$$\eta_1 = \sum_i w_i x_i \qquad \text{composite}$$
$$\eta_1 = \beta\xi + \zeta \qquad \text{dependent}$$
$$\Rightarrow \sum_i w_i x_i = \beta\xi + \zeta$$

This model shows that the β parameter depends on the covariance between the x_i components and the reflective independent latent variable ξ. Hence, given good indicator reliability, the PLS estimate of β as a function of the covariances between x_i and y_i is precisely the intended one. Rather than showing a failure in PLS, Aguirre-Urreta and Marakas (2014) showed that the model that most researchers assume for specifying formative or composite endogenous variables is not in fact correct.

While Aguirre-Urreta and Marakas (2014) suggested that the observed covariances between the y_i and x_i that form the effect could be due to omitted common causes, they are more likely due to precisely the intended nature of the formative construct: ξ (and, thus, its highly correlated indicators) has an effect on the components x_i of η_1. Indeed, Cadogan and Lee (2013) already made this point in convincingly arguing that components must necessarily mediate effects on composites and, thus, essentially endorsed Rigdon et al.'s (2014) variable-level specification).

To avoid the confusion that this debate showcases, researchers must clearly state their assumptions and the precise statistical model they intend to estimate. While the graphical user interfaces that PLS software provides make it very easy to perform a PLS analysis, this simplicity comes with the downside that one does not readily know the actual equations that form the statistical model. Moreover, commonly used PLS software provides very little documentation (SmartPLS, 2020) on how they translate path diagrams into equations, which causes considerable ambiguity on what exactly researchers estimated when publications present the results only in a graphical format.

R10: If researchers use PLS with dependent formative constructs, they should verify and publish the assumed and estimated statistical model (i.e., the equations) to avoid any confusion.

Given that formative indicators mediate the effects from a latent variable to a formative latent variable, estimating only the total effect rather than the individual mediator effects represents a wasted research opportunity. As a result, one might want to use the latent variable specification (Cadogan & Lee, 2013; Rigdon et al., 2014). One could also obtain comparable results after PLS estimation by exporting the composite scores for ξ from the analysis and regressing each indicator x_i on these scores.

4.9 Endogeneity and 2SLS

Researchers have traditionally implemented PLS as a two-phase method with ordinary least squares (OLS) regression in its second phase. OLS has two important restrictions. First, it requires the regressors to be independent of the regressands' error terms in order for the parameter estimates to be unbiased. Second, it requires a recursive system of regression equations (i.e., the structural part of the PLS path model must be acyclical).

Researchers often violate the first restriction in practice, which leads to "endogeneity" (Antonakis et al., 2010). Omitted variables (i.e., variables that affect both the regressor and regressand) are typical causes of endogeneity. When one omits them from the estimated model, the regressor becomes correlated with the error term (Antonakis et al., 2010). The second restriction prevents researchers from estimating an entire class of models that incorporate feedback cycles between constructs or composites.

However, researchers use OLS in the second PLS phase for convenience and habit, not necessity. Dijkstra and Henseler (2015a) used two-stage-least-squares (2SLS) regression to estimate the structural parameters from the bias-adjusted correlation or covariance matrix in PLSc "whether recursively or with feedback patterns" (p. 14). As such, they relaxed the second restriction in PLS and allow the modeler to describe cyclical structural models. One can also use 2SLS to address endogeneity. Benitez-Amado et al. (2016) and Hult et al. (2018) have recommended 2SLS in order to use the instrumental variable approach for addressing endogeneity provided researchers have suitable instrumental variables ("instruments"). An instrument refers to a variable that strongly correlates with the problematic endogenous regressor but is independent of the error term in the regression model.

Hult et al.'s (2018) approach has two limitations. First, the authors recommended the usual Hausman test for identifying endogeneity and the Sargan test for identifying the instrument's suitability. Both tests are parametric tests that make distributional assumptions, which runs counter to the typical goal to keep PLS a distribution-free method. Moreover, these techniques do not consider the unknown effect of the sampling distribution of PLS. Hult et al. (2018) recommended that future research should investigate using bootstrapping. A second issue, which Hult et al. (2018) did not address, concerns the first stage in PLS and, specifically, that it computes composite scores in a way that relies on OLS and may be subject to endogeneity. Researchers need to conduct more work to examine whether and to what extent the way PLS computers composite scores has any impact on composites and, hence, parameter estimates.

R11: If researchers use PLS and suspect endogeneity, they should address it using 2SLS.

R12: If researchers use PLS and a non-recursive structural model, they should use 2SLS rather than OLS.

5 Alternatives to PLS

When first developed, Herman Wold positioned PLS as an alternative to Karl Jöreskog's nascent covariance-based analysis (Wold, 1982). Given the computing abilities at the time, Wold's PLS worked faster, which allowed one to interactively explore data. However, given its estimation bias and consistency-at-large property, PLS clearly could not be a valid alternative to covariance-based analysis for inferential, theory-driven research work with the common factor model. Yet, perhaps based on the historical context, researchers often view covariance-based SEM as the primary alternative to PLS (e.g., Reinartz et al., 2009; Goodhue et al., 2012; Sarstedt et al., 2016; Rigdon et al., 2017; Hair et al., 2019).

Given that most researchers now explicitly accept PLS's composite nature and position it primarily as a composite analysis tool, we can clearly see that covariance-based SEM does not in fact represent the primary alternative to PLS. Instead, alternatives to PLS are other composite methods, the simplest of which is unweighted summed scales. Hence, recent work has compared PLS primarily to summed scales (Rönkkö & Ylitalo, 2010; Rönkkö & Evermann, 2013; Rönkkö, 2014). Composites' reliability represents the key consideration in composite modeling (i.e., how well the estimated composite scores correlate with the true scores). While one can certainly construct scenarios where PLS's weighting algorithm outperforms unweighted summed scales by a significant amount, it by no means represents the typical case. More cases where summed scales provide composites as reliable or more so than PLS weighted composites exist. In many cases, PLS offers only a small advantage and, given the limitations and cautions around PLS that we note in the previous section, PLS likely represents an unacceptable trade-off. In situations with highly correlated indicators, one can gain little reliability by differential item weighting (Rönkkö et al., 2015; Rönkkö et al., 2016a). In fact, Rönkkö, Evermann, and Aguirre-Urreta (2016) found that indicator weighting provides little advantage even at a low mean indicator correlation of 0.4, well below the typical level of indicator correlations in IS research.

Does indicator weighting make a difference in practice? Examining published standardized indicator loadings for the TAM scales across 43 published studies, Evermann and Tate (2014) found only small differences in the mean loadings of different indicators. Re-analyzing an empirical study, Rönkkö, McIntosh, and Antonakis (2015) shows PLS composites that are nearly perfectly correlated with unweighted composites.. Extending this analysis, Rönkkö et al. (2016) re-analyzed another empirical dataset to calculate different PLS composites and also concluded that "PLS indicator weights do not generally provide a meaningful improvement in reliability" (p. 6). Rönkkö et al. (2021) analyzed two datasets that researchers have commonly used to demonstrate PLS and found that the correlations between PLS composites and corresponding unit-weighted composites exceeded 0.99 for all but one composite. A single very poorly performing item explained the one outlier correlation. After eliminating that item, as one would do in the factor analysis stage in scale development (Hair et al., 2010), the correlation between the corresponding PLS composite exceeded 0.99 in line with the other correlations in that study.

One can easily compare simple summed scale with PLS composites by exporting PLS composite scores and examining their correlation with the summed item scores. Rönkkö et al. (2021) formalized this comparison as the composite equivalence index (CEI) statistic. Researchers should adopt this analysis as a basic diagnostic for any PLS analysis: if they find no meaningful differences, they can hardly justify using PLS composites over simpler unweighted composites given the challenges associated with PLS (e.g., in terms of the potential for capitalization on chance and parameters testing (see Section 4.2)). If researchers do find differences, they should explain them because differing weights can indicate a problem such as a cross-loading (Rönkkö et al. 2021).

R13: Researchers should compare PLS composites to unweighted composites to demonstrate any possible advantage that the PLS composites might have.

Hwang and Takane (2004, 2014) developed another composite analysis method called generalized structure component analysis (GSCA) explicitly as an alternative to PLS. It has an explicit objective function (squares of the measurement errors and structural residuals) that GSCA optimizes to estimate model parameters. In contrast, PLS has no explicit optimizing criterion and researchers have not clearly and explicitly defined its frequently claimed weight "optimality" (Rönkkö et al., 2015; Rönkkö et al., 2016a). Recent advances in GSCA include regularization (Hwang, 2009), latent interactions (Hwang et al., 2010a), multilevel analysis (Hwang et al., 2007), and non-linear models (Hwang & Takane, 2009). A bias correction inspired by consistent PLS has been implemented as GSCA_M (Hwang et al., 2017). Curiously, GSCA has found little adoption in IS research even though various software implementations are available, such as an R package, an Excel add-on, and a Web-based graphical implementation.

We find the fact that IS researchers have scarcely adopted GSCA and their focus on PLS all the more puzzling as, in many situations, research has shown GSCA to outperform PLS. In their initial application, Hwang and Takane (2004) showed GSCA estimates to be very close to PLS estimates but with smaller standard errors. Examining the quality of composites, Tenenhaus (2008) compared GSCA, PLS, unweighted summed scales, and principal components and concluded that "all methods yield to comparable components" (p. 15) and "when the blocks are good, the computation of the components does not depend upon the method use." (p. 15) (i.e., for unidimensional and strongly correlated indicators). Here too the simple method of unweighted summed scales proved as good as complex weighting methods. Focusing on parameter recovery in misspecified models, Hwang, Malhotra et al. (2010b) also claimed that GSCA has an advantage over PLS and concluded that "we recommend the adoption of generalized structured component analysis as a sensible alternative to partial least squares...[since] [it] performed better than or as well as partial least squares in parameter recovery" (p. 710) Comparing the bias corrected GSCA_M to consistent PLS, Hwang et al. (2017) found that, with equal weights, GSCA_M biases parameter estimates positively, while PLSc biases them negatively. Again, they found consistently smaller standard errors for GSCA_M than PLSc. For heterogeneous weights, GSCA_M and PLSc exhibit bias in the same direction for loading estimates but in opposite directions for structural path estimates (negative for GSCA_M, positive for PLSc).

In a systematic simulation study, Hair et al. (2017b) showed that GSCA produces smaller estimation errors for measurement model parameters than PLS independent of samples size, number of indicators, and indicator weights. For equal weights, they found that summed scales outperformed both GSCA and PLS independent of sample sizes and number of indicators. Here, the complexities of PLS and GSCA actually prove detrimental and lead to worse outcomes. For unequal weights, and especially for few indicators and large sample sizes, summed scales show a large error, while GSCA and PLS perform about equally well. In a recent working paper, Hwang et al. (2020a) combined GSCA and GSCA_M for mixed composite and factor models ("integrated GSCA" or "IGSCA"). In their first simulation, they found that, as theoretically expected, GSCA and PLS "provided positively-biased estimates of the factor loadings" (p. 20) while "PLSc and IGSCA were the only approaches that provided unbiased estimates of all the loadings and path coefficients" (p. 20). Their second simulation study with a wider range of conditions pitted PLSc against IGSCA. Hwang et al. (2020a, p. 27) concluded that:

> *IGSCA always recovered loadings better than PLSc regardless of the experimental factors, while the two approaches recovered path coefficients similarly in most conditions except for the case of a misspecified model with cross component loadings, where IGSCA largely recovered path coefficients better.*

GSCA also features prominently in a recent special issue on composite analysis (Sarstedt & Hwang, 2020). Among a GSCA application (Jung et al., Rönkkö 2020) and a concept analysis of the literature (Hwang et al., 2020b), Cho and Choi (2020) presented a simulation study in which they compared the parameter recovery performance and statistical power of GSCA and PLS. They found similar results to Hair et al. (2017b) in that GSCA outperformed PLS in recovering measurement model parameters while they performed at a similar level in recovering structural parameters.

In addition to summed scales and GSCA, another alternative to PLS is regularized generalized canonical correlation analysis (RGCCA) (Tenenhaus & Tenenhaus, 2011, 2014; Tenenhaus et al., 2017). It builds on elements of both generalized canonical correlation analysis (GCCA) and PLS. While formally well developed, we know little about its comparative performance since few simulation studies have compared RGCCA to GSCA and PLS.

We write this section not to evaluate alternative component-based statistical models in detail but rather to ensure IS researchers know about these alternatives. The choice researchers must make is not one between covariance-based SEM and PLS. Instead, unweighted summed scales should be the primary alternative to PLS since it performs as well or better than complex indicator weighting schemes in many situations (Tenenhaus, 2008; Hair et al., 2017b). We should prefer simplicity over complexity when the outcomes are essentially the same. Simple methods are more robust in that they make fewer assumptions that one can violate, and researchers typically understand them better..

R14: Faced with multiple, equivalent methodological options, researchers should use the simplest method.

6 Conclusions

Overall, we have seen significant methodological developments that our 2013 critique at least partially drove (Rönkkö & Evermann, 2013). Many developments have focused on addressing concerns that we raised in that critique, such as estimation bias and model and parameter testing. Since then, the field has seen a plethora of methodological developments and simulation studies that examine these methodological developments. This work, which has significantly advanced our understanding of PLS, is very much welcome.

However, much of that work, such as bias-corrected PLSc and global fit assessment using model-implied covariance matrices, appears to have tried very hard to retrofit additions to PLS that come naturally and for free in covariance-based SEM. In contrast to covariance-based SEM, which rests on a unified foundation of covariance algebra, optimal fit functions, and maximum likelihood estimates, PLS looks increasingly like a hodgepodge of kludges added upon kludges. Many of the recent developments, such as model fit and model selection criteria, are based on ideas originally applied to covariance-based SEM, such as RMSEA, model-implied matrices, and the model likelihood. One has to wonder what, despite all these improvements, can be gained by using PLS over well-established, better understood, and conceptually simpler alternatives? Again, we leave this question as a rhetorical one for readers to ponder.

Some PLS researchers advocate a retreat to composite models and acknowledge that, while we have made advances in PLS, other methods better suit the common factor model. Yet, researchers have extensively debated whether to use formative and composite models, their place in applied research, and their ontology (Edwards & Bagozzi, 2000; Howell et al., 2007b, 2007a; Bagozzi 2007; Bollen, 2007; Wilcox et al., 2008; Marakas et al., 2007, 2008; Hardin et al., 2008; MacKenzie et al., 2011; Edwards, 2011; Markus, 2018). Moreover, while researchers have made important contributions to composite modeling, we still lack simulation research to indicate that PLS does indeed represent the best tool for this model type with the kinds of dataset that IS researchers use. As we note in Section 5, GSCA appears to be superior to PLS in many composite modeling applications, and so do simple, unweighted summed scales. The re-focusing of PLS on composite models also does not mean that researchers can simply declare or claim their model to be a composite one. The hypothesized relationships between observed variables and composites have to be theoretically plausible. Table 1 shows that many prominent IS theories deal with human psychology and include latent variables that represent beliefs, attitudes, or similar constructs. For such constructs, subjects' responses to measurement questions clearly causally depend on the measured psychological construct, which leads to a common factor model. While it may be possible to reframe these theories using composites, those composites would have very different meanings than the latent variables that reflect psychological constructs (Evermann & Tate, 2012).

Research has shown prediction from PLS estimated models to be superior to predictions from covariance-based SEM in many situations. However, the restrictions that PLS or otherwise estimated models impose limit prediction from them. The linearity of the models, the absence of certain paths in the model because they lack theoretical motivation, or the imposition of composite or factor structures represent constraints that necessarily reduce such models' predictive performance in practical applications compared to a theoretical prediction techniques such as neural networks or random forests. If users want to predict individual values for a particular application, they may be unwilling to accept lower predictive performance when they only gain a theoretically plausible, linear model. This trade-off between prediction and explanation is specific to every application but will not likely always go well for PLS (or any other linear structural equation modeling method).

In summary, PLS finds itself between more than one proverbial rock and a hard place. For the common factor model, it does not perform as well as the better understood, conceptually simpler, and statistically grounded covariance-based SEM. For the composite model, summed scales in some applications and GSCA in the remaining situations outperform PLS. For predictive modeling, it is threatened by various contemporary atheoretical prediction methods, such as deep learning networks, which outperform traditional statistical learning techniques in many different applications.

Additionally, PLS can naturally only model composites and cannot truly model latent variables that have no associated indicators. Such latter models include latent growth curves models (Bollen & Curran, 2006) and latent difference score models (McArdle, 2009; Grimm et al., 2012). While composites can approximate latent variables in some models, researchers in many research disciplines (which includes the IS discipline) have increasingly begun using models with truly latent variables (i.e., without directly associated observed variables) (e.g., Serva et al., 2011; Bala & Venkatesh 2013).

On a positive note, PLS researchers and critics agree on many issues. For instance, they universally acknowledge sample size requirements as an invalid reason to choose PLS. They also agree that parameter estimates in the original PLS algorithm exhibit bias when applied to the common factor model. Furthermore, they agree that, as Rigdon (2016, p. 602) points out, weakly correlated or uncorrelated constructs are “outside of its [PLS] boundary conditions”, though researchers have paid little attention to this requirement in IS applications. Finally, they agree that one should use bootstrap CIs over t-tests, although IS researchers also rarely do so.

With this paper, we contribute to minimum methodological competence standards and began by asking what such standards might look like. To give an example, one would hope to not see a paper that uncritically used traditional mode A for a common factor model, had a sample size of 30 while measuring five constructs using 20 indicators, employed t-tests to test a weak effect size for significance, and used CR and AVE criteria to justify model fit. On the other hand, a paper that used PLSc, had done a priori power analyses to arrive at a sample size of 500, demonstrated good model fit using SRMR and d_G, and used percentile bootstrapping confidence intervals around a strong parameter estimate would give a lot more confidence in the results.

Similarly, it would be inappropriate for researchers to claim they conducted a predictive study but then proceed with a common factor model, to evaluate model fit and factorial structure, and to report the R^2 values for the dependent latent variables. It would be more convincing for such researchers to motivate the importance of prediction, relegate model tests and evaluation of the factor structure to secondary considerations or omit them entirely, and present out-of-sample predictive statistics such as mean squared error (MSE) for a 10-fold cross-validated model. As another example, when employing composite models, researchers should explicitly justify the departure from the still prevalent measurement using common factor models. In other words, they cannot simply declare a construct to be a composite; they should theoretically justify the choice, and the manifest variables should be demonstrably useful in creating a meaningful composite. They can also not base a subsequent evaluation on traditional factor analytic criteria either.

In summary, we are pleased to see that our 2013 critique has spurred many constructive developments in PLS and how researchers use it. Among other reasons, we wrote this paper because some authors still cite papers on PLS from 25 or 30 years ago while being apparently unaware of both the criticisms and the recent advances. We write this paper to help IS researchers remain current with important methodological developments and choose statistical methods based on a sound foundation.

References

Agarwal, R., & Karahanna, E. (2000). Time Flies when you're having fun: Cognitive absorption and beliefs about information technology usage. *MIS Quarterly*, *24*(4), 665-694.

Aguirre-Urreta, M. I., & Marakas, G. M. (2008). The use of PLS when analyzing formative constructs: Theoretical analysis and results from simulations. In *Proceedings of the International Conference on Information Systems.*

Aguirre-Urreta, M. I., & Marakas, G. M. (2014). Research note: Partial least squares and models with formatively specified endogenous constructs: A cautionary note. *Information Systems Research*, *25*(4), 761-778.

Aguirre-Urreta, M. I., & Rönkkö, M. (2018). Statistical inference with PLSc using bootstrap confidence intervals. *MIS Quarterly, 42*(3), 1001-1020.

Aguirre-Urreta, M. I., & Rönkkö, M. (2015). Sample size determination and statistical power analysis in PLS using R: An annotated tutorial. *Communications of the Association for Information Systems*, *36,* 33-51.

Aguirre-Urreta, M. I., Rönkkö, M., & McIntosh, C. N. (2019). A cautionary note on the finite sample behavior of maximal reliability. *Psychological Methods*, *24*(2), 236-252.

Ali, F., Rasoolimanesh, S. M., Sarstedt, M., Ringle, C. M., & Ryu, K. (2018). An assessment of the use of partial least squares structural equation modeling (PLS-SEM) in hospitality research. *International Journal of Contemporary Hospitality Management, 30*(1), 514-538.

Antonakis, J., Bendahan, S., Jacquart, P., & Lalive, R. (2010). On making causal claims: A review and recommendations. *The Leadership Quarterly, 21*(6), 1086-1120.

Bagozzi, R. P. (2007). On the meaning of formative measurement and how it differs from reflective measurement: Comment on Howell, Breivik, and Wilcox (2007). *Psychological Methods*, *12*(2), 229-237.

Bala, H., & Venkatesh, V. (2013). Changes in employees' job characteristics during an enterprise system implementation: A latent growth modeling perspective. MI*S Quarterly, 37*(4), 1113-1140.

Bassellier, G., & Benbasat, I. (2004). Business competence of information technology professionals: Conceptual development and influence on IT-business partnerships. *MIS Quarterly, 28*(4), 673-694.

Becker, J.-M., Ringle, C. M., & Sarstedt, M. (2018). Estimating moderating effects in PLS-SEM and PLSc-SEM: interaction term generation*data treatment. *Journal of Applied Structural Equation Modeling, 2*(2), 1-21.

Benitez-Amado, J., Henseler, J., & Roldán, J. L. (2016). How to address endogeneity in partial least squares path modeling. In *Proceedings of the 22nd Americas Conference on Information Systems*.

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.

Bhattacherjee, A., & Premkumar, G. (2004). Understanding changes in belief and attitude toward information technology usage: A theoretical model and longitudinal test. *MIS Quarterly, 28*(2), 229-254.

Bollen, K. A. (2007). Interpretational confounding is due to misspecification, not to type of indicator: Comment on Howell, Breivik, and Wilcox (2007). *Psychological Methods, 12*(2), 219-228.

Bollen, K. A., & Curran, P. J. (2006). *Latent growth curve models: A structural equation approach.* Wiley.

Cadogan, J. W., & Lee, N. (2013). Improper use of endogenous formative variables. *Journal of Business Research, 66*(2), 233-241.

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). Lawrence Erlbaum Associates.

Chin, W. W., & Gopal, A. (1995). Adoption intention in GSS: Relative importance of beliefs. *ACM SIGMIS Database: The DATABASE for Advances in Information Systems, 26*(2-3), 42-64.

Chin, W. W., Marcolin, B. L., & Newsted, P. R. (1996). A PLS latent variable modelling approach for measuring interaction effects: Results from a Monte-Carlo simulation study and voice mail emotion/adoption study. In *Proceedings of the Seventeenth International Conference on Information Systems.*

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). Sage.

Cho, G., & Choi, J. Y. (2020). An empirical comparison of generalized structured component analysis and partial least squares path modeling under variance-based structural equation models. *Behaviormetrika, 47*(1), 243-272.

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

Compeau, D., & Higgins, C. A. (1995). Computer self-efficacy: Development of a measure and initial test. *MIS Quarterly, 19*(2), 189-211.

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

Devlieger, I., Mayer, A., & Rosseel, Y. (2015). Hypothesis testing using factor score regression: A comparison of four methods. *Educational and Psychological Measurement*, *76*(5), 741-770.

Devlieger, I., & Rosseel, Y. (2017). Factor score path analysis. *Methodology*, *13*(1), 31–38.

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

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.

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

Dijkstra, T. K., & Henseler, J. (2015b). Consistent and asymptotically normal PLS estimators for linear structural equations. *Computational Statistics & Data Analysis, 81*, 10-23.

Dijkstra, T. K., & Schermelleh-Engel, K. (2014). Consistent partial least squares for nonlinear structural equation models. *Psychometrika*, *79*(4), 585-604.

Edwards, J. R. (2011). The fallacy of formative measurement. *Organizational Research Methods, 14*(2), 370–88.

Edwards, J. R., & Bagozzi, R. P. (2000). On the nature and direction of relationships between constructs and measures. *Psychological Methods, 5*(2), 155-174.

Evermann, J., & Tate, M. (2010). Testing models or fitting models? Identifying model misspecification in PLS. In *Proceedings of the International Conference on Information Systems.*

Evermann, J., & Tate, M. (2012). An ontology of structural equation models with application to computer self-efficacy. In *Proceedings of the International Conference on Information Systems.*

Evermann, J., & Tate, M. (2014). Bayesian structural equation models for cumulative theory building in information systems—a brief tutorial using BUGS and R. *Communications of the AIS*, *34,* 1481-1514.

Evermann, J., & Tate, M. (2016). Assessing the predictive performance of structural equation model estimators. *Journal of Business Research, 69*(10), 4565-4582.

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.

Ghasemy, M., Jamil, H., & Gaskin, J. E. (2020). Have your cake and eat it too: PLSe2 = ML + PLS. *Quality & Quantity*, *55*, 497-541.

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 *Proceedings of the Americas Conference on Information Systems.*

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.

Grimm, K. J., An, Y., McArdle, J. J., Zonderman, A. B., & Resnick, S. M. (2012). Recent changes leading to subsequent changes: Extensions of multivariate latent difference score models. *Structural Equation Modeling: A Multidisciplinary Journal, 19*(2), 268-292.

Guyon, H. (2018). The fallacy of the theoretical meaning of formative constructs. *Frontiers in Psychology*. Retrieved from https://doi.org/10.3389/fpsyg.2018.00179

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate data analysis* (7th ed.). Pearson.

Hair, J. F, Ringle, C. M., & Sarstedt, M. (2011). PLS-SEM: Indeed a silver bullet. *Journal of Marketing Theory and Practice, 19*(2), 139-152.

Hair, J. F, Sarstedt, M., Ringle, C. M., & Mena, J. A. (2012a). 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.

Hair, J. F., Hollingsworth, C. L., Randolph, A. B., & Chong, A. Y. L. (2017a). An updated and expanded assessment of PLS-SEM in information systems research. *Industrial Management & Data Systems, 117*(3), 442-458.

Hair, J. F., Hult, G. T. M., Ringle, C. M., Sarstedt, M., & Thiele, K. O. (2017b). Mirror, mirror on the wall: A comparative evaluation of composite-based structural equation modeling methods. *Journal of the Academy of Marketing Science, 45*(5), 616-632.

Hair, J. F., Ringle, C. M., & Sarstedt, M. (2013). Partial least squares structural equation modeling: Rigorous applications, better results and higher acceptance. *Long Range Planning, 46*(1-2), 1-12.

Hair, J. F., Sarstedt, M., Pieper, T. M., & Ringle, C. M. (2012b). The use of partial least squares structural equation modeling in strategic management research: A review of past practices and recommendations for future applications. *Long Range Planning, 45*(5-6), 320-340.

Hair, J. F., Sarstedt, M., & Ringle, C. M. (2019). Rethinking some of the rethinking of partial least squares. *European Journal of Marketing*, *53*(4), 566-584.

Hardin, A. (2017). A call for theory to support the use of causal-formative indicators: A commentary on Bollen and Diamantopoulos (2017). *Psychological Methods*, *22*(3), 597-604.

Hardin, A. M., Chang, J. C.-J., & Fuller, M. A. (2008). Formative vs. reflective measurement: Comment on Marakas, Johnson, and Clay (2007). *Journal of the Association for Information Systems, 9*(9), 519-535.

Henseler, J. (2017). Bridging design and behavioral research with variance-based structural equation modeling. *Journal of Advertising, 46*(1), 178-192.

Henseler, J., Dijkstra, T. K., Sarstedt, M., Ringle, C. M., Diamantopoulos, A., Straub, D. W., Ketchen, D. J., Hair, J. F., Hult, G. T. M., & Calantone, R. J. (2014). Common beliefs and reality about PLS: Comments on Rönkkö and Evermann (2013). *Organizational Research Methods, 17*(2), 182-209.

Henseler, J., & Fassott, G. (2010). Testing moderating effects in PLS path models: An illustration of available procedures. In V. Esposito Vinzi, W. W. Chin, J. Henseler, & H. Wang (Eds.), *Handbook of partial least squares* (pp. 713-735). Springer.

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.

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. In R. R. Sinkovics & P. N. Ghauri (Eds.), *New challenges to international marketing* (pp. 277-319). Emerald.

Henseler, J., & Sarstedt, M. (2013). Goodness-of-fit indices for partial least squares path modeling. *Computational Statistics*, *28*(2), 565-580.

Howell, R. D., Breivik, E., & Wilcox, J. B. (2007a). Is formative measurement really measurement? Reply to Bollen (2007) and Bagozzi (2007). *Psychological Methods, 12*(2), 238-245.

Howell, R. D., Breivik, E., & Wilcox, J. B. (2007b). Reconsidering formative measurement. *Psychological Methods*, *12*(2), 205-218.

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.

Huang, W. (2013). *PLSe: Efficient estimators and tests for partial least squares* (doctoral dissertation). University of California, Los Angeles, CA.

Hui, B. S., & Wold, H. (1982). Consistency and consistency at large of partial least squares estimates. In K. Jöreskog & H. Wold (Eds.), *Systems under indirect observation: Causality, structure, prediction, Amsterdam.* North-Holland.

Hulland, J. (1999). Use of partial least squares (PLS) in strategic management research: A review of four recent studies. *Strategic Management Journal, 20*(2), 195-204.

Hult, G. T. M., Hair, J. F., Jr., Proksch, D., Sarstedt, M., Pinkwart, A., & Ringle, C. M. (2018). Addressing endogeneity in international marketing applications of partial least squares structural equation modeling. *Journal of International Marketing, 26*(3), 1-21.

Hwang, H. (2009). Regularized generalized structured component analysis. *Psychometrika, 74*(3), 517-530.

Hwang, H., Cho, G., Jung, K., Falk, C., Flake, J., Jin, M. J., & Lee, S. H. (2020a). An approach to structural equation modeling with both factors and components: Integrated generalized structured component analysis. P*sychological Methods*, *26*(3), 273-294.

Hwang, H., Ho, M.-H. R., & Lee, J. (2010a). Generalized structured component analysis with latent interactions. *Psychometrika, 75*(2), 228-242.

Hwang, H., Malhotra, N. K., Kim, Y., Tomiuk, M. A., & Hong, S. (2010b). A comparative study on parameter recovery of three approaches to structural equation modeling. *Journal of Marketing Research*, *47*(4), 699-712.

Hwang, H., Sarstedt, M., Cheah, J. H., & Ringle, C. M. (2020b). A concept analysis of methodological research on composite-based structural equation modeling: Bridging PLSPM and GSCA. *Behaviormetrika*, *47*(1), 219-241.

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

Hwang, H., & Takane, Y. (2009). Nonlinear generalized structured component analysis. *Behaviormetrika, 37*(1), 1-14.

Hwang, H., & Takane, Y. (2014). *Generalized structured component analysis: A component-based approach to structural equation modeling.* CRC Press.

Hwang, H., Takane, Y., & Jung, K. (2017). Generalized structured component analysis with uniqueness terms for accommodating measurement error. *Frontiers in Psychology.* Retrieved from https://www.frontiersin.org/articles/10.3389/fpsyg.2017.02137/full

Hwang, H., Takane, Y., & Malhotra, N. (2007). Multilevel generalized structured component analysis. *Behaviormetrika, 34*(2), 95-109.

Jöreskog, K. G., & Wold, H. (1982). The ML and PLS techniques for modeling with latent variables. In K. G. Jöreskog & H. Wold (Ed.), *Systems under indirect observation: Causality, structure, prediction*. North-Holland.

Jung, K., Cho, S. S., Lee, J., Kim, S., & Ryoo, J. H. (2020). An illustrative application of generalized structured component analysis for brain connectivity research. *Behaviormetrika, 47*(1), 273-289.

Karahanna, E., Straub, D. W., & Chervany, N. L. (1999). Information technology adoption across time: A cross-sectional comparison of pre-adoption and post-adoption beliefs. *MIS Quarterly, 23*(2), 183-213.

Kline, R. B. (2011). *Principles and practice of structural equation modeling* (3rd ed.). Guilford Press.

Liengaard, B. D., Sharma, P. N., Hult, G. T. M., Jensen, M. B., Sarstedt, M., Hair, J. F., & Ringle, C. M. (2020). Prediction: Coveted, yet forsaken? Introducing a cross-validated predictive ability test in partial least squares path modeling. *Decision Sciences*, *52*(2), 362-392

Lohmöller, J.-B. (1989). *Latent variable path modeling with partial least squares.* Physica-Verlag.

MacKenzie, S. B, Podsakoff, P. M., & Podsakoff, N. P. (2011). Construct measurement and validation procedures in MIS and behavioral research: Integrating new and existing techniques*. MIS Quarterly, 35*(2), 293-334.

Marakas, G., Johnson, R., & Clay, P. F. (2007). The Evolving Nature of the Computer Self-Efficacy Construct: An Empirical Investigation of Measurement Construction, Validity, Reliability and Stability over Time. Journal of the Association for Information Systems 8 (1), 2.

Marakas, G., Johnson, R. D., & Clay, P. F. (2008). Formative vs. reflective measurement: A reply to Hardin, Chang, and Fuller. *Journal of the Association for Information Systems*, *9*(9),

Markus, K. A. (2018). Three conceptual impediments to developing scale theory for formative scales. *Methodology, 14*(4), 156-164.

McArdle, J. J. (2009). Latent variable modeling of differences and changes with longitudinal data. An*n*ual *Review of Psychology, 60,* 577-605.

McIntosh, C. N., Edwards, J. R., & Antonakis, J. (2014). Reflections on partial least squares path modeling. *Organizational Research Methods, 17*(2), 210-251.

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.

Petter, S. (2018). Haters gonna hate: PLS and information systems research. *ACM SIGMIS Database: The DATABASE for Advances in Information Systems, 49*(2), 10-13.

Rademaker, M. E., Schuberth, F., & Dijkstra, T. K. (2019). Measurement error correlation within blocks of indicators in consistent partial least squares. *Internet Research*, *29*(3), 448-463.

Reinartz, W., 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.

Rhemtulla, M., van Bork, R., & Borsboom, D. (2020). Worse than measurement error: Consequences of inappropriate latent variable measurement models. *Psychological Methods*, *25*(1), 30-45.

Rigdon, E. E. (2016). Choosing PLS path modeling as analytical method in European management research: A realist perspective. *European Management Journal*, *34*(6), 598-605.

Rigdon, E. E. (2012). Rethinking partial least squares path modeling: In praise of simple methods. *Long Range Planning*, *45*(5-6), 341-358.

Rigdon, E. E, Becker, J.-M., Rai, A., Ringle, C. M., Diamantopoulos, A., Karahanna, E., Straub, D. W., & Dijkstra, T. K. (2014). Conflating antecedents and formative indicators: A comment on Aguirre-Urreta and Marakas. *Information Systems Research*, *25*(4), 780-784.

Rigdon, E. E, Sarstedt, M., Ringle, C. M. (2017). On comparing results from CB-SEM and PLS-SEM: Five perspectives and five recommendations. *Marketing* ZFP, *39*(3), 4-16.

Ringle, C, M, Sarstedt, M., Mitchell, R., & Gudergan, S. P. (2020). Partial least squares structural equation modeling in HRM research. *The International Journal of Human Resource Management*, *31*(12), 1617-1643.

Ringle, C. M., Sarstedt, M., & Straub, D. W. (2012). A critical look at the use of PLS-SEM in MIS Quarterly. *MIS Quarterly*, *36*(1), iiv-xiv.

Rosseel, Y. (2020). Small sample solutions for structural equation modeling. In R. van de Schoot & M. Miočević (Eds.), *Small sample size solutions: A guide for applied researchers and practitioners* (pp. 226-238). Taylor & Francis.

Rouse, A., & Corbitt, B. (2008). There's SEM and SEM: A critique of the use of PLS regression in information systems research. In *Proceedings of the Australasian Conference on Information Systems.*

Rönkkö, M. (2014). The effects of chance correlations on partial least squares path modeling. *Organizational Research Methods, 17*(2), 164-181.

Rönkkö, M. (2020). Introduction to Matrixpls. https://www.researchgate.net/publication/303936869_Introduction_to_matrixpls_100

Rönkkö, M., Lee, N., Evermann, J., McIntosh, C., & Antonakis, J. (2021) Marketing of Methodology? Exposing the Fallacies of PLS with Simple Demonstration. In press at *European Journal of Marketing*.

Rönkkö, M., & Cho, E. (2020). An updated guideline for assessing discriminant validity. *Organizational Research Methods, 25*(1), 6-14.

Rönkkö, M., & Evermann, J. (2013). A critical examination of common beliefs about partial least squares path modeling. *Organizational Research Methods*, *16*(3), 425-448.

Rönkkö, M., Evermann, J., & Aguirre-Urreta, M. I. (2016a). *Estimating formative measurement models in IS research: Analysis of the past and recommendations for the future*. Retrieved from http://urn.fi/URN:NBN:fi:aalto-201605031907

Rönkkö, M., McIntosh, C. N., & Aguirre-Urreta, M. I. (2016b). *Improvements to PLSc: Remaining problems and simple solutions.* Retrieved from https://core.ac.uk/download/pdf/80718697.pdf

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.

Rönkkö, M., McIntosh, C. N., Antonakis, J., & Edwards, J. R. (2016c). Partial least squares path modeling: Time for some serious second thoughts. *Journal of Operations Management*, *47-48*(1), 9-27.

Rönkkö, M., Parkkila, K., & Ylitalo, J. (2012). Use of partial least squares as a theory testing tool—an analysis of information systems papers. In *Proceedings of the European Conference on Information Systems.*

Rönkkö, M., & Ylitalo, J. (2010). Construct validity in partial least squares path modeling. In *Proceedings of the International Conference on Information Systems.*

Sarstedt, M., Hair, J. F., Ringle, C. M., Thiele, K. O., & Gudergan, S. P. (2016). Estimation issues with PLS and CBSEM: Where the bias lies! *Journal of Business Research*, *69*(10), 3998-4010.

Sarstedt, M., & Hwang, H. (2020). Advances in composite-based structural equation modeling. *Behaviormetrika, 47,* 213-217.

Sarstedt, M., Ringle, C. M., Cheah, J.-H., Ting, H., Moisescu, O. I., & Radomir, L. (2020). Structural model robustness checks in PLS-SEM. *Tourism Economics, 26*(4), 531-554.

Schuberth, F., Henseler, J., & Dijkstra, T. K. (2018). Confirmatory composite analysis. *Frontiers in Psychology.* Retrieved from https://www.frontiersin.org/articles/10.3389/fpsyg.2018.02541/full

Serva, M. A., Kher, H., & Laurenceau, J. (2011). Using latent growth modeling to understand longitudinal effects in MIS theory: A primer. *Communications of the Association for Information Systems*, *28*, 213-232.

Sharma, P. N., Shmueli, G., Sarstedt, M., Danks, N., & Ray, S. (2019a). Prediction-oriented model selection in partial least squares path modeling. *Decision Sciences*, *52*(3), 567-706.

Sharma, P., Sarstedt, M., Shmueli, G., Kim, K. H., & Thiele, K. O. (2019b). PLS-based model selection: The role of alternative explanations in information systems research. *Journal of the Association for Information Systems, 20*(4), 346-397.

Shmueli, G., Ray, S., Estrada, J. M. V., & Chatla, S. B. (2016). The elephant in the room: Predictive performance of PLS models. *Journal of Business Research*, *69*(10), 4552-4564.

SmartPLS. (2020) *SmartPLS—frequently asked questions.* Retrieved from https://www.smartpls.com/faq/documentation/documentation-or-manual

Spearman, C. (1904). The proof and measurement of association between two things. *The American Journal of Psychology, 15*(1), 72-101.

Tenenhaus, A., & Tenenhaus, M. (2011). Regularized generalized canonical correlation analysis. *Psychometrika, 76*(2), 257-284.

Tenenhaus, A., & Tenenhaus, M. (2014). Regularized generalized canonical correlation analysis for multiblock or multigroup data analysis. *European Journal of Operational Research*, *238*(2), 391-403.

Tenenhaus, M. (2008). Component-based structural equation modelling. *Total Quality Management, 19*(7-8), 871-886.

Tenenhaus, M., Amato, S., & Vinzi, V. E. (2004). A global goodness-of-fit index for PLS structural equation modelling. In *Proceedings of the XLII SIS Scientific Meeting*, *1*, 739-742.

Tenenhaus, M., Tenenhaus, A., & Groenen, P. J. F. (2017). Regularized generalized canonical correlation analysis: A framework for sequential multiblock component methods. *Psychometrika, 82*(3), 737-777.

Teo, H.-H., Wei, K. W., & Benbasat, I. (2003). Predicting intention to adopt interorganizational linkages: An institutional perspective. *MIS Quarterly, 27*(1), 19-49.

Thatcher, J. B., & Perrewé, P. L. (2002). An empirical examination of individual traits as antecedents to computer anxiety and computer self-efficacy. *MIS Quarterly*, *26*(4), 381-396.

Do Valle, P. O., & Assaker, G. (2016). Using partial least squares structural equation modeling in tourism research: A review of past research and recommendations for future applications. *Journal of Travel Research, 55*(6), 695-708.

Vandenberg, R. J. (2006). Introduction: statistical and methodological myths and urban legends: Where, pray tell, did they get this idea? *Organizational Research Methods*, *9*(2), 194-201.

Venkatesh, V. (2000). Determinants of perceived ease of use: Integrating control, intrinsic motivation, and emotion into the technology acceptance model. *Information Systems Research, 11*(4), 342-365.

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly, 27*(3), 425-478.

Wilcox, J. B, Howell, R. D., & Breivik, E. (2008). Questions about formative measurement. *Journal of Business Research*, *61*(12), 1219-1228.

Wixom, B. H., & Todd, P. A. (2005). A theoretical integration of user satisfaction and technology acceptance. *Information Systems Research, 16*(1), 85-102.

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). North-Holland.

Wold, H. (1985). Systems analysis by partial least squares. In P. Nijkamp, L. Leitner, & N. Wrigley (Eds.), *Measuring the unmeasurable* (pp. 221-252). Marinus Nijhoff.

Yuan, K.-H., Wen, Y., & Tang, J. (2020). Regression analysis with latent variables by partial least squares and four other composite scores: Consistency, bias and correction. *Structural Equation Modeling: A Multidisciplinary Journal*, *27*(3), 333-350.

About the Authors

Joerg Evermann received his PhD in Information Systems from the University of British Columbia. Prior to being a faculty member at Memorial University, he was a lecturer in Information Systems with the School of Information Management at the University of Wellington, New Zealand. Joerg's research interests are in business process management, statistical research methods, and information integration. He has published his research in more than 70 peer-reviewed publications. His work has appeared in high-quality journals, such as *Decision Support Systems*, *IEEE Transactions on Software Engineering*, *IEEE Transactions on Knowledge and Data Engineering*, *IEEE Transactions on Services Computing*, *Journal of Business Research*, *Communications of the* AIS, *Organizational Research Methods*, *Structural Equation Modeling*, *Journal of the AIS*, *Information systems*, *Business and Information Systems Engineering* and *Information Systems Journal*. He has presented his work at international conferences, such as ICIS, AMCIS, CAiSE, ER, among others. He serves on the editorial board of *Communications of the AIS*.

Mikko Rönkkö is associate professor of entrepreneurship at Jyväskylä University School of Business and Economics (JSBE) and a docent at Aalto University School of Science in the field of statistical methods in management research. His current research interests are technology entrepreneurship and quantitative research methods in management research. He has published in *MIS Quarterly*, *Journal of Operations Management*, *Psychological Methods*, and *Organizational Research Methods* among other journals. He is a department editor at the Journal of Operations Management handling methodological articles and is on the editorial boards of Organizational Research Methods and Entrepreneurship Theory and Practice.