Communications of the Association for Information Systems

Research Paper DOI: 10.17705/1CAIS.044XX ISSN: 1529-3181

Recent Developments in PLS

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

Mikko Rönkkö
University of Jyväskylä
Jyväskylä, Finland

Abstract:

Partial Least Squares path modeling (PLS) is a method for estimating linear structural equation models. Widely used in the information systems (IS) discipline, there has been considerable argument over its relative merits compared to simple summed scores or to covariance-based estimation of structural equation models. This paper comments 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. The paper briefly reviews the mechanisms of PLS, its well-known properties of PLS, 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 practices and use of PLS. That critique proved the driver for many advances in the field, which are discussed extensively and used to make 14 recommendations for how and when to use PLS or alternatives.

Keywords: Partial least squares, PLS, structural equation modeling, statistics, research methods.

[Department statements, if appropriate, will be added by the editors. Teaching cases and panel reports will have a statement, which is also added by the editors.]

[Note: this page has no footnotes.]

This manuscript underwent [editorial/peer] review. It was received xx/xx/20xx and was with the authors for XX months for XX revisions. [firstname lastname] served as Associate Editor. **or** The Associate Editor chose to remain anonymous.]

1 Introduction

Partial Least Squares path modelling (PLS) 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 developed by Karl Jöreskog (Jöreskog and Wold 1982). IS researchers were among the earliest adopters of PLS in the social sciences (Chin and Gopal 1995; Chin, Marcolin, and Newsted 1996, 2003; Chin and Newsted 1999). PLS has been increasingly popular, not only in IS research (Ringle, Sarstedt, and Straub 2012) but in many other disciplines, such as marketing (Henseler, Ringle, and Sinkovics 2009; Hair et al. 2012), tourism (Valle and Assaker 2016), and hospitality (Ali et al. 2018). The technique has also been introduced to other business disciplines such as operations management (Peng and Lai 2012) and strategic management (Hulland 1999; Hair, Sarstedt, et al. 2012), but gained less traction there.

Beginning in 2012, realizing that many of the claims by PLS proponents were only weakly substantiated and noting the increasing voices of caution on the use of PLS (Goodhue, Lewis, and Thompson 2007, 2012; Aguirre-Urreta and Marakas 2008; Rouse and Corbitt 2008; Reinartz, Haenlein, and Henseler 2009; Evermann and Tate 2010; Rönkkö and Ylitalo 2010), we independently began to empirically investigate the behaviour of PLS. Becoming aware of each others' research at ICIS 2010 (Evermann and Tate 2010; Rönkkö and Ylitalo 2010), we joined forces to present a critical assessment of PLS, using a very basic structural equation model, in a journal article published in 2013 (Rönkkö and Evermann 2013). Motivated by our critique, there have been many positive developments around PLS in recent years, both to further investigate its performance, but also, 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 the use of PLS. Instead, we highlight recent developments to ensure that IS researchers have up-to-date methodological knowledge of PLS if they decide to use it. As presented below, there has been a veritable flood of new developments in the past 5–10 years. This paper also helps to set minimum standards methodological competence when using PLS. In essence, *if* researchers choose to use PLS, they must be mindful of the latest advances for the method and use best practices.

So what do we mean by ”minimum standards”? For example, 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 a significance tests on weak path estimates and justifies the use of PLS by appealing to its predictive capabilities, all the while reporting R^2 and focusing on properties of the factor model 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 supporting the use of these procedures inspires significantly more confidence in its validity. Often, however, researchers will find themselves between these two ideals. In this situation it is important to be cognizant of the pros and cons and of the trade-offs inherent in every statistical technique, and argue persuasively based on substance (i.e. what is known based on methodological research), not by appeal to authority or prior studies.

Again, we do not intend to argue for or against the use of PLS compared to other methods; it is left to readers to come to their own conclusions. Moreover, in reviewing recent developments, the paper does not engage with the many ”review & guidelines” papers (Sarstedt et al. 2020; Ringle et al. 2020; Ali et al. 2018; Valle and Assaker 2016; Hulland 1999, 1999; Hair, Sarstedt, et al. 2012; Hair et al. 2017; Henseler, Ringle, and Sinkovics 2009; Hair et al. 2012; Peng and Lai 2012; Ringle, Sarstedt, and Straub 2012) and polemic ”advocacy” pieces (Hair, Ringle, and Sarstedt 2011, 2013; Petter 2018; Hair, Sarstedt, and Ringle 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.

The remainder of the paper is structured as follows: Section 2 is a review of linear structural equation models, PLS, and the history of PLS use in IS research. Section 3 revisits our 2013 critique of PLS as a starting point for this commentary. Section 4 presents recent methodological developments of PLS and makes recommendations for researchers based on the presented arguments and evidence. Section 5 discusses alternatives to PLS and Section 6 conclude with a brief summary and discussion.

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

Structural equation modeling is a set of methods to estimate 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

There are different types of structural equation models. While models can be classified in different ways, for our purposes it is important to distinguish by the way in which observed and unobserved variables are connected.

Consider the common factor model shown in Figure 1. The variables ξ and η are latent variables, each of the x_i are manifest variables. For every dependent variable (or every regression equation), there is an error term (ϵ_i, ζ). The figure shows parameters of the linear regression relationships on the arrows. Additionally, every independent variable (ϵ_i, ξ, ζ) has a variance; variances are not usually shown, and there exist covariances between all exogenous variables, shown by double-headed arrows. A typical assumption (model constraint) is that most of these covariances are zero, for example those between the ϵ_i and ξ or between ξ and ζ. However, this assumption may be relaxed for a subset of the indicators as long as the model remains statistically identified.

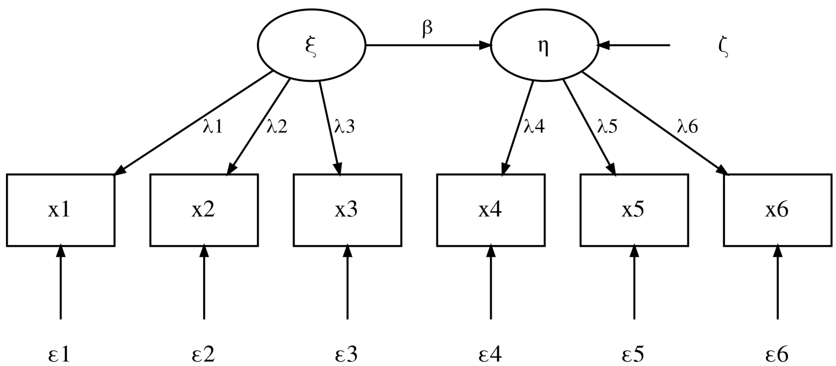

Figure 1. Common factor model, two latent variables each with three indicators and measurement errors

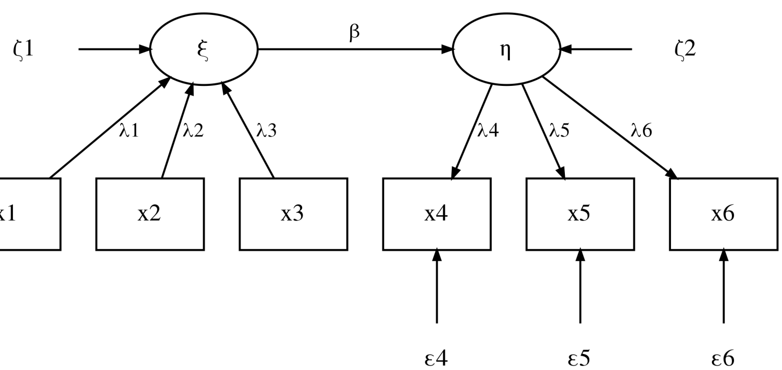

Figure 2. Mixed formative-reflective model (correlations among the exogenous x indicators are omitted for clarity)

Another type of model is the mixed formative-reflective model, such as that in Figure 2. ξ is modelled as 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, their covariances need not be estimated, but are observed, and are not parameters in the model. Note also that ξ is now a dependent variable and therefore has an associated error term ζ_1.

There is considerable debate in the literature about the usefulness and appropriateness of different types of models for various research applications (Cadogan and Lee 2013; Edwards 2011; Howell, Breivik, and Wilcox 2007b, 2007a; Bagozzi 2007; Wilcox, Howell, and Breivik 2008; Hardin 2017; Markus 2018; Guyon 2018), also in the IS discipline (Marakas, Johnson, and Clay 2007, 2008; Hardin, Chang, and Fuller 2008). However, that debate is a topic for another paper; this paper assumes that a researcher has identified a suitable statistical model and is choosing a method to estimate its parameters.

2.2 Principles of PLS

Partial least squares path modelling is a method for estimating structural equation models. In stating this, it is acknowledged that there is significant discussion about whether it is a good technique for doing so. Nonetheless, PLS is widely used by IS researchers to estimate all of these model types (Ringle, Sarstedt, and Straub 2012). This section describes classic PLS; consistent PLS (”PLSc”) is a significant extension described in Sec. 4.1 below.

PLS is a two-phase method. The first phase 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 composite model shown in Figure 3, where the composites are used as approximations of 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 the absence of error terms on ξ, while the error term ζ on η is supposed to represent the error term of the structural regression only (cf. Section 4.8 for further discussion). It is clear that the composites can only be imperfect approximations (”proxies”) of the latent variables in the common factor and the formative model: Fig. 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. Fig. 2 shows that formative latent variables have an associated error term, which is not part of the PLS composite.

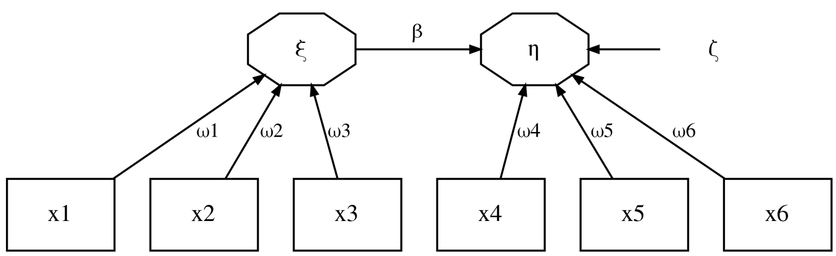

Figure 3. Composite model (correlations among the x indicators are omitted for clarity)

Composites are formed by weighting the manifest variables using the following iterative algorithm. PLS 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”. Outer weights are determined either by 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”, typically by examining the sign of the covariance of connected composites, and then forms the ”inner approximation” as a weighted sum of the earlier outer approximations. 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 the composite-proxies and indicators.

The only difference between PLS and regression with simple (unweighted) summed scales are the indicator weights. 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 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). This was showcased in a MISQ editorial 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" (Gefen, Rigdon, and Straub 2011, v). However, as there is no set of perfectly reliable indicators, no weighting system can complete completely exclude measurement error (Henseler et al. 2014). To what extent PLS weights are effective in reducing the effects of measurement error is addressed in Section 3.1.

2.3 Known Limitations

The inclusion of measurement error in the independent variables in a regression attenuates (biases, reduces) the estimate of regression coefficients, a well-known fact first discussed by Spearman (1904) 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, PLS estimates have been known to be biased when the algorithm was developed, before even the earliest application in IS research. Herman Wold, the original developer of PLS acknowledged this as early as 1982 (Hui and Wold 1982). Specifically, the bias decreases and the estimates approach the true value as sample size increases and the number of indicators for each latent variable increases. Comparing PLS and SEM, Jöreskog and Wold (1982 pg. 266) write that ”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 pg. 28) writes: ”Sacrificing optimality, PLS rests content with consistency, albeit in the qualified sense of consistency at large”, and also in (Wold 1985 pg. 213): ”The PLS estimates of parameters and of LV case values are inconsistent inasmuch 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, 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] $R^2(q_i, f_i)$”. (pg. 81) In his seminal work on PLS, Lohmöller (1989 pg. 207) demonstrates 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 (Chin 1998 pg. 330): ”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, an extensive and systematic simulation study, conducted by Reinartz, Haenlein, and Henseler (2009), concludes with respect to the quality of the estimates that ”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.” (pg. 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 1980’s, one has to wonder why the IS discipline thought that adopting a designed-to-be-biased estimator for the common factor model was a good idea. This is of course a rhetorical question, left 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 by Chin (1998 pg. 311): ”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.” The small sample argument was the focus of a subsequent chapter by Chin and Newsted (1999) in which they 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 by Goodhue, Lewis, and Thompson (2012) to 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).” (pg. 998). In the following year, our study (Rönkkö and Evermann 2013) pointed out how Chin and Newsted (1999) had misinterpreted their evidence; this was further confirmed later by Rönkkö (2014).

The current consensus is that it is often true that PLS provides parameter estimates for very small samples, whereas other methods may not converge. For example, Henseler et al. (2014 pg. 199) conclude that ”PLS can be applied in many instances of small samples when other methods fail”. But should it? The more important question is not whether parameters are estimated, but whether the estimates are trustworthy. Rigdon (2016 pg. 600) states this succinctly: ”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, Lewis, and Thompson 2012; Reinartz, Haenlein, and Henseler 2009). In fact, Henseler, Hubona, and Ray (2016 pg. 14) unequivocally state that ”PLS does not need fewer observations than other techniques when it comes to inference statistics.”

Given that the asymptotic consistency of the estimator for large samples was noted even in the original works of Wold, one has to wonder why the IS discipline thought that it would be a good idea to use PLS particularly for small sample sizes. Again, this rhetorical question is left 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 the creator of the method pointed out early on. How much damage has been done in these 25 years? IS researchers were one of the earliest adopters of PLS (Ringle, Sarstedt, and Straub 2012) and many of our main theories are built on empirical research that uses PLS. Table 1 lists early, prominent IS theories whose foundational study is based on PLS analysis. Sample sizes in these studies ranged from 52 to 929 and all studies are based 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 the authors of those studies acted on the best available knowledge and advice at the time. While these seminal studies have undergone a rigorous peer-review, that review process was also limited by our early lack of understanding.

In retrospect it becomes apparent that the adoption of PLS in the IS discipline was premature and this now casts significant doubts on the validity of seminal studies. Of course, these theories have been expanded, extended, adopted, or replicated, sometimes with other statistical methods, and that 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 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. Citations
User acceptance	Venkatesh et al. (2003)	28000
Computer self efficacy	Compeau and Higgins (1995)	7000
Technology acceptance	Venkatesh (2000)	6100
Cognitive absorption and IT use	Agarwal and Karahanna (2000)	5000
IT adoption	Karahanna, Straub, and Chervany (1999)	4300
User satisfaction and technology acceptance	Wixom and Todd (2005)	3000
Interorganizational IT adoption	Teo, Wei, and Benbasat (2003)	1800
IT usage	Bhattacherjee and Premkumar (2004)	1800
Computer self efficacy	Thatcher and Perrewé (2002)	800
IT–business linkage	Bassellier and Benbasat (2004)	700

3 Our 2013 Critique

In 2013 we presented a widely noted critical assessment of PLS (Rönkkö and Evermann 2013), which was followed by 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, Edwards, and Antonakis 2014). Our critique focused on six points that we called ”statistical and methodological myths and urban legends”, in the spirit of Vandenberg (2006). These were at the time widely accepted, but not well substantiated – in some cases it was difficult to find any evidence to support the belief and in other cases the belief was rooted on misinterpretation of 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 PLS be used to validate measurement models?
4. Can PLS be used 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 of these in the negative. Much of the discussion centered on the two issues of estimation bias, due to inclusion of measurement error, and on capitalization on chance due to sample error correlations (Rönkkö 2014). This section does not revisit the arguments in their entirety but only briefly presents our main findings and the main arguments of the subsequent exchange with Henseler et al. (2014) and McIntosh, Edwards, and Antonakis (2014).

In our critique, we noted that a large number of 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, McIntosh, and Antonakis 2015). Instead, our claim is that this labeling has misled and continues to mislead researchers to project the abilities of other SEM methods onto PLS (Rönkkö et al. 2016). Instead of 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 is counter to the claims that the weighting mechanism favours more reliable indicators, thus 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 advantages are too small to be relevant, even if ideal weights were known (Rönkkö, McIntosh, and Antonakis 2015; Rönkkö et al. 2016). Moreover, even if indicator weighting can increase reliability, techniques other than PLS, such as maximal reliability composites (Aguirre-Urreta, Rönkkö, and McIntosh 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 was the only term entering the weighting mechanism. We noted that this ”is problematic, because as a result of sampling variation, the errors are never exactly uncorrelated” (Rönkkö and Evermann 2013 pg. 434). For mode B weights, we noted that highly correlated indicators, as typically found in the common factor model, can lead to instability of the regression results. We demonstrated these issues empirically, showing that mode B weights generally lead to much lower reliability of the composites than mode A weights, which in turn are slightly worse than simple summed scales.

The indicator weighting problems are propagated 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 more serious with smaller sample sizes, we termed this behaviour capitalization on chance correlations. This finding was further expanded in (Rönkkö 2014).

Our study was not the first to suggest this idea. In fact, capitalization on chance in PLS was first suspected in the context of interaction analysis by Goodhue, Lewis, and Thompson (2007), who concluded that PLS ’capitalizes on chance’ by taking advantage of PLS’s ability to weight interaction indicators based on their correlation with the dependent variable Y, but hid away this important insight in their online appendix D. 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 reliability of composites depends strongly on sample size, the size of the true parameter, and the loadings of the indicators. While mode B composites are dominated also in their study, mode A composites appear to have a marginal advantage over simple summed scale for large samples ($n = 500$), strong effects ($\beta = 0.5$), or higher loadings (or some combination of these). The sample size effect is expected, as chance error correlations diminish with larger sample size. The higher loading of the most reliable indicator in the extension by Henseler et al. (2014) consequently led to better reliability of the resulting composite, as it is more highly weighted. Nevertheless, in these ideal scenarios, the differences between the techniques were small, in the third decimal (McIntosh, Edwards, and Antonakis 2014) .

In line with the conclusions by Henseler et al. (2014) about the requirement of a ”broader nomological net” (pg. 190), Rigdon (2016 pg. 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 mechanism of the iterative weight computation described in Section 2, this is valid point. It is however unfortunate the PLS ’breaks’ in precisely that condition when it is typically used by applied researchers, i.e. when the very existence of an effect is uncertain and may be small or zero. Indeed, a review by Goodhue, Lewis, and Thompson (2015) concluded that such models are common in IS research and thus a large share of PLS-based IS research may be susceptible to this issue.

3.2 Model Testing

Our 2013 critique also highlighted an issue with model testing. We investigated a set of traditionally recommended model quality metrics such as CR (composite reliability), AVE (average variance extracted), GoF (goodness-of-fit) and relative GoF, as well as the SRMR (standardized root mean squared residual) on their performance in identifying misspecified models. Our findings, in line with earlier work by Evermann and Tate (2010), indicated that none of these metrics can reliably identify misspecified models and accept correctly specified models. As McIntosh, Edwards, and Antonakis (2014 pg. 266) explain, ”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” and that ”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”. (pg. 225) They caution that ”researchers should not interpret measures based on parameter magnitude as indicating model fit, which is the province of the chi-square test and SRMR”. (pg. 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”. (pg. 195) Additionally, they demonstrated that SRMR and what they termed the ”test of exact fit” can indeed distinguish between misspecified and correctly specified models. Section 4.4 discusses this in more detail. Ironically, these tests are largely ignored by IS researchers who tend to favor the CR and AVE statistics despite their demonstrated incapability to differentiate good models from bad ones.

3.3 Parameter Testing

As noted above, we found that with no or very weak effects, estimates of structural parameters are distinctly non-normally distributed. Hence, a key requirement for using the t-test for significance is not satisfied. We concluded that the t statistic and the t-test will be biased and should not be used. We also noted potential issues with bootstrapping the parameter estimates, pointing out that the bootstrapped estimates do not seem to follow the original distribution, thus violating a key assumption of bootstrapping. 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 BCa (bias corrected and accelerated), but not basic bootstrapping, provides acceptable inference. They also showed it to be reasonably robust even in the absence of a true effect. Bootstrap performance improves for larger sample sizes and a more extended latent variable model. They state 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 pg. 198). McIntosh, Edwards, and Antonakis (2014) agree that ”the routine use of the t distribution to test parameters is questionable” (pg. 229) and caution that ”significance tests are meaningful only if the estimates can be trusted to be consistent” (pg. 229). The latter is of course questionable, especially in the case of no effect ($\beta = 0$). Indeed, a recent simulation by Aguirre-Urreta and Rönkkö (2018) demonstrated that while bootstrap-based confidence intervals performed well when testing an existing path, they could not always be trusted in the no-effect scenarios.

3.4 Exploratory Research

Our 2013 critique noted that in light of its shortcomings PLS might not be appropriate for what we understood as exploratory modeling. We argued that if theory is tentative, 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 pg. 200) argued that ”a blind faith in covariance-based SEM’s exploratory capabilities appear both unwise and unjustified”. We concur with this, in that blind faith should never be applied towards any scientific method or theory. As McIntosh, Edwards, and Antonakis (2014 pg. 233) state, ”conventional SEM modification indices (MIs) show suboptimal performance”. However they also note 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”. (pg. 234) Importantly, ”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.”. (pg. 234) In fact, the words ”can” in the prior quote should be replaced 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” (pg. 443). At the same time, we agree with the conclusions by Henseler et al. (2014 pg. 202) who suggest 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, Edwards, and Antonakis (2014 pg. 235) suggest 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". This latter point was also emphasized by Henseler et al. (2014) in their response and has led to important recent developments, discussed in the next section.

4 Recent Developments in PLS

Our 2013 critique has spurred interesting and important developments of the PLS method and its use. Some of the developments were already foreshadowed in the response by Henseler et al. (2014) or developed concurrently, but have seen further development since then.

4.1 Consistent PLS

Consistent PLS (PLSc) was developed by Dijkstra (Dijkstra and Henseler 2015b, 2015a) to correct for the bias of structural regression parameter estimates due to the inclusion of 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 estimate of the composite reliability of a composite variable proxy.

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

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

In a first step, the traditional PLS algorithm is performed. This 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

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

Next, the reliability ρ_A is computed for each latent variable proxy. The corrected latent variable correlations are then estimated using the well-known correction for attenuation introduced by Spearman (1904).

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

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

Initial simulation studies by Dijkstra and Henseler (2015b) showed that PLSc provides estimates close to that of the maximum likelihood estimator, with little bias and comparable precision for the structural parameters but less precision for the item loadings. A later comparison using a real data set confirms that PLSc estimates are close to covariance-based estimates; there, PLSc provided results close to ULS-based covariance estimation (Henseler 2017).

However, not all studies have found similarly encouraging effects. Both Huang (2013) and Rönkkö et al. (2016) find PLSc loading estimates that are less precise and also more biased than traditional ML estimates. They show that PLSc tends to overestimate small correlations and underestimate large correlations. The positive bias may be due to capitalization on chance (Rönkkö et al. 2016) and the negative bias due to overestimating reliability (Aguirre-Urreta, Rönkkö, and McIntosh 2019). PLSc also produces more non-convergent or inadmissible results. Most recently, Yuan, Wen, and Tang (2020 pg. 334) argue that there is no clear advantage of PLSc over using unweighted scales and disattenuating with coefficient (Cronbach's) alpha. Finally, Dijkstra and Henseler (2015a pg. 309) note that "among the consistent techniques, PLSc typically had the lowest statistical power." The same conclusion was drawn in the other studies as well.

PLSc has also been examined on its performance in estimating models with interaction terms by Dijkstra and Schermelleh-Engel (2014) and Becker, Ringle, and Sarstedt (2018). Using a simulation study with sample size $n = 500$, the latter investigated different methods for modeling and estimating the interaction term and also compare PLS to PLSc. They find that the bias correction of PLSc is also evident in the interaction terms. This is as expected, as the interaction term is estimated using an OLS regression based on the product of the bias-corrected main effects composite scores (Henseler and Fassott 2010). Their findings extend to common factor models as well as mixed formative-reflective models.

A further extension to address the bias from ignoring correlated measurement errors by Rademaker, Schuberth, and Dijkstra (2019) demonstrates a small improvement over the original PLSc method, although the authors conclude that ”original PLSc is comparatively robust with respect to misspecification.” (pg. 459) Rademaker, Schuberth, and Dijkstra (2019) suggest that a particular area where handling correlated measurement errors may help is with the product-indicator approach to interaction modeling in PLS (Henseler and Fassott 2010). However, noting the results by Becker, Ringle, and Sarstedt (2018), the PLSc corrected two-step method might be preferable. On the other hand, Becker, Ringle, and Sarstedt (2018) only examined PLSc without the extension by Rademaker, Schuberth, and Dijkstra (2019). Hence, identifying the best way to estimate interaction models in PLS remains an open issue.

While PLSc is a significant improvement over traditional PLS when applied to the common factor model, further simulation research remains to clearly identify the limits of its applicability and robustness to a range of different conditions. In particularly, given the most recent findings by Rönkkö et al. (2016) and Yuan, Wen, and Tang (2020), its advantage over the simpler approach of using indicator means and disattenuating the correlations with coefficient alpha or CR (Devlieger and Rosseel 2017; Devlieger, Mayer, and Rosseel 2015; Rosseel 2020) requires further investigation.

RECOMMENDATION 1

If a researcher uses PLS for factor models (reflective), consistent PLS with measurement error correction should be used.

4.2 t-tests are inappropriate

Our earlier critique (Rönkkö and Evermann 2013) 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 that percentile, bias corrected (BC) and bias corrected and accelerated (BCa) bootstrapping are strongly preferred over normal-distribution based bootstrap. This is not surprising, as Rönkkö and Evermann (2013) noted that the underlying parameter estimates are not normally distributed. While 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, percentile intervals are slightly less efficient but are conservative and should therefore be preferred. Aguirre-Urreta and Rönkkö (2018 pg. 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”. The implications of their results is that researchers should use conservative *percentile* bootstrapping of PLSc estimates, rather than relying on BC/BCa based bootstrapping; the latter was recommended by Henseler et al. (2014) and is the default in many PLS software tools.

RECOMMENDATION 2

If a researcher uses PLS, t-tests should be avoided; percentile bootstrapping should be used for statistical inference.

Aguirre-Urreta and Rönkkö (2018) also show that all bootstrap methods struggle when the latent variables are weakly or not at all correlated, especially for models with a small number or sparsely connected latent variables. Thus, the problem of weakly connected latent variables, which leads to a bimodal parameter estimate distribution and inappropriateness of the t-test, is not entirely addressed. The problem is of course that parameter significance tests are most useful when the researcher is uncertain that there is a relationship at all. Aguirre-Urreta and Rönkkö (2018) also recommend diagnostics such as inspecting the histogram of bootstrap replications, available in most modern PLS software, for a severely non-normal shape. As a remedy for potential problems, they suggest using equal weights to calculate the problematic composites.

RECOMMENDATION 3

If a researcher uses bootstrapping for significance testing, the bootstrap estimates should be examined for severely non-normal distribution.

4.3 Power estimation using simulation

Recent work by Aguirre-Urreta and Rönkkö (2015) discusses how to select the appropriate sample size for a PLS study. Sample size is a particularly important aspect of using PLS as it affects the consistency of the estimator, its convergence behaviour, and the confidence intervals for parameter significance testing. Examining the latter, Aguirre-Urreta and Rönkkö (2015) provide a guide to using simulation studies for power calculations.

In general, traditional rules-of-thumbs like ten times the largest number of indicators on a construct (Chin 1998), or using regression based power calculations, e.g. the popular GPower software, are inappropriate. Aguirre-Urreta and Rönkkö (2015) discuss the questionable origin of the ”ten-times” rule at length and note that regression-based power computations assume perfectly reliable regressors and regressands. They are therefore incompatible with models that use multiple indicators that are assumed to contain measurement error.

Aguirre-Urreta and Rönkkö (2015) provide a step-by-step guide using a simple script for the R statistical system. One caveat is that the method as reported still uses uncorrected mode A weights by default and the t-test for significance testing in the power calculations, which should be avoided and replaced with percentile confidence intervals, as noted in the previous section. Recent versions of the `matrixpls` package for the R system, on which the tutorial is built, offer the choice of percentile confidence intervals (using the `citype="perc"` parameter for the `matrixpls.sim` function) and the use of PLSc (using the `disattenuate=TRUE` and `parametersReflective=estimator.plscLoadings` parameters for the `matrixpls.sim` function). Usage of `matrixpls` has changed since the publication of Aguirre-Urreta and Rönkkö (2015), but the user manual (Rönkkö 2020) provides an updated version of their example.

Given that simulation studies for PLS are easy to perform with modern PLS software tools, researchers should explicitly hypothesize about a range of likely effect sizes and data distribution assumptions, set these up for a simulation study, and identify the statistical power for various sample sizes *before* collecting data.

RECOMMENDATION 4

If a researcher uses PLS, a simulation study should be used to a-priori determine appropriate sample size and study power.

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 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 the model is misspecified. Needless to say, simulation studies confirmed that these indices are unable to detect misspecified, i.e. non-fitting, models (Evermann and Tate 2010; Rönkkö and Evermann 2013; Henseler and Sarstedt 2013; Henseler et al. 2014).

Model testing adequacy can be demonstrated 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, it is easy to come up with models that are surely misspecified, e.g. by combining two factors into one, assigning indicators to obviously incorrect factors, or adding more latent variables to the model. If a model that is surely incorrect is not rejected by the model quality indices used, this is strong evidence against these indices. This kind of analysis is demonstrated by Rönkkö, Parkkila, and Ylitalo (2012), who evaluated various incorrectly specified models based on a number of models published in leading IS journals. Their conclusion was not encouraging: "Based on the analysis of the goodness of fit indices we can conclude that [...] any model fits approximately as well as the original model".

RECOMMENDATION 5

If using PLS-based indices for model testing or fit evaluation, test or evaluate also models that are assumed to be incorrect for the data to ensure that the model quality indices can actually lead to the rejection of incorrect models.

Early in the development of PLS, Lohmöller (1989) proposed fit indices and reliability indices, some of them based on the residual covariances and on the maximum-likelihood χ^2 fit function. However, these were never picked up by the early PLS literature. Only later were these ideas further developed by Henseler et al. (2014) and Dijkstra and Henseler (2015b) who propose three ways to evaluate the overall model fit by comparing the model-implied to the observed covariance matrix, a technique adopted from covariance-based SEM. They define three distance metrics as follows.

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

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

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

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

As McIntosh, Edwards, and Antonakis (2014) point 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. The interpretation of SRMR is the same as in covariance-based SEM, where values < 0.08 are assumed to indicate acceptable model fit (Hu and Bentler 1999), although there is continuing debate around the appropriateness and usefulness of fit indices in general and simple cutoff rules-of-thumb in particular (Kline 2011, chap. 8).

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

RECOMMENDATION 6

If a researcher uses PLS, adequate model fit should be established using d_G prior to interpreting path estimates or assessing factorial structure.

Model selection is related to identifying misspecified models and assessing model fit. It is used to identify the best model of a set of equally well-fitting, competing models. Model selection should never select illfitting over well-fitting models. In covariance-based SEM, model selection can be done with nested models using the χ^2 difference, providing a strong and well-understood statistical test. When models are not nested, 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 have recently also been applied to PLS estimated models (Sharma et al. 2019; Sharma et al. 2019; Danks, Sharma, and Sarstedt 2020). These criteria penalize models for complexity to achieve a balance between fit and parsimony. As these criteria are based on the likelihood of a model, they are natural in covariance-based SEM, which uses a closed form of the likelihood function in its estimation procedure. 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. 2019 pg. 351). Note that this is a significant distributional assumption for PLS, which is typically presented as a distribution-agnostic or assumption-free method.

A simulation study by Sharma et al. (2019) showed that AIC, BIC and GM are significantly better at successfully identifying the correct model than previous ”quality criteria” such as GoF, R^2 and Q^2 metrics, confirming earlier results noted above: ”Overall, none of the PLS criteria performed satisfactorily” (pg. 356). Sharma et al. (2019) find that BIC and GM provide the best success rates in identifying the correct model. Their findings hold also for relatively small samples of 100, and are robust across sample sizes with performance improving with increasing sample size. Similarly, their findings hold also for small effect sizes and performance improves with increasing effect size. They recommend that BIC or GW should replace traditional model ”quality criteria”. While Sharma et al. (2019) used the lowest information criterion value for model selection, Danks, Sharma, and Sarstedt (2020) uses 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” (pg. 13).

Note that model selection is a comparative method while identifying model misfit is based on a single model and the observed data. Thus, while Sharma et al. (2019) and Danks, Sharma, and Sarstedt (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 is not a substitute for identifying model misfit.

Recommendation 7

If a researcher uses PLS to compare multiple, well-fitting models, model selection may be done using information-theoretic criteria. Good model fit must be established and demonstrated before any comparison.

4.5 R^2 does not measure predictive power

PLS is frequently argued to be particularly useful for prediction, rather than explanation, though few researchers appear to follow through with this (Evermann and Tate 2016; Rönkkö et al. 2016). Shmueli et al. (2016) point out two key differences between explanation and prediction. At the data level, prediction is concerned with predicting *new* values for specific observations (”out-of-sample prediction”) whereas explanation is concerned with analyzing patterns in the existing values. At the application level, prediction is concerned with making *individual-level* decisions, while explanation focuses on average or population level decisions. Metrics such as R^2 or Q^2 are not measuring 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) present a method to produce out-of-sample, individual-level predictions from PLS. Adopting the concept of ”operative prediction” from Lohmöller (1989), Shmueli et al. (2016 pg. 4554) describe prediction from manifest variables to manifest variables which 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.”

Evermann and Tate (2016) compare the predictive performance of PLS, covariance-based estimation, and an a-theoretic prediction method (EM and linear regression, operating 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, their first study on purely reflective models necessarily uses the Q^2 statistics with blindfolding to evaluate prediction performance. Their study also examined PLSc, but found that traditional PLS with mode A performs better than PLSc. Their second study examined mixed formative-reflective models of the type shown in Figure 2. Here, it is possible to identify observed exogenous predictor variables and observed endogenous predicted variables. Evermann and Tate (2016) use RMSE (root mean squared error) 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.

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

In summary, the point here is not to argue for or against PLS for predictive studies over any other technique (and there are a multitude of others, many of them a-theoretic ones rooted in machine learning, such as support vector machines or deep neural networks), but to remind researchers that simply stating that a study is predictive in nature does not in fact make it predictive. Evermann and Tate (2016 pg. 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 theory-testing that requires explanatory models. If predictive modeling is relevant, researchers should clearly explain the purpose and usefulness of the predictions and use operative prediction (i.e. mixed formative-reflective) to assess out-of-sample prediction of their models at the individual level, as recommended by Shmueli et al. (2016) and Evermann and Tate (2016).

RECOMMENDATION 8

If a researcher uses PLS for prediction, the study 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 of AVE and CR cannot identify model misspecification. Their argument focused on cross-loadings, which affect discriminant validity metrics. Their results were confirmed by Henseler et al. (2014) and are the impetus behind the development of a new technique for assessing discriminant validity. Taking the multitrait-multimethod matrix (MTMM) as a as starting point, Henseler, Ringle, and Sarstedt (2015) derive the HTMT, the hetero-trait-mono-trait (HTMT) criterion for assessing discriminant validity. They propose two hard cutoff criteria but also recommend bootstrapping to test whether the HTMT is less than one. Henseler, Ringle, and Sarstedt (2015) show that the HTMT is related to coefficient alpha and a recent study by Rönkkö and Cho (2020) proves that HTMT is in fact a correlation between unweighted composites that is disattenuated with standardized alpha.

In a simulation study, Henseler, Ringle, and Sarstedt (2015) replicate our poor results for the AVE and CR criteria and show that the HTMT method is able to detect low discriminant validity. However, a more comprehensive simulation study by Rönkkö and Cho (2020) indicates that these results reflect the problems of the AVE criterion more than any capability of HTMT. They further show that the HTMT method makes more assumptions than traditional confirmatory factor analysis. It provides little advantage when these assumptions hold and performs less well in other conditions (Rönkkö and Cho 2020).

Because HTMT is a disattenuated correlation, its use 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 is likely to confuse researchers and readers alike; for most researchers, simply estimating a traditional factor analysis might be an easier and more effective alternative.

RECOMMENDATION 9

Do not use the AVE and CR based criteria to establish discriminant validity. The HTMT can be used if its assumptions can be justified. Otherwise traditional factor analysis techniques should be preferred.

4.7 PLS is intended for composite models

In light of the increasingly critical evaluations of PLS (Chin 1998; Chumney 2013; Goodhue, Lewis, and Thompson 2012; Reinartz, Haenlein, and Henseler 2009), PLS proponents could no longer simply ignore or sweep away the issue of biased estimates of PLS for the common factor model. Prior to development of PLSc, Rigdon (2012) argued for a refocusing of 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". (pg. 342) Rigdon (2016) suggests that PLS is misunderstood and misapplied to estimate the common factor model. A similar point is made by Sarstedt et al. (2016) who also argue for a focus on composite models. The common argument is that PLS should not be used to estimate models for which the data generating process was the common factor model and that any evaluation of PLS on these terms is therefore misguided. This may well be true, but at the same time, the common factor model is still the one most frequently estimated by applied IS researchers using PLS (Ringle, Sarstedt, and Straub 2012), and underpins prominent IS theories (Table 1).

The article by Rigdon (2012) attracted a number of responses, many of which rejected the proposed turning 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 like PLS. In PLSe1, the PLS estimates are used as starting values in a conventional CBSEM estimation, and in PLSe2 the PLS estimates are used to build a weight matrix for GLS estimation. While it is unclear what advantage such a parameterization of the GLS estimator would provide, these techniques have nevertheless been promoted in the recent literature (Ghasemy, Jamil, and Gaskin 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, Edwards, and Antonakis (2014) pointed out that the model was neither a composite model nor statistically identified, it was later refined by Schuberth, Henseler, and Dijkstra (2018). They show that the composite model can be rewritten 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, Henseler, and Dijkstra (2018) propose a model fit test using the SRMR, d_{LS} and d_G metrics (Section 3.2). Their simulation study shows that d_G in particular can correctly identify misspecifications of the composite model, with an error rate restricted to the theoretical expectation at the 0.05 and 0.1 significance levels. They also note that sample sizes of more than 350 are required to consistently detect misspecifications. However, Schuberth, Henseler, and Dijkstra (2018) did not estimate their model using PLS but used GCCA (generalized canonical correlation analysis). Whether PLS weights work as well remains to be demonstrated.

There are two arguments for composite models in the literature. First, a misspecified latent variable model can produce severely biased estimates and though biased, estimates from composite models are less bad (Rhemtulla, Bork, and Borsboom 2020). While this is true, a better course of action is to avoid misspecifications in the first place by engaging in proper diagnostics of the model. Second, PLS is presented as an ideal technique when the population follows a composite model instead of a common factor model. Unfortunately, the literature has yet to provide a compelling argument for why and when one should expect the phenomenon of interest to follow a composite model. While analyzing the merits of this argument would be important, it is better left for another paper and we refrain from making a recommendation on the use of PLS at this time. Nevertheless, adopting such a perspective would require the development of an entirely different measurement theory, as well as abandoning Cronbach's alpha, CR, and AVE as reliability measures.

4.8 PLS and dependent formative models

Recent work by Aguirre-Urreta and Marakas (2014) examined the way PLS is used 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 is not what is actually estimated. They argue that the statistical model that should be inferred from the path diagram in Figure 4 based on SEM path diagram rules is

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

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

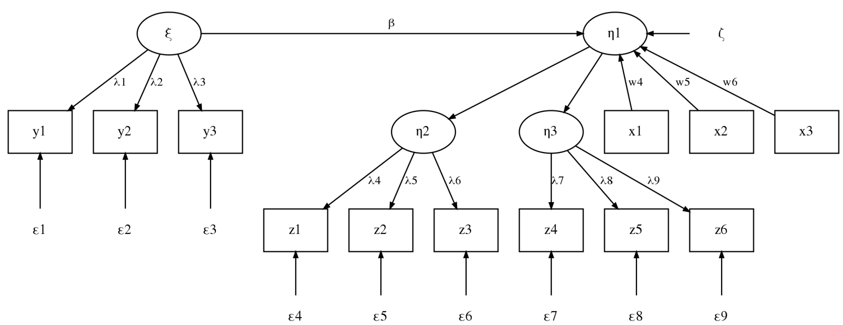

Figure 4. Formative dependent latent variable (η_1) (correlations among the x indicators are omitted for clarity)

In response, Rigdon et al. (2014) note that the graphical notation may be misleading. They point 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 the following:

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

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) have shown 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) suggest 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. This point had already been made by Cadogan and Lee (2013), who argue convincingly that effects on composites must necessarily be mediated by the components, essentially endorsing the variable-level specification by Rigdon et al. (2014).

To avoid the confusion showcased by this debate, researchers must be very clear about their assumptions and the precise statistical model they intend to estimate. While the graphical user interfaces of modern PLS software make it very easy to perform a PLS analysis, this simplicity comes with the downside that the actual equations that form the statistical model are not obvious. Moreover, commonly used PLS software provide very little documentation (SmartPLS 2020) on how path diagrams are translated to equations, causing considerable ambiguity on what exactly was estimated when publications present the results only in a graphical format.

RECOMMENDATION 10

If a researcher uses PLS with dependent formative constructs, researchers should verify and publish the statistical model (i.e. the equations) that is assumed and estimated to avoid any confusion.

Given that the effects from a latent variable to a formative latent variable are mediated by the formative indicators, estimating only the total effect rather than the individual mediator effects is a wasted research opportunity. This would suggest using the latent variable specification (Cadogan and Lee 2013; Rigdon et al. 2014); comparable results can also be obtained 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

PLS is traditionally implemented as a two-phase method with ordinary least squares (OLS) regression for 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 the system of regression equations to be recursive, i.e. there must be no cycles in the structural part of the PLS path model.

The first restriction is often violated in practice, leading to "endogeneity" (Antonakis et al. 2010). A typical cause of endogeneity are omitted variables, i.e. variables that affect both the regressor and regressand. By omitting these 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, OLS is used in the second phase of PLS for convenience and out of habit, not by necessity. Dijkstra and Henseler (2015a) use 2SLS (two-stage-least-squares) regression for estimating the structural parameters from the bias-adjusted correlation or covariance matrix in PLSc, "whether recursively or with feedback patterns" (pg. 14), thereby relaxing the second restriction in PLS, and allowing the modeler to describe cyclical structural models. 2SLS is also useful for addressing endogeneity. Benitez-Amado, Henseler, and Roldán (2016) and Hult et al. (2018) recommend 2SLS in order to use the instrumental variable approach for addressing endogeneity, provided the researcher has suitable instrumental variables ("instruments"). An instrument is a variable that is strongly correlated with the problematic endogenous regressor, but independent of the error term in the regression model.

There are two limitations with the approach by Hult et al. (2018). First, the authors recommend the usual Hausman test for identifying endogeneity and the Sargan test for identifying the suitability of the instrument. Both tests are parametric tests that make distributional assumptions. This runs counter to the typical aim of keeping PLS a distribution free method. Moreover, these techniques do not take into account the unknown effect of the sampling distribution of PLS. Hult et al. (2018) recommend investigating the use of bootstrapping in future research. A second issue, not addressed by Hult et al. (2018), is that the first stage of PLS, the computation of composite scores, also relies on OLS and may be subject to endogeneity. Whether and to what extent this has any impact on composites and hence parameter estimates, must be investigated further.

RECOMMENDATION 11

If a researcher uses PLS and endogeneity is suspected, it should be addressed using 2SLS.

RECOMMENDATION 12

If a researcher uses PLS and a non-recursive structural model, 2SLS should be used instead of OLS.

5 Alternatives to PLS

When first developed, Herman Wold positioned PLS as an alternative to the nascent covariance-based analysis by Karl Jöreskog (Wold 1982). Given the computing abilities at the time, Wold's PLS was faster and this allowed for an interactive exploration of data. However, it was clear that, given its estimation bias and consistency-at-large property, PLS 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, covariance-based SEM is often viewed as the primary alternative to PLS (e.g. Reinartz, Haenlein, and Henseler 2009; Goodhue, Lewis, and Thompson 2012; Sarstedt et al. 2016; Rigdon, Sarstedt, and Ringle 2017; Hair, Sarstedt, and Ringle 2019).

The explicit acceptance of the composite nature of PLS and its positioning as primarily a composite analysis tool make it clear that covariance-based SEM is not in fact 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 and Ylitalo 2010; Rönkkö and Evermann 2013; Rönkkö 2014). The key consideration in composite modeling is the reliability of the composites, i.e. how well the estimated composite scores correlate with the true scores. While it is possible to construct scenarios where the weighting algorithm of PLS outperforms unweighted summed scales by a significant amount, this is by no means the typical case. There are more cases where summed scales provides composites as reliable or more so than PLS weighted composites. In many cases, the advantage of PLS is a small one and, given the limitations and cautions around PLS in the previous section, it is not clear that the use of PLS is an acceptable trade-off. When indicators are highly correlated, little reliability can be gained by differential item weighting (Rönkkö, McIntosh, and Antonakis 2015; Rönkkö et al. 2016). In fact, Rönkkö, Evermann, and Aguirre-Urreta (2016) find that indicator weighting provides little advantage even at a low mean indicator correlation of 0.4, which is well-below a 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) find 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, also concluding that ”PLS indicator weights do not generally provide a meaningful improvement in reliability” (pg. 6). Rönkkö et al. (2021) analyzed two datasets that are 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. This one outlier correlation was explained by one very poorly performing item. After eliminating that item, as would be done in the factor analysis stage of scale development (Hair et al. 2010, Chapter 3), the correlation between the corresponding PLS composite exceeded 0.99 in line with the other correlations in that study.

A comparison of simple summed scale with PLS composites is easy to implement by exporting PLS composite scores and examining their correlation with the summed item scores. Rönkkö et al. (2021) formalized this comparison as the CEI (composite equivalence index) statistic. This analysis should be adopted as a basic diagnostic for any PLS analysis: If there are no meaningful differences, it is difficult to justify the use of PLS composites over simpler unweighted composites, given the challenges of PLS, e.g. in terms of the potential for capitalization on chance and parameters testing (Section 4.2). If there are differences, they should be explained because differing weights can indicate a problem such as a cross-loading (Rönkkö et al. 2021)

RECOMMENDATION 13

PLS composites should be compared to unweighted composites to demonstrate any possible advantage that the PLS composites might have.

Generalized structure component analysis (GSCA) is another composite analysis method, developed by Hwang and Takane (2004, 2014) explicitly as an alternative to PLS. It has an explicit objective function (squares of the measurement errors and structural residuals) that is optimized to estimate model parameters. In contrast, PLS has no explicit optimizing criterion and its frequently claimed ”optimality” of weights is not clearly and explicitly defined (Rönkkö, McIntosh, and Antonakis 2015; Rönkkö et al. 2016). Recent advances in GSCA include regularization (Hwang 2009), latent interactions (Hwang, Ho, and Lee 2010), multilevel analysis (Hwang, Takane, and Malhotra 2007), and non-linear models (Hwang and Takane 2009). A bias correction inspired consistent PLS has been implemented as GSCA_M (Hwang, Takane, and Jung 2017). Curiously, GSCA has found little adoption in IS research, despite being available in a number of software implementations, such as an R package, an Excel add-on and a web-based graphical implementation.

The low adoption of GSCA and focus on PLS by IS researchers is all the more puzzling as in many situations GSCA has been shown 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, concluding that ”all methods yield to [sic] comparable components” (pg. 15) and ”when the blocks are good, the computation of the components does not depend upon the method used.” (pg. 15; i.e. for unidimensional and strongly correlated indicators). Here too the simple method of unweighted summed scales proved to be as good as complex weighting methods. Focusing on parameter recovery in misspecified models, Hwang, Malhotra, et al. (2010) also claim an advantage of GSCA over PLS and conclude that ”we recommend the adoption of generalized structured component analysis as a sensible alternative to partial least squares ... [it] performed better than or as well as partial least squares in parameter recovery.” (pg. 710) Comparing the bias corrected GSCA_M to consistent PLS, Hwang, Takane, and Jung (2017) show that, with equal weights, GSCA_M biases parameter estimates positively, while the PLSc bias is negative. Again, standard errors are consistently smaller for GSCA_M than PLSc. For heterogeneous weights, GSCA_M and PLSc are biased in the same direction for loading estimates but in opposite directions for structural path estimates (negative for GSCA_M, positive for PLSc).

A systematic simulation study (Hair, Hult, et al. 2017) showed that estimation errors of GSCA for measurement model parameters are smaller than of PLS, independent of samples size, number of indicators and indicator weights. For equal weights, summed scales outperformed both GSCA and PLS, independent of sample sizes and number of indicators. Here, the complexities of PLS and GSCA are actually detrimental, leading 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. A recent working paper (Hwang, Cho, et al. 2020) combines GSCA and GSCA_M for mixed composite and factor models (”integrated GSCA”, ”IGSCA”). Their first simulation study shows that, as theoretically expected, GSCA and PLS ”provided positively-biased estimates of the factor loadings” (pg. 20) while ”PLSc and IGSCA were the only approaches that provided unbiased estimates of all the loadings and path coefficients” (pg. 20). Their second simulation study with a wider range of conditions pitted PLSc against IGSCA. Hwang, Cho, et al. (2020 pg. 27) conclude 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 and Hwang 2020). Among a GSCA application (Jung et al. 2020) and a concept analysis of the literature (Hwang, Sarstedt, et al. 2020), Cho and Choi (2020) present a simulation study comparing the parameter recovery performance and statistical power of GSCA and PLS. Their results mirror those by Hair, Hult, et al. (2017), in that GSCA outperforms PLS in recovering measurement model parameters while they are evenly matched for recovery of structural parameters.

In addition to summed scales and GSCA, regularized generalized canonical correlation analysis (RGCCA) (Tenenhaus and Tenenhaus 2011, 2014; Tenenhaus, Tenenhaus, and Groenen 2017) is another alternative to PLS. It builds on elements of both generalized canonical correlation analysis (GCCA) and PLS. While formally well developed, little is known about its comparative performance as there is a lack of simulation studies that compare RGCCA to GSCA and PLS.

The point of this section is not to evaluate alternative component-based statistical models in detail. Instead, we wish to make IS researchers aware of these alternatives. The choice is not one between covariance-based SEM and PLS. Instead, unweighted summed scales should be the primary alternative to PLS, performing as well or better than complex indicator weighting schemes in many situations (Tenenhaus 2008; Hair, Hult, et al. 2017). 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 can be violated, and are typically better understood.

RECOMMENDATION 14

Faced with multiple, equivalent methodological options, the simplest method should be preferred.

6 Conclusions

Overall, there has been a wealth of methodological developments on PLS that were at least partially driven by our 2013 critique (Rönkkö and Evermann 2013). Many of these strive to address the concerns 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 be trying 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, this is best left as a rhetorical question for the reader to ponder.

Some PLS researchers advocate a retreat to composite models, acknowledging that while advances are being made on PLS, other methods are better suited for the common factor model. Yet, there is considerable debate about the use of formative and composite models, their place in applied research, and their ontology (Edwards and Bagozzi 2000; Howell, Breivik, and Wilcox 2007b, 2007a; Bagozzi 2007; Bollen 2007; Wilcox, Howell, and Breivik 2008; Marakas, Johnson, and Clay 2007, 2008; Hardin, Chang, and Fuller 2008; MacKenzie, Podsakoff, and Podsakoff 2011; Edwards 2011; Markus 2018). Moreover, while important contributions to composite modelling have been made, there is still a paucity of simulation research to indicate that PLS is indeed the best tool for this type of model with the kinds of dataset that IS researchers use. As noted earlier, GSCA appears to be superior to PLS in many composite modeling applications, and so are 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 are clearly causally dependent on the measured psychological construct, leading to a common factor model. While it may be possible to reframe these theories using composites, those composites would have very different meaning than the latent variables that reflect psychological constructs (Evermann and Tate 2012).

Prediction from PLS estimated models has been shown to be superior to prediction from covariance-based SEM in many situations. However, prediction from PLS or otherwise estimated models is limited by the restrictions that the models impose. The linearity of the models, the absence of certain paths in the model because they are not theoretically motivated, or the imposition of composite or factor structures are constraints that necessarily reduce the predictive performance of such models in practical applications compared to a-theoretical prediction techniques such as neural networks or random forests. If the aims is to predict individual values for a particular application, users of predictive models may be unwilling to accept lower predictive performance when all that is gained is a theoretically plausible, linear model. This trade-off between prediction and explanation is specific to every application, but is unlikely to always go well for PLS (or any other linear structural equation modelling method).

In summary, PLS finds itself between more than one proverbial rock and 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, it is outperformed by summed scales in some applications and GSCA in the remaining situations. For predictive modeling, it is threatened by a range of modern a-theoretical prediction methods, such as deep learning networks, which outperform traditional statistical learning techniques in a wide variety of applications.

Additionally, PLS is naturally limited to modeling composites and cannot model truly latent variables that have no associated indicators. Examples of the latter type of models are latent growth curves models (Bollen and Curran 2006) and latent difference score models (McArdle 2009; Grimm et al. 2012). While the idea is that the composites can approximate latent variables in some models, the use of models with truly latent variables, i.e. without directly associated observed variables, is increasing in many research disciplines, including in Information Systems (e.g. Serva, Kher & Laurenceau 2011; Bala & Venkatesh 2013).

On a positive note, PLS researchers and critics agree on many issues. Sample size requirements are universally acknowledged as being an invalid reason to choose PLS. There is universal acknowledgment that parameter estimates in the original PLS algorithm are biased when applied to the common factor model. There is agreement that, as Rigdon (2016 pg. 602) points out, weakly correlated or uncorrelated constructs are ”outside of its [PLS] boundary conditions”, though little attention has been paid to this requirement in IS applications. There is agreement that bootstrap CIs should be preferred over t-tests, although this too is rarely done in IS research.

This paper aims to contribute to minimum methodological competence standards. We opened our article by asking what might such standards look like? To give an example, one would hope to not see a paper that uncritically uses traditional mode A for a common factor model, has a sample size of 30 while measuring 5 constructs using 20 indicators, employs t-tests to test a very weak effect size for significance and uses CR and AVE criteria to justify model fit. On the other hand, a paper that uses PLSc, had done a-priori power analyses to arrive at a sample size of 500, demonstrates good model fit using SRMR and d_G, and uses 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 a predictive study, but then proceed with a common factor model, to evaluate model fit and factorial structure, and report the R^2 values for the dependent latent variables. It would be more convincing for that researcher 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 MSE (mean squared error) 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, it is insufficient simply to declare a construct to be a composite; this choice should be theoretically justified, and the manifest variables should be demonstrably useful in creating a meaningful composite. Subsequent evaluation can then not be based on traditional factor analytic criteria either.

In summary, it is good to see that our 2013 critique spurred many constructive developments in PLS and the way it is used. One driver for the present paper is the fact that some manuscripts still cite papers on PLS that are 25 or 30 years old with their authors apparently unaware of both the criticisms and the recent advances. This paper is intended to help IS researcher remain current with important methodological developments and to provide a sound foundation upon which to base their choice of statistical methods.

References

Agarwal, Ritu, and Elena Karahanna. 2000. “Time Flies When You’re Having Fun: Cognitive Absorption and Beliefs About Information Technology Usage.” *MIS Quarterly* 24 (4): 665–94.

Aguirre-Urreta, Miguel I, and George M Marakas. 2008. “The Use of PLS When Analyzing Formative Constructs: Theoretical Analysis and Results from Simulations.” In *Proceedings of the International Conference on Information Systems, ICIS 2008, Paris, France, December 14-17, 2008*, 185.

Aguirre-Urreta, Miguel I, and George M Marakas. 2014. “Research Note — Partial Least Squares and Models with Formatively Specified Endogenous Constructs: A Cautionary Note.” *Information Systems Research* 25 (4): 761–78.

Aguirre-Urreta, Miguel I, and Mikko Rönkkö. 2018. “Statistical Inference with PLSc Using Bootstrap Confidence Intervals.” *MIS Quarterly* 42 (3): 1001–20.

Aguirre-Urreta, Miguel I., and Mikko Rönkkö. 2015. “Sample Size Determination and Statistical Power Analysis in PLS Using R: An Annotated Tutorial.” *Communications of the Association for Information Systems* 36 (1). http://aisel.aisnet.org/cais/vol36/iss1/3.

Aguirre-Urreta, Miguel I., Mikko Rönkkö, and Cameron N. McIntosh. 2019. “A Cautionary Note on the Finite Sample Behavior of Maximal Reliability.” *Psychological Methods* 24 (2): 236–52. https://doi.org/10.1037/met0000176.

Ali, Faizan, S Mostafa Rasoolimanesh, Marko Sarstedt, Christian M Ringle, and Kisang Ryu. 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): 514–38.

Antonakis, John, Samuel Bendahan, Philippe Jacquart, and Rafael Lalive. 2010. “On Making Causal Claims: A Review and Recommendations.” *The Leadership Quarterly* 21 (6): 1086–1120.

Bagozzi, RP. 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): 229–37.

Bala, H., & Venkatesh, V. (2013). Changes in Employees' Job Characteristics During an Enterprise System Implementation: A Latent Growth Modeling Perspective. *MIS Quarterly* 37(4), 1113-1140.

Bassellier, Genevieve, and Izak Benbasat. 2004. “Business Competence of Information Technology Professionals: Conceptual Development and Influence on It-Business Partnerships.” *MIS Quarterly* 28 (4): 673–94.

Becker, Jan-Michael, Christian M Ringle, and Marko Sarstedt. 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, Jose, Jörg Henseler, and José Luis Roldán. 2016. “How to Address Endogeneity in Partial Least Squares Path Modeling.” In *22nd Americas Conference on Information Systems, AMCIS 2016, San Diego, ca, Usa, August 11-14, 2016*. Association for Information Systems.

Bentler, Peter M, and Wenjing Huang. 2014. “On Components, Latent Variables, PLS and Simple Methods: Reactions to Rigdon’s Rethinking of PLS.” *Long Range Planning* 47 (3): 138–45.

Bhattacherjee, A., and G Premkumar. 2004. “Understanding Changes in Belief and Attitude Toward Information Technology Usage: A Theoretical Model and Longitudinal Test.” *MIS Quarterly* 28 (2): 229–54.

Bollen, KA. 2007. “Interpretational Confounding Is Due to Misspecification, Not to Type of Indicator: Comment on Howell, Breivik, and Wilcox (2007).” *Psychological Methods* 12 (2): 219.

Bollen, KA, and PJ Curran. 2006. *Latent Growth Curve Models: A Structural Equation Approach*. Hoboken, NJ: Wiley.

Cadogan, John W, and Nick Lee. 2013. “Improper Use of Endogenous Formative Variables.” *Journal of Business Research* 66 (2): 233–41.

Chin, Wynne W. 1998. “The Partial Least Squares Approach to Structural Equation Modeling.” In *Modern Methods for Business Research*, edited by George A. Marcoulides, 295–336. Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Chin, Wynne W, and Abhijit Gopal. 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, Wynne W., Barbara L. Marcolin, and Peter R. Newsted. 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, Cleveland, Ohio, Usa, December 16-18, 1996*, 6.

Chin, Wynne W, Barbara L Marcolin, and Peter R Newsted. 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, Wynne W., and Peter R. Newsted. 1999. “Structural Equation Modeling Analysis with Small Samples Using Partial Least Squares.” In *Statistical Strategies for Small Sample Research*, edited by Rick H. Hoyle, 307–42. Thousand Oaks, CA: Sage Publications.

Cho, Gyeongcheol, and Ji Yeh Choi. 2020. “An Empirical Comparison of Generalized Structured Component Analysis and Partial Least Squares Path Modeling Under Variance-Based Structural Equation Models.” *Behaviormetrika* 47 (1): 243–72.

Chumney, Frances. 2013. “Structural Equation Models with Small Samples: A Comparative Study of Four Approaches.” Doctoral dissertation, Lincoln, NE: University of Nebraska-Lincoln. http://digitalcommons.unl.edu/cehsdiss/189.

Compeau, Deborah, and Christopher A. Higgins. 1995. “Computer Self-Efficacy: Development of a Measure and Initial Test.” *MIS Quarterly* 19 (2): 189–211.

Danks, Nicholas P, Pratyush N Sharma, and Marko Sarstedt. 2020. “Model Selection Uncertainty and Multimodel Inference in Partial Least Squares Structural Equation Modeling (PLS-SEM).” *Journal of Business Research* 113: 13–24.

Devlieger, Ines, Axel Mayer, and Yves Rosseel. 2015. “Hypothesis Testing Using Factor Score Regression A Comparison of Four Methods.” *Educational and Psychological Measurement*, September, 0013164415607618. https://doi.org/10.1177/0013164415607618.

Devlieger, Ines, and Yves Rosseel. 2017. “Factor Score Path Analysis.” *Methodology* 13 (Supplement 1): 31–38. https://doi.org/10.1027/1614-2241/a000130.

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

Dijkstra, Theo K. 2014. “PLS’ Janus Face – Response to Professor Rigdon’s ‘Rethinking Partial Least Squares Modeling: In Praise of Simple Methods’.” *Long Range Planning*, Rethinking Partial Least Squares Path Modeling: Looking Back and Moving Forward, 47 (3): 146–53. https://doi.org/10.1016/j.lrp.2014.02.004.

Dijkstra, Theo K., and Jörg Henseler. 2015a. “Consistent Partial Least Squares Path Modeling.” *MIS Quarterly* 39 (2): 297–316.

Dijkstra, Theo K., and Jörg Henseler. 2015b. “Consistent and Asymptotically Normal PLS Estimators for Linear Structural Equations.” *Computational Statistics & Data Analysis* 81 (January): 10–23. https://doi.org/10.1016/j.csda.2014.07.008.

Dijkstra, Theo K., and Karin Schermelleh-Engel. 2014. “Consistent Partial Least Squares for Nonlinear Structural Equation Models.” *Psychometrika* 79 (4): 585–604. https://doi.org/10.1007/s11336-013-9370-0.

Edwards, Jeffrey R. 2011. “The Fallacy of Formative Measurement.” *Organizational Research Methods* 14 (2): 370–88.

Edwards, Jeffrey R, and Richard P Bagozzi. 2000. “On the Nature and Direction of Relationships Between Constructs and Measures.” *Psychological Methods* 5 (2): 155.

Evermann, Joerg, and Mary Tate. 2010. “Testing Models or Fitting Models? Identifying Model Misspecification in PLS.” In *Proceedings of the International Conference on Information Systems, ICIS 2010, Saint Louis, Missouri, Usa, December 12-15, 2010*, edited by Rajiv Sabherwal and Mary Sumner, 21. Association for Information Systems.

Evermann, Joerg, and Mary Tate. 2012. “An Ontology of Structural Equation Models with Application to Computer Self-Efficacy.” In *Proceedings of the International Conference on Information Systems, Icis 2012, Orlando, Florida, Usa, December 16-19, 2012*.

Evermann, Joerg, and Mary Tate. 2014. “Bayesian Structural Equation Models for Cumulative Theory Building in Information Systems – a Brief Tutorial Using BUGS and R.” *Communications of the AIS* 34: 76. http://aisel.aisnet.org/cais/vol34/iss1/76.

Evermann, Joerg, and Mary Tate. 2016. “Assessing the Predictive Performance of Structural Equation Model Estimators.” *Journal of Business Research* 69 (10): 4565–82.

Gefen, David, E. E Rigdon, and Detmar W. Straub. 2011. “An Update and Extension to SEM Guidelines for Administrative and Social Science Research.” *MIS Quarterly* 35 (2): iii–xiv.

Ghasemy, Majid, Hazri Jamil, and James E. Gaskin. 2020. “Have Your Cake and Eat It Too: PLSe2 = ML + PLS.” *Quality & Quantity*, July. https://doi.org/10.1007/s11135-020-01013-6.

Goodhue, Dale L., William Lewis, and Ron Thompson. 2012. “Does PLS Have Advantages for Small Sample Size or Non-Normal Data.” *MIS Quarterly* 36 (3): 981–1001.

Goodhue, Dale L., William Lewis, and Ron Thompson. 2015. “PLS Pluses and Minuses in Path Estimation Accuracy.” In *AMCIS 2015 Proceedings*. San Juan, Puerto Rico. http://aisel.aisnet.org/amcis2015/ISPhil/GeneralPresentations/3.

Goodhue, Dale L., W. Lewis, and R. Thompson. 2007. “Statistical Power in Analyzing Interaction Effects: Questioning the Advantage of PLS with Product Indicators.” *Information Systems Research* 18 (2): 211–27. https://doi.org/10.1287/isre.1070.0123.

Grimm, Kevin J, Yang An, John J McArdle, Alan B Zonderman, and Susan M Resnick. 2012. “Recent Changes Leading to Subsequent Changes: Extensions of Multivariate Latent Difference Score Models.” *Structural Equation Modeling: A Multidisciplinary Journal* 19 (2): 268–92.

Guyon, Hervé. 2018. “The Fallacy of the Theoretical Meaning of Formative Constructs.” *Frontiers in Psychology* 9. https://doi.org/10.3389/fpsyg.2018.00179.

Hair, Joseph F, William C Black, Barry J Babin, and Rolph E Anderson. (2010). Multivariate data analysis. 7th ed. Pearson Education.

Hair, Joe F, Christian M Ringle, and Marko Sarstedt. 2011. “PLS-SEM: Indeed a Silver Bullet.” *Journal of Marketing Theory and Practice* 19 (2): 139–52.

Hair, Joe F, Marko Sarstedt, Christian M Ringle, and Jeannette A Mena. 2012. “An Assessment of the Use of Partial Least Squares Structural Equation Modeling in Marketing Research.” *Journal of the Academy of Marketing Science* 40 (3): 414–33.

Hair, Joe, Carole L Hollingsworth, Adriane B Randolph, and Alain Yee Loong Chong. 2017. “An Updated and Expanded Assessment of PLS-SEM in Information Systems Research.” *Industrial Management & Data Systems* 117 (3): 442–58.

Hair, Joseph F, G Tomas M Hult, Christian M Ringle, Marko Sarstedt, and Kai Oliver Thiele. 2017. “Mirror, Mirror on the Wall: A Comparative Evaluation of Composite-Based Structural Equation Modeling Methods.” *Journal of the Academy of Marketing Science* 45 (5): 616–32.

Hair, Joseph F, Christian M Ringle, and Marko Sarstedt. 2013. “Partial Least Squares Structural Equation Modeling: Rigorous Applications, Better Results and Higher Acceptance.” *Long Range Planning* 46 (1-2): 1–12.

Hair, Joseph F, Marko Sarstedt, Torsten M Pieper, and Christian M Ringle. 2012. “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): 320–40.

Hair, Joseph F, Marko Sarstedt, and Christian M Ringle. 2019. "Rethinking Some of the Rethinking of Partial Least Squares." *European Journal of Marketing* 53 (4): 566–84.

Hardin, Andrew. 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. https://doi.org/10.1037/met0000115.

Hardin, Andrew M, Jerry Cha-Jan Chang, and Mark A Fuller. 2008. "Formative Vs. Reflective Measurement: Comment on Marakas, Johnson, and Clay (2007)." *Journal of the Association for Information Systems* 9 (9): 3.

Henseler, Jörg. 2017. "Bridging Design and Behavioral Research with Variance-Based Structural Equation Modeling." *Journal of Advertising* 46 (1): 178–92.

Henseler, Jörg, Theo K. Dijkstra, Marko Sarstedt, Christian M. Ringle, Adamantios Diamantopoulos, Detmar W. Straub, David J. Ketchen, Joseph F. Hair, G. Tomas M. Hult, and Roger J. Calantone. 2014. "Common Beliefs and Reality About PLS: Comments on Rönkkö and Evermann (2013)." *Organizational Research Methods* 17 (2): 182–209. https://doi.org/10.1177/1094428114526928.

Henseler, Jörg, and Georg Fassott. 2010. "Testing Moderating Effects in PLS Path Models: An Illustration of Available Procedures." In *Handbook of Partial Least Squares*, edited by V. Esposito Vinzi, W. W. Chin, J. Henseler, and H. Wang, 713–35. Springer.

Henseler, Jörg, Geoffrey Hubona, and Pauline Ash Ray. 2016. "Using PLS Path Modeling in New Technology Research: Updated Guidelines." *Industrial Management & Data Systems* 116 (1): 2–20. https://doi.org/10.1108/IMDS-09-2015-0382.

Henseler, Jörg, Christian M Ringle, and Marko Sarstedt. 2015. "A New Criterion for Assessing Discriminant Validity in Variance-Based Structural Equation Modeling." *Journal of the Academy of Marketing Science* 43 (1): 115–35.

Henseler, Jörg, Christian M Ringle, and Rudolf R Sinkovics. 2009. "The Use of Partial Least Squares Path Modeling in International Marketing." In *New Challenges to International Marketing*, 277–319. Emerald Group Publishing Limited.

Henseler, Jörg, and Marko Sarstedt. 2013. "Goodness-of-Fit Indices for Partial Least Squares Path Modeling." *Computational Statistics* 28 (2): 565–80.

Howell, Roy D, Einar Breivik, and James B Wilcox. 2007a. "Is Formative Measurement Really Measurement? Reply to Bollen (2007) and Bagozzi (2007)." *Psychological Methods* 12 (2): 238–45.

Howell, Roy D, Einar Breivik, and James B Wilcox. 2007b. "Reconsidering Formative Measurement." *Psychological Methods* 12 (2): 205.

Hu, Li-tze, and Peter M Bentler. 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, Wenjing. 2013. "PLSe: Efficient Estimators and Tests for Partial Least Squares." Doctoral dissertation, Los Angeles, CA: University of California. http://escholarship.org/uc/item/2cs2g2b0.

Hui, B. S., and Herman Wold. 1982. "Consistency and Consistency at Large of Partial Least Squares Estimates." In *Systems Under Indirect Observation: Causality, Structure, Prediction*, edited by K Jöreskog and Herman Wold. Amsterdam: North-Holland.

Hulland, John. 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 Tomas M, Joseph F Hair Jr, Dorian Proksch, Marko Sarstedt, Andreas Pinkwart, and Christian M Ringle. 2018. "Addressing Endogeneity in International Marketing Applications of Partial Least Squares Structural Equation Modeling." *Journal of International Marketing* 26 (3): 1–21.

Hwang, Heungsun. 2009. "Regularized Generalized Structured Component Analysis." *Psychometrika* 74 (3): 517–30.

Hwang, Heungsun, Gyeongcheol Cho, Kwanghee Jung, Carl Falk, Jessica Flake, and Min Jin. 2020. “An Approach to Structural Equation Modeling with Both Factors and Components: Integrated Generalized Structured Component Analysis.”

Hwang, Heungsun, Moon-Ho Ringo Ho, and Jonathan Lee. 2010. “Generalized Structured Component Analysis with Latent Interactions.” *Psychometrika* 75 (2): 228–42.

Hwang, Heungsun, Naresh K Malhotra, Youngchan Kim, Marc A Tomiuk, and Sungjin Hong. 2010. “A Comparative Study on Parameter Recovery of Three Approaches to Structural Equation Modeling.” *Journal of Marketing Research* 47 (4): 699–712.

Hwang, Heungsun, Marko Sarstedt, Jun Hwa Cheah, and Christian M Ringle. 2020. “A Concept Analysis of Methodological Research on Composite-Based Structural Equation Modeling: Bridging PLSPM and GSCA.” *Behaviormetrika* 47 (1): 219–41.

Hwang, Heungsun, and Yoshio Takane. 2004. “Generalized Structured Component Analysis.” *Psychometrika* 69 (1): 81–99.

Hwang, Heungsun, and Yoshio Takane. 2009. “Nonlinear Generalized Structured Component Analysis.” *Behaviormetrika* 37 (1): 1–14.

Hwang, Heungsun, and Yoshio Takane. 2014. *Generalized Structured Component Analysis: A Component-Based Approach to Structural Equation Modeling*. CRC Press.

Hwang, Heungsun, Yoshio Takane, and Kwanghee Jung. 2017. “Generalized Structured Component Analysis with Uniqueness Terms for Accommodating Measurement Error.” *Frontiers in Psychology* 8: 2137.

Hwang, Heungsun, Yoshio Takane, and Naresh Malhotra. 2007. “Multilevel Generalized Structured Component Analysis.” *Behaviormetrika* 34 (2): 95–109.

Jöreskog, Karl. G., and Herman Wold. 1982. “The ML and PLS Techniques for Modeling with Latent Variables.” In *Systems Under Indirect Observation: Causality, Structure, Prediction*, edited by Karl G. Jöreskog and Herman Wold. Amsterdam: North-Holland.

Jung, Kwanghee, Sang Soo Cho, Jaehoon Lee, Seungman Kim, and Ji Hoon Ryoo. 2020. “An Illustrative Application of Generalized Structured Component Analysis for Brain Connectivity Research.” *Behaviormetrika* 47 (1): 273–89.

Karahanna, Elena, Detmar W. Straub, and Norman L. Chervany. 1999. “Information Technology Adoption Across Time: A Cross-Sectional Comparison of Pre-Adoption and Post-Adoption Beliefs.” *MIS Quarterly* 23 (2): 183–213.

Kline, Rex B. 2011. *Principles and Practice of Structural Equation Modeling*. 3rd ed. New York, NY: Guilford Press.

Liengaard, Benjamin Dybro, Pratyush Nidhi Sharma, G Tomas M Hult, Morten Berg Jensen, Marko Sarstedt, Joseph F Hair, and Christian M Ringle. 2020. “Prediction: Coveted, yet Forsaken? Introducing a Cross-Validated Predictive Ability Test in Partial Least Squares Path Modeling.” *Decision Sciences*.

Lohmöller, Jan-Bernd. 1989. *Latent Variable Path Modeling with Partial Least Squares*. Heidelberg: Physica-Verlag.

MacKenzie, Scott B, Philip M Podsakoff, and Nathan P Podsakoff. 2011. “Construct Measurement and Validation Procedures in Mis and Behavioral Research: Integrating New and Existing Techniques.” *MIS Quarterly* 35 (2): 293–334.

Marakas, George, Richard Johnson, and Paul F Clay. 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, George, Richard D Johnson, and Paul F Clay. 2008. “Formative Vs. Reflective Measurement: A Reply to Hardin, Chang, and Fuller.” *Journal of the Association for Information Systems* 9 (9): 2.

Markus, Keith A. 2018. “Three Conceptual Impediments to Developing Scale Theory for Formative Scales.” *Methodology* 14 (4): 156–64. https://doi.org/10.1027/1614-2241/a000154.

McArdle, John J. 2009. “Latent Variable Modeling of Differences and Changes with Longitudinal Data.” *Annual Review of Psychology* 60: 577–605.

McIntosh, Cameron N., Jeffrey R. Edwards, and John Antonakis. 2014. “Reflections on Partial Least Squares Path Modeling.” *Organizational Research Methods* 17 (2): 210–51. https://doi.org/10.1177/1094428114529165.

Peng, David Xiaosong, and Fujun Lai. 2012. “Using Partial Least Squares in Operations Management Research: A Practical Guideline and Summary of Past Research.” *Journal of Operations Management* 30 (6): 467–80.

Petter, Stacie. 2018. “"Haters Gonna Hate": PLS and Information Systems Research.” *ACM SIGMIS Database: The DATABASE for Advances in Information Systems* 49 (2): 10–13.

Rademaker, Manuel E., Florian Schuberth, and Theo K. Dijkstra. 2019. “Measurement Error Correlation Within Blocks of Indicators in Consistent Partial Least Squares.” *Internet Research* 29 (3): 448–63.

Reinartz, Werner, Michael Haenlein, and Jörg Henseler. 2009. “An Empirical Comparison of the Efficacy of Covariance-Based and Variance-Based SEM.” *International Journal of Research in Marketing* 26 (4): 332–44.

Rhemtulla, Mijke, Riet van Bork, and Denny Borsboom. 2020. “Worse Than Measurement Error: Consequences of Inappropriate Latent Variable Measurement Models.” *Psychological Methods* 25 (1): 30–45. https://doi.org/10.1037/met0000220.

Rigdon, Edward E. 2016. “Choosing PLS Path Modeling as Analytical Method in European Management Research: A Realist Perspective.” *European Management Journal* 34 (6): 598–605.

Rigdon, Edward E. 2012. “Rethinking Partial Least Squares Path Modeling: In Praise of Simple Methods.” *Long Range Planning* 45 (5–6): 341–58. https://doi.org/10.1016/j.lrp.2012.09.010.

Rigdon, Edward E, Jan-Michael Becker, Arun Rai, Christian M Ringle, Adamantios Diamantopoulos, Elena Karahanna, Detmar W Straub, and Theo K Dijkstra. 2014. “Conflating Antecedents and Formative Indicators: A Comment on Aguirre-Urreta and Marakas.” *Information Systems Research* 25 (4): 780–84.

Rigdon, Edward E, Marko Sarstedt, and Christian M Ringle. 2017. “On Comparing Results from CB-SEM and PLS-SEM: Five Perspectives and Five Recommendations.” *Marketing Zfp* 39 (3): 4–16.

Ringle, Christian M, Marko Sarstedt, Rebecca Mitchell, and Siegfried P Gudergan. 2020. “Partial Least Squares Structural Equation Modeling in HRM Research.” *The International Journal of Human Resource Management* 31 (12): 1617–43.

Ringle, Christian M., Marko Sarstedt, and Detmar W. Straub. 2012. “A Critical Look at the Use of PLS-SEM in MIS Quarterly.” *MIS Quarterly* 36 (1): iiv–8.

Rosseel, Yves. 2020. “Small Sample Solutions for Structural Equation Modeling.” In *Small Sample Size Solutions: A Guide for Applied Researchers and Practitioners*, edited by Rens van de Schoot and Milica Miočević, 226–38. Taylor & Francis. https://doi.org/10.4324/9780429273872-19.

Rouse, Anne, and Brian Corbitt. 2008. “There’s SEM and ”SEM”: A Critique of the Use of PLS Regression in Information Systems Research.” *ACIS 2008 Proceedings*, 81.

Rönkkö, Mikko. 2014. “The Effects of Chance Correlations on Partial Least Squares Path Modeling.” *Organizational Research Methods* 17 (2): 164–81.

Rönkkö, Mikko. 2020. “Introduction to Matrixpls.”

Rönkkö, Mikko, et al. (2021) “What does PLS do and is it of any use? A practical demonstration of the critical problems with PLS.” Working paper.

Rönkkö, Mikko, and Eunseong Cho. 2020. “An Updated Guideline for Assessing Discriminant Validity.” *Organizational Research Methods* forthcoming.

Rönkkö, Mikko, and Joerg Evermann. 2013. “A Critical Examination of Common Beliefs About Partial Least Squares Path Modeling.” *Organizational Research Methods* 16 (3): 425–48. https://doi.org/10.1177/1094428112474693.

Rönkkö, Mikko, Joerg Evermann, and Miguel I. Aguirre-Urreta. 2016. “Estimating Formative Measurement Models in IS Research: Analysis of the Past and Recommendations for the Future.” *Unpublished Working Paper*.

Rönkkö, Mikko, Cameron N McIntosh, Miguel I Aguirre-Urreta, and others. 2016. “Improvements to PLSc: Remaining Problems and Simple Solutions.”

Rönkkö, Mikko, Cameron N. McIntosh, and John Antonakis. 2015. “On the Adoption of Partial Least Squares in Psychological Research: Caveat Emptor.” *Personality and Individual Differences* 87 (December): 76–84. https://doi.org/10.1016/j.paid.2015.07.019.

Rönkkö, Mikko, Cameron N. McIntosh, John Antonakis, and Jeffrey R. Edwards. 2016. “Partial Least Squares Path Modeling: Time for Some Serious Second Thoughts.” *Journal of Operations Management* 47–48 (1): 9–27. https://doi.org/10.1016/j.jom.2016.05.002.

Rönkkö, Mikko, Kaisa Parkkila, and Jukka Ylitalo. 2012. “Use of Partial Least Squares as a Theory Testing Tool – an Analysis of Information Systems Papers.” In *ECIS 2012 Proceedings*. http://aisel.aisnet.org/ecis2012/145.

Rönkkö, Mikko, and Jukka Ylitalo. 2010. “Construct Validity in Partial Least Squares Path Modeling.” In *Proceedings of the International Conference on Information Systems, ICIS 2010, Saint Louis, Missouri, Usa, December 12-15, 2010*, edited by Rajiv Sabherwal and Mary Sumner, 155. Association for Information Systems.

Sarstedt, Marko, Joseph F Hair, Christian M Ringle, Kai O Thiele, and Siegfried P Gudergan. 2016. “Estimation Issues with PLS and CBSEM: Where the Bias Lies!” *Journal of Business Research* 69 (10): 3998–4010.

Sarstedt, Marko, and Heungsun Hwang. 2020. “Advances in Composite-Based Structural Equation Modeling.” *Behaviormetrika* 47: 213–17.

Sarstedt, Marko, Christian M Ringle, Jun-Hwa Cheah, Hiram Ting, Ovidiu I Moisescu, and Lacramioara Radomir. 2020. “Structural Model Robustness Checks in Pls-Sem.” *Tourism Economics* 26 (4): 531–54.

Schuberth, Florian, Jörg Henseler, and Theo K Dijkstra. 2018. “Confirmatory Composite Analysis.” *Frontiers in Psychology* 9: 2541.

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, pp-pp. https://doi.org/10.17705/1CAIS.02814

Sharma, Pratyush Nidhi, Galit Shmueli, Marko Sarstedt, Nicholas Danks, and Soumya Ray. 2019. “Prediction-Oriented Model Selection in Partial Least Squares Path Modeling.” *Decision Sciences*.

Sharma, Pratyush, Marko Sarstedt, Galit Shmueli, Kevin H Kim, and Kai Oliver Thiele. 2019. “PLS-Based Model Selection: The Role of Alternative Explanations in Information Systems Research.” *Journal of the Association for Information Systems* 20 (4): 4.

Shmueli, Galit, Soumya Ray, Juan Manuel Velasquez Estrada, and Suneel Babu Chatla. 2016. “The Elephant in the Room: Predictive Performance of PLS Models.” *Journal of Business Research* 69 (10): 4552–64.

SmartPLS. 2020 “SmartPLS – Frequently Asked Questions.” retrieved from https://www.smartpls.com/faq/documentation/documentation-or-manual. Last access December 22, 2020.

Spearman, C. 1904. “The Proof and Measurement of Association Between Two Things.” *The American Journal of Psychology* 15 (1): 72–101. http://www.jstor.org/stable/1412159.

Tenenhaus, Arthur, and Michel Tenenhaus. 2011. “Regularized Generalized Canonical Correlation Analysis.” *Psychometrika* 76 (2): 257.

Tenenhaus, Arthur, and Michel Tenenhaus. 2014. “Regularized Generalized Canonical Correlation Analysis for Multiblock or Multigroup Data Analysis.” *European Journal of Operational Research* 238 (2): 391–403.

Tenenhaus, Michel. 2008. “Component-Based Structural Equation Modelling.” *Total Quality Management* 19 (7-8): 871–86.

Tenenhaus, Michel, Silvano Amato, and Vincenzo Esposito Vinzi. 2004. “A Global Goodness-of-Fit Index for PLS Structural Equation Modelling.” In *Proceedings of the Xlii Sis Scientific Meeting*, 1:739–42. 2.

Tenenhaus, Michel, Arthur Tenenhaus, and Patrick JF Groenen. 2017. “Regularized Generalized Canonical Correlation Analysis: A Framework for Sequential Multiblock Component Methods.” *Psychometrika* 82 (3): 737–77.

Teo, Hock-Hai, Kwok Kee Wei, and Izak Benbasat. 2003. “Predicting Intention to Adopt Interorganizational Linkages: An Institutional Perspective.” *MIS Quarterly* 27 (1): 19–49.

Thatcher, Jason Bennett, and Pamela L. Perrewé. 2002. “An Empirical Examination of Individual Traits as Antecedents to Computer Anxiety and Computer Self-Efficacy.” *MIS Quarterly* 26 (4): 381–96.

Valle, Patrícia Oom do, and Guy Assaker. 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, Viswanath. 2000. “Determinants of Perceived Ease of Use: Integrating Control, Intrinsic Motivation, and Emotion into the Technology Acceptance Model.” *Information Systems Research* 11 (4): 342–65.

Venkatesh, Viswanath, Michael G. Morris, Gordon B. Davis, and Fred D. Davis. 2003. “User Acceptance of Information Technology: Toward a Unified View.” *MIS Quarterly* 27 (3): 425–78.

Wilcox, James B, Roy D Howell, and Einar Breivik. 2008. “Questions About Formative Measurement.” *Journal of Business Research* 61 (12): 1219–28.

Wixom, Barbara H, and Peter A Todd. 2005. “A Theoretical Integration of User Satisfaction and Technology Acceptance.” *Information Systems Research* 16 (1): 85–102.

Wold, Herman. 1982. “Soft Modeling: The Basic Design and Some Extensions.” In *Systems Under Indirect Observation: Causality, Structure, Prediction*, edited by K. G. Jöreskog and Svante Wold, 1–54. Amsterdam: North-Holland.

Wold, Herman. 1985. “Systems Analysis by Partial Least Squares.” In *Measuring the Unmeasurable*, edited by P. Nijkamp, L Leitner, and N Wrigley, 221–52. Dordrecht, Germany: Marinus Nijhoff.

Yuan, Ke-Hai, Yong Wen, and Jiashan Tang. 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): 333–50. https://doi.org/10.1080/10705511.2019.1647107.

About the Authors

Joerg Evermann. Joerg 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. Mikko 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.