

JYX



This is a self-archived version of an original article. This version may differ from the original in pagination and typographic details.

Author(s): Rönkkö, Mikko; Lee, Nick; Evermann, Joerg; McIntosh, Cameron; Antonakis, John

Title: Rejoinder : fractures in the edifice of PLS

Year: 2023

Version: Published version

Copyright: © Mikko Rönkkö, Nick Lee, Joerg Evermann, Cameron McIntosh and John Antonak

Rights: CC BY 4.0

Rights url: <https://creativecommons.org/licenses/by/4.0/>

Please cite the original version:

Rönkkö, M., Lee, N., Evermann, J., McIntosh, C., & Antonakis, J. (2023). Rejoinder : fractures in the edifice of PLS. *European Journal of Marketing*, 57(6), 1626-1640.

<https://doi.org/10.1108/ejm-07-2022-0508>

Rejoinder: fractures in the edifice of PLS

Rejoinder

Mikko Rönkkö

*Jyväskylä University School of Business and Economics, Jyväskylä University,
Jyväskylä, Finland*

Nick Lee

Department of Marketing, Warwick Business School, Coventry, UK

Joerg Evermann

*Department of Information Systems, Memorial University of Newfoundland,
Saint John's, Canada*

Cameron McIntosh

*Strategic Planning and Methodology Division,
Employment and Social Development Canada Gatineau, Gatineau, Canada, and*

John Antonakis

*Department of Organizational Behavior, Faculty of Business and Economics,
University of Lausanne, Lausanne, Switzerland*

Received 8 July 2022
Revised 26 September 2022
3 March 2023
Accepted 10 March 2023

Abstract

Purpose – This study aims to provide a response to the commentary by Yuan on the paper “Marketing or Methodology” in this issue of EJM.

Design/methodology/approach – Conceptual argument and statistical discussion.

Findings – The authors find that some of Yuan’s arguments are incorrect, or unclear. Further, rather than contradicting the authors’ conclusions, the material provided by Yuan in his commentary actually provides additional reasons to avoid partial least squares (PLS) in marketing research. As such, Yuan’s commentary is best understood as additional evidence speaking against the use of PLS in real-world research.

Research limitations/implications – This rejoinder, coupled with Yuan’s comment, continues to support the strong implication that researchers should avoid using PLS in marketing and related research.

Practical implications – Marketing researchers should avoid using PLS in their work.

Originality/value – This rejoinder supports the earlier conclusions of “Marketing or Methodology,” with additional argumentation and evidence.

Keywords PLS, Composites, Structural equation models, Methodology, Measurement, Partial least squares

Paper type Conceptual paper

© Mikko Rönkkö, Nick Lee, Joerg Evermann, Cameron McIntosh and John Antonakis. Published by Emerald Publishing Limited. This article is published under the Creative Commons Attribution (CC BY 4.0) licence. Anyone may reproduce, distribute, translate and create derivative works of this article (for both commercial and non-commercial purposes), subject to full attribution to the original publication and authors. The full terms of this licence may be seen at <http://creativecommons.org/licenses/by/4.0/legalcode>

As an editorial item, this article has not been subject to double blind peer review.



Before addressing Yuan's commentary, we start with a few general points. The single most important point in our paper *Marketing or Methodology: Exposing the Fallacies of PLS with Simple Demonstrations (MoM)* is to some extent not specifically about partial least squares (PLS) at all: *The first decision a researcher must make when selecting an analysis method for multi-item scale data is whether they are going to work with composites [1] or latent variables in their model.* Once this decision is made, one can select between the various estimation methods available for the chosen task. If this process is followed, the false equivalence that is drawn in much pro-PLS literature between PLS and ML-SEM vanishes. If researchers made explicit their decision to use composites or latent variables and justified that decision clearly, many of the problems that are so evident in existing research practice that we point out in MoM would be far less prevalent.

To be very clear, PLS is *not* a method to directly estimate latent variable models in the way that ML-SEM or factor analysis is. It is a method to *construct composites*, and decisions to use PLS should be made among the different composite methods, *after* deciding to use composites rather than latent variables in the model. As we make clear in MoM, composites have their uses. However, to have a viable place in a researcher's toolkit, PLS must have *useful advantages over other composite methods*, rather than be compared against latent variable modeling methods like factor analysis.

Unfortunately, it appears that many researchers do not understand that composites and latent variables are not interchangeable in models, and that there are important implications of using one or the other. We do not wish to enter in any depth the debate about the conceptualization of constructs here (although see Lee and Cadogan's two papers in this special issue, and associated commentaries, for a more focused treatment). However, it stands to reason that if a theory includes concepts that are characterized as latent (i.e. not directly observable; as are many in marketing and related fields), then latent variable methods such as ML-SEM, or common factor analysis, should be the first choice of operationalization. Such methods are not directly conceptually interchangeable with composite methods, and therefore, if one wishes to use a composite method in place of a latent variable method, the choice should be justified (perhaps through the need for computational simplicity in parameter estimation). Once the reason for *this* decision to use composites is established, one must justify the decision to use any given composite method, of which there are many available. Almost no existing literature using PLS provides any kind of justification for using composites, and simply uses PLS in the place of ML-SEM, presuming modeled concepts and estimators are interchangeable. This is simply not the case.

Even if the choice to use composites is well-justified, one must still justify which method is to be used to create those composites. In MoM, we showed clearly that claims of advantages of PLS over other composite methods either:

- were not based on any evidence;
- were based on invalid evidence or incorrect interpretation of evidence; and
- were evident only to a trivial degree and/or in highly unrealistic settings.

Further, we showed that even if PLS *did* have the claimed advantages, they were heavily outweighed by the clear and well-established drawbacks of PLS. [Table 1](#) summarizes the claims that we made about PLS in MoM, regarding both its advantages and disadvantages, and whether they are supported by evidence.

To make things crystal clear for readers, there is nothing in the comments of Yuan that convincingly rebuts any of our points (although there are points certainly worthy of

Point ^a	Additional evidence in existing literature ^b	Counter-evidence in existing literature ^b	Conclusion
PLS does not maximize R^2 or explained variance	Simulation evidence in Rönkkö (2020)	<i>None.</i> Existing PLS literature does not provide evidence to support this claim Indeed, it is simple to show many methods that can outperform PLS on any specific criteria of maximization	PLS does not maximize R^2 or explained variance. The claim itself makes little sense and no supporting proofs exist
Improving reliability by differential indicator weighting is not a reason to use PLS	Simulation evidence in Rönkkö and Ylitalo (2010) , Rönkkö and Evermann (2013) and Rönkkö et al. (2016)	<i>None under likely real-world analysis conditions</i>	It is unclear why a researcher should favor a method, which shows a trivial reliability improvement only in situations of very low inter-item correlation, and at the expense of proven serious drawbacks. Standard scale development procedures recommend against items with low inter-correlations, and where they <i>should</i> be included (e.g. formative indices); internal consistency is irrelevant
PLS weights bias composite correlations:	The simulations in Henseler et al. (2014) also show that in most situations, PLS leads to a loss of reliability Decades of evidence show that differential indicator weights generally provide only trivial advantages at best	PLS may offer small reliability improvements in simulation studies that are designed with conditions ideally favorable to PLS, such as extremely low inter-item correlations: e.g. Simulations in Henseler et al. (2014) show a <1% improvement in reliability for situations expressly designed to favor PLS	It is impossible for researchers to know composites are weakly correlated <i>a priori</i> That PLS is not robust to departures from this assumption should be pointed out in PLS introductory texts
a) If scales are weakly correlated	Simulations by Goodhue et al. (2015) , Rönkkö (2014) and Rönkkö and Evermann (2013)	<i>None.</i> Rigdon (2016) claims weakly correlating composites are a known violation of PLS assumptions However, this is clearly not a well-known violation of specified PLS prerequisites, as we are not aware of any published guidelines in primer or introductory PLS literature that state this should be tested	
b) Where there are cross-loadings or correlated errors between items in different scales			
c) Particularly when sample size is small			
AVE and CR should never be used with PLS/PLS should not be used to validate measures	Simulations by Evermann and Tate (2010) , Rönkkö and Evermann (2013) , Rönkkö and Cho (2022)	<i>None.</i> HTMT has been proposed as an improvement, but it is not a PLS-specific method, and CFA works better more generally. Evidence, which suggests HTMT generally outperforms CFA (e.g., Voorhees et al., 2016), is based on incorrect use of CFA (Rönkkö and Cho, 2022)	HTMT is a better method than using AVE with PLS. However, HTMT is not a PLS method

(continued)

Table 1.
Technical points in MoM and additional points made in this rejoinder

Point ^a	Additional evidence in existing literature ^b	Counter-evidence in existing literature ^b	Conclusion
	These results are corroborated even by PLS advocates' research (Henseler <i>et al.</i> , 2014; McIntosh <i>et al.</i> , 2014)		PLS introductory texts should remove mention of AVE and CR as measure validation and model assessment tools. Factor analysis should be used to test the assumptions of HTMT
<i>Additional point not in MoM:</i> The bootstrap "sign-change" options in PLS programs can produce 100% false positive rate	Simulation evidence in Rönkkö <i>et al.</i> (2015)	<i>None.</i> In fact, Henseler <i>et al.</i> (2016) recommend abandoning the sign-change corrections	Unfortunately, the damage to statistical decision-making has likely already been done, and is perhaps still continuing. The sign-change corrections only serve to increase false positive rate and the use of this feature should be discontinued

Notes: ^aAll points made in MoM are supported by numerical illustrations, to prioritize understandability for non-methodological readers. ^bSources of evidence and counter evidence are considered in terms of a hierarchy of strength. While we recognize that for different purposes, different forms of evidence are more or less appropriate, the strength of evidence for or against the sort of claims we make in MoM can be ascertained according to the following hierarchy: the strongest evidence is a mathematical proof, followed by appropriate simulations, followed by numerical illustrations (e.g. using real data). Rhetoric alone is not considered to be evidence for or against these claims, and hence, we do not include sources that only rely on rhetoric here. We also make the conceptual point that PLS is not a latent variable method at all despite referred to as such in the literature. In fact, some of the current PLS literature argues that PLS is not intended to estimate common factor-based population models and is in fact most suitable for examining "composite-based population models" (Dijkstra, 2017; Hair and Sarstedt, 2019; Sarstedt *et al.*, 2016). Yet, as we show in MoM, in research practice PLS is nearly exclusively used to examine factor-model based conceptualizations. Indeed, it is clear that recent PLS work still claims that PLS can estimate reflective models (Schubert, 2021), and even the most recent edition of Hair *et al.*'s (2021), PLS primer text clearly indicates that PLS can handle reflective models, which from a measurement theory perspective are essentially equivalent to factor models, and certainly are not composite models (Markus and Borsboom, 2013)

Source: Authors thinking specifically for this paper

Table 1.

discussion). Yuan's comment (Y21) contains copious statistical detail, which certainly *looks* impressive. However, none of the points made by Yuan invalidate the points made in MoM (or even appear that they are intended to do so), or provides any strong evidence to support the continued use of PLS. Indeed, as we will show, Yuan's results can be most correctly understood as speaking *against* the use of PLS in typical marketing (and related) research studies.

We will next discuss Yuan's commentary, which diverges significantly from our key points but still contains a lot of material that needs to be addressed. We finish with a general summary and set of conclusions for how best to move forward. We show again that there is no reason at all to use PLS, given the numerous superior alternatives already available.

Response to Yuan's commentary

We were surprised and intrigued to see that Yuan had written a commentary on MoM, and naturally wondered why he chose now to enter a discussion about PLS. Y21 contains several interesting points, most of which are drawn from his recent work in the area, some of which was unpublished at the time we wrote MoM. However, it is not correct for Y21 to claim that we were unaware of his results, given that we did not include them in MoM. We did not include Yuan's work in our original paper because it was not relevant to the points that we wanted to make. Similarly, Yuan's current comment does not offer any evidence to discount

the points made in MoM, and some of his arguments are simply incorrect. Yuan's paper does bring up some broader issues, which can be discussed, but they are tangential to our more specific points. Still, below, we attempt to group Yuan's main points into a set of overarching themes of relevance to our main points and respond to Y21 in relation to them.

Theme 1: optimality

Y21 focuses much of his comments on various ways to expound on the "optimality" properties of PLS. This issue is only peripherally related to our core points (see [Table 1](#)). According to Y21's derivations, when the full population is analyzed and a factor model holds for the data, the indicator weights under PLS Mode B are equivalent to those in the formulation of the Bartlett factor score. These findings are not entirely new to the PLS literature but are a welcome formalization of previous results ([Schubert et al., 2022](#)).

Y21 also makes two entirely new claims: That:

- (1) PLS Mode B is equivalent to "the normal-distribution-based maximum likelihood (ML) estimator/predictor of the latent trait" ([Yuan and Deng, 2021](#)); and
- (2) the composite following PLS-SEM Mode B enjoys the optimal statistical properties of an ML estimator (see e.g. [Casella and Berger, 2001](#)).

However, to the best of our knowledge at least, the first claim is not in fact made in [Yuan and Deng \(2021\)](#), at least not explicitly. We also went through [Casella and Berger \(2002\)](#) and did not find clear support for the second part of the claim there. Specifically, [Casella and Berger \(2002\)](#) present consistency, efficiency and asymptotic normality as desirable properties of estimators (Chapter 10) and make it clear that efficiency and normality do not follow from consistency (p. 473) but need to be proven separately. Indeed, there many estimators that are consistent but inefficient. While there is little evidence to support Y21's claims, the first claim may be correct in the sense that once we know the factor score weights, those weights can be used to calculate optimal predictions for individual observations. However, the claim does not mean that PLS Mode B would be an optimal way to calculate the weights themselves beyond being consistent. For example, efficiency would still need to be proven.

Leaving aside the fact that very few applications of PLS in marketing and related literature use either Bartlett scores or PLS Mode B, Y21's points – even if correct – do not demonstrate any substantive advantages of PLS. Indeed, Y21 acknowledge that PLS is not an ideal technique for estimating Bartlett scores from sample data. Specifically, although both ML-based and PLS Mode B Bartlett scores may be biased by measurement error, there are many other reasons to prefer ML-based Bartlett scores, if one is to use them. Importantly, PLS Mode B Bartlett scores are susceptible to bias caused by capitalization on chance, as established over a decade ago by [Rönkkö and Ylitalo \(2010\)](#). Thus, there is seemingly no clear reason why a researcher in a typical situation of using multi-item scale data would choose this approach.

Further, Y21 states that his results hold only under models with no cross-loadings or correlated errors across the constructs (blocks of indicators). This assumption severely limits the usefulness of these results in empirical research. It is well established that correlated errors and cross-loadings are virtually inevitable in real-world applications of multi-item scale data ([Asparouhov et al., 2015](#); [Marsh et al., 2020](#); [Muthén and Asparouhov, 2012](#)). Thus, if a researcher insists on using a composite method, for computational simplicity perhaps, the obvious choice is not PLS, but GSCA(m), which has matrices to convey cross-block information at the indicator level, among other advantages. However, the advantage over simpler alternatives such as unit weights should be justified. This is

where the composite equivalence index (CEI) that we introduced in MoM becomes very useful. Indeed, there is no conceivable situation where PLS would be preferred, and as we noted earlier, even core PLS proponents such as Sarstedt and Ringle are moving toward advocating GSCA instead of PLS (Cho *et al.*, 2020, 2022).

Theme 2: partial least squares weights and reliability

From the above, Y21 also derives the implication that – parallel to the Bartlett factor scores – composites under PLS Mode B are the most reliable among all weighted averages of the observed indicators (Yuan and Deng, 2021). This claim is true only if factor scores are calculated one indicator block at a time, which Y21 does not mention. In more general conditions, regression factor scores outperform Bartlett scores in terms of reliability, but this comes with the cost of producing scores that are biased by other factors.

Further, PLS Mode B composites and Bartlett scores are only asymptotically equivalent for correctly specified models. In finite samples (i.e. real samples, not theoretical infinite ones), PLS Mode B may be better than some other composite scores, but all composite scores should perform *worse* than an ML-based Bartlett score. Indeed, the point of Bartlett scores is to produce scores that are not biased by other factors (univocality; see Harman, 1976, p. 387). Yet, PLS, regardless of whether Mode A or Mode B is used, weights the indicators based on their correlations with indicators of other factors, virtually guaranteeing that the scores are biased in small samples (Rönkkö, 2014; Rönkkö and Evermann, 2013; Rönkkö and Ylitalo, 2010). In this light, the point made here by Y21 is very weak with respect to applied uses of PLS.

Moreover, PLS Mode B models will almost invariably be misspecified because of the assumption that all cross-block indicator correlations (i.e. cross-loadings and error correlations) are channeled through the composite correlation, and PLS cannot handle a violation of that assumption. In common factor models, such inherent data features can be accommodated (Asparouhov *et al.*, 2015; Asparouhov and Muthén, 2021; Muthén and Asparouhov, 2012), which produces factor scores that are not biased by these misspecifications (although they will also contain measurement error). But, we again emphasize that the use of empirically derived indicator weights should be justified by:

- demonstrating that they differ meaningfully from unweighted scores (e.g. by the use of the CEI); and
- explaining why a specific set of weights makes sense considering the theory as we explain in MoM (see Figure 6 in MoM).

Taken together, these points result in the conclusion that one should *not use PLS* in most real-world situations, because it cannot handle the inherent features of real-world data (i.e. cross-loadings) and is highly susceptible to capitalization on chance even under correct causal specification.

Y21 speculates that “The criticism against the weights in the PLS-fallacy article might be because analytical results regarding the reliabilities of the composites under PLS-SEM were available only recently” (p. 6) This is incorrect for two reasons. First, Dijkstra (1981) proved already more than 40 years ago that asymptotically PLS Mode B produces the “most likely values” of latent variables, which is the same as maximizing reliability. Second, and more importantly as explained in MoM, there exists decades of research that demonstrate that practical advantages of differential indicator weighting are trivial even if ideal weights are known. Starting from Rönkkö and Ylitalo (2010), this has been demonstrated with PLS as well. The example by Y21 of using indicators with reliabilities of 0.16, 0.16 and 0.81 does not invalidate this. In practice, the poor indicators would be just thrown away as we explain in

MoM (e.g. p. 10) and as recommended as a best practice in the PLS literature: “Indicators with very low loadings (below 0.40) should, however, always be eliminated from the measurement model (Hair, Hult, Ringle, & Sarstedt, 2022)”. (Hair *et al.*, 2021, p. 77). This leaves us comparing a single indicator with reliability of 0.81 and a composite with reliability of 0.823. The difference is trivial and in practice when the PLS weights are not calculated from population values like Y21 does, but estimated from sample data, PLS composites rarely outperform simple summed scales (or the use of a single indicator in this case) as explained in MoM.

The key evidence in Y21 comes from Yuan *et al.* (2020). However, Yuan *et al.* (2020) do not provide any direct evidence of reliability differences between the indicators. What they show is that both PLS and summed scale estimates ($\hat{\beta}_{pls}$ and $\hat{\beta}_s$ in their Tables 3 and 5, respectively) are *all* negatively biased, but PLS estimates are *less* so. But this is not because of any reliability advantage of the method. Instead, this results from how PLS capitalizes on chance, creating a positive bias that to some extent cancels out the effect of measurement error, as has been documented in many articles as explained in MoM, a fact that Y21 for some reason ignore.

The positive bias is clearly evident in two features of the results. First, the positive bias due to capitalization on chance decreases with sample size, but measurement error bias does not and, as a consequence, we see an overall increase in negative bias as sample size increases, converging to the same level that summed scales have (Rönkkö, 2014). If PLS did indeed provide a reliability advantage, we should observe that PLS estimates were consistently *less* biased and the level of bias would *not* be affected by sample size. Second, the PLSc estimates ($\hat{\beta}_{plsc}^c$ and $\hat{\beta}_{plsc}^{dlc}$) are positively biased in small samples. This happens because the effect of random measurement error has been eliminated by correction for attenuation, but the capitalization on chance effect is not addressed, causing a positive bias, which too has been documented in the prior literature (Rönkkö *et al.*, 2016).

Y21 also claims that although PLS Mode A composites can be less reliable than equally weighted composites, but that those Mode A weights can be transformed to Mode B weights using a non-iterative method. Y21 further claims that these transformed weights enjoy the same statistical properties as Mode B weights. However, given the lack of situations where one would choose a PLS Mode B composite over a common factor-based method for obtaining scores, as detailed in the previous paragraph, there seems no real reason for readers to particularly care about this feature. Finally, Y21’s arguments here that weights sometimes make a difference, and sometimes do not, further support the use of the CEI to compare different composite methods.

Theme 3: bias, explained variance and “signal-to-noise”

Y21 makes some quite interesting points as regards the idea of comparing different estimators as to their “bias.” Unfortunately, the discussion in Y21 conflates two different issues, and in doing so makes some misleading points. First, we accept the point that without knowledge of “true” values, we cannot technically speak about “bias.” However, this does not appear to us to justify Y21’s blanket rejection of the entire notion of quantifying the bias of estimators against population parameters in SEM methods, simply because choosing scales for latent variables is necessary for the models to be identified. In fact, Y21’s argument readily extends to composites as well because regression coefficients depend on the composite weights, which are also chosen by the researcher. Taken to an extreme, the argument would also apply to physical measurements. For example, if we regress a person’s weight on their height, we get very different results depending on whether kilograms and centimeters or inches and pounds were used. We take the point that scaling choices and

metrics are often specific to a particular simulation, but it is not clear why this should invalidate the notion of within-study comparisons for example.

In practice, latent variable scales are not arbitrary; by constraining the first indicator's loading to 1, the latent variable inherits the scale of the first indicator (Little *et al.*, 2006). Nevertheless, we emphasize that researchers should carefully choose scaling methods and be aware of the impact of this decision for analysis and interpretation. Numerous sources are easily available in the literature already to help researchers understand the implications of scaling constraints (Gonzalez and Griffin, 2001; Klopp and Klößner, 2021; Klößner and Klopp, 2019; Schweizer and Troche, 2019; Steiger, 2002).

Second, after rejecting the idea of bias as a meaningful concept for evaluating the performance of SEM estimators, Y21 promotes the use of what is termed in Y21 the “signal to noise ratio,” which Y21 equates with “effect size.” Y21 further claims that, in a situation with two latent variables, PLS Mode B always yields a greater signal to noise ratio than ML-SEM for estimating the regression coefficient between the two latent variables (Yuan and Fang, 2021). This argument has two major problems. First, the signal-to-noise ratio discussed by Yuan and Fang (2021) is nothing more than a t -statistic, defined as the ratio of an estimate and its standard error. The t -statistic is not an effect size measure according to any common definition (Kelley and Preacher, 2012). By contrast, the t -statistic and the related p -value are measures of statistical significance. Indeed, the recent ASA guidelines on using and interpreting statistical significance clearly state “A p -value, or statistical significance, does not measure the size of an effect or the importance of a result” (Wasserstein and Lazar, 2016, p. 132).

Second, it seems to us that Y21's signal-to-noise analysis is incorrect in at least two different ways. The standard errors are taken from the OLS regression analysis that is applied to the PLS composites to obtain the path coefficient. The same error is present in Deng and Yuan (2023), which can be verified from their source code. As shown in Figure 4 of MoM, the variance of PLS estimates can be much greater than the variance of OLS estimates. This problem has been explained in prior research on PLS (Aguirre-Urreta and Rönkkö, 2018; Goodhue *et al.*, 2006) and is now textbook knowledge:

Parametric significance tests used in regression analyses cannot be applied to test whether coefficients such as outer weights, outer loadings, and path coefficients are significant. Instead, PLS-SEM relies on a nonparametric bootstrap procedure (Hair *et al.*, 2014, p. 130).

Perhaps even more problematically, as Schuberth *et al.* (2022) demonstrate, in a more realistic scenario with more than one predictor variable, the inconsistency of the PLS estimator can lead to incorrect results, failing to detect relationships that exist and detecting non-existent relationships, in contrast to ML-SEM that identifies the relationships correctly. As such, the claim by Y21 that an “inconsistent estimator can be more preferred if the purpose is to confirm a relationship between two constructs” (p. 3) is simply not true.

Still, the claims made in Y21 about signal to noise are interesting and worth some more discussion, lest they lead to the creation of yet another PLS myth. Specifically, to avoid issues of scaling, which we point out above, Y21's indicator of “effect size” is dimensionless. However, even leaving aside the problem that their effect size measure is in reality not an effect size measure, it is still not clear what we can learn from Y21's discussion. Specifically, whether or not the effect size indicator promoted in Y21 and Yuan and Fang (2021) is the most appropriate benchmark, and therefore that all prior simulations are meaningless (which we believe is a conclusion without solid grounding), Y21 tells us what we already know; in their own words: “Our empirical results indicate that *PLS-SEM tends to have an*

inflated effect size even with normally distributed data” (Y21 this issue, emphasis added). It is thus unclear why anyone would use PLS over another method.

Therefore, it is hard to reconcile any claim that PLS has some kind of “optimal signal-to-noise ratio,” with the claim that it *also* has “inflated effect sizes” and/or “inflated type I errors and R-square values.” It appears that Y21 anticipates this objection, as they state that “maximization of R^2 and capitalization on chance cannot be separated.” This statement may of course be true, but that is the case for any optimization problem, so the statement is disingenuous at best. In fact, some methods are more robust and less susceptible to this issue than others, and many are more robust than PLS (e.g. unit-weighted composites and sum scores, where the weights are fixed and thus immune to sampling fluctuations). Again, we return to the only logical conclusion; there is very little use for PLS in applied research. In [Table 2](#), we provide a list of the main points made in Y21 (of relevance to MoM) and how they provide yet more evidence against the use of PLS.

Moreover, when discussing capitalization on chance, it is also impossible to avoid noting that the high prevalence of statistically significant results in the PLS literature is likely also due to the bootstrap sign-corrections implemented in some PLS software. Briefly, these procedures selectively flip the signs of the outputs (e.g. indicator weights and regression coefficients) within the bootstrap resamples to maintain consistency with the signs obtained from analyzing the entire, original data set. However, as demonstrated by [Rönkkö et al. \(2015\)](#), this “trick” (as it is best described) will lead to drastically inflated false-positive rates. In fact, with the individual sign-change correction that makes each and every bootstrap quantity has a sign that is consistent with the original estimate, one achieves a 100% false-positive rate! We are aware of only one publication stating that this approach “should be considered as deprecated” ([Henseler et al., 2016](#), p. 15, Note 3), but have no indication of how many applied PLS articles have heeded the warning. Therefore, there could be many contaminated results, hence many incorrect conclusions and recommendations, in the literature.

Discussion and conclusions

In our paper MoM, we brought together well over a decade’s worth of critique of PLS, using very simple examples, easily reproducible by anyone using software to run PLS or any other statistical analysis. We did not aim to introduce any new points to the discourse, strong evidence for the problems of PLS abounds, but it is available mainly in methodological journals and is therefore perhaps inaccessible to many applied marketing and management researchers. The existing critiques of PLS already provide more than enough evidence to conclude that PLS as it is used in marketing and related disciplines offers no meaningful advantages over ML SEM or GSCA(m); however, it has enough serious disadvantages that it should be avoided as a general rule. We cannot envisage a single realistic marketing research situation where PLS would be the preferred analytic method on any criterion other than convenience – although in such a case, one may as well use the *most* convenient option: sum scores and OLS regression [2]. Worse, as we have pointed out in MoM and here, PLS is currently used in marketing and related fields in such a way that it is harmful to scientific progress.

Our intention with MoM was to reach as broad an audience as possible in a clear way and to reiterate that PLS is not a viable analysis method for typical marketing and related field research problems. Obviously, we expected to see some pushback. While we were pleased to engage with Y21, which we think have helped clarify a number of important issues, we were disappointed not to see engagement with our core points by any other commenters. Of course, there would have been a very simple way to effectively respond to

Claim in Y21	Conclusion
Indicator weights from PLS Mode B are equivalent to Bartlett factor scores, but only when assuming no cross-loadings or correlated errors across scales, which ML-SEM and Bayesian SEM can accommodate	In real-world analysis situations, we cannot assume cross-loadings or correlated errors are non-existent <i>Therefore:</i> PLS should not be used in situation where cross-loadings or correlated errors may exist, as it has no way to account for these features Even if a model was correct, in real-world finite samples, advantages of differential item weighting are trivial as long as very bad items are first dropped from the data
PLS Mode B composites are most reliable among all weighted averages of observed indicators, for correctly specified models and where all cross-loadings and error correlations are completely channeled through the composite correlation(s)	<i>Therefore:</i> Differentially weighted composites should be always compared against unit weighted ones using the CEI. Unless meaningful differences are found and can be explained, unit-weighted composites should be chosen for their simplicity Y21 confuses effect size with statistical significance. While PLS may lead to higher statistical significance, and thus a greater likelihood of finding an effect, this comes at the expense of a higher chance of false positives. The claim by Y21 does not hold in more realistic models with more than one predictor variable, where the inconsistency of PLS can lead to incorrect conclusions about the existence of an effect <i>Therefore:</i> Results from a PLS analysis are more likely to be false positives than those from other methods such as ML-SEM, <i>ceteris paribus</i>
In a situation with two latent variables, PLS Mode B always yields a greater signal to noise ratio than ML-SEM for estimating the regression coefficient between the two latent variables	PLS should especially not be used with small samples or low quality data
<i>However:</i> "PLS-SEM tends to have an inflated effect size even with normally distributed data ..." "PLS-SEM may have inflated type I errors and R-square values even with normally distributed data" "[PLS] needs a large enough sample size and good quality of data for reliable model/parameter inference (Marcoulides and Saunders, 2006). In particular, samples with heavy-tails or data contamination can strongly affect the goodness of the estimates by the LS method"	

Table 2.
Evidence *against*
PLS provided by Y21

Source: Authors thinking specifically for this paper

our main concerns; simply present an empirical data set where PLS weights make a substantial difference to analysis results, and provide a clear explanation of why the PLS weights make more sense than unit weights in the particular context. That no defenders of PLS were willing or able to do so provides a strong indication that situations where PLS makes an explainable beneficial difference in real-world analysis situations, are either extremely rare or non-existent.

Indeed, as we show above, in attempting to defend PLS, Y21 actually provides *additional* evidence of its lack of suitability for handling typical, real-world multi-item scale data features such as cross-loadings or correlated errors. Y21 also states that relationships

estimated in PLS tend to have inflated effect sizes, R^2 s and Type 1 errors (all obviously interconnected), even with normally distributed data. We are also grateful that Y21 may help in finally debunking the strangely persistent myth that PLS is suited to small samples. This myth continues even in the face of large amounts of evidence to the contrary, and no actual evidence to support it. Alarming, even the J-B Steenkamp Award judging committee for the 2021 *International Journal of Research in Marketing* states that the simulations in Reinartz *et al.* (2009) winning paper comparing SEM and PLS “show that PLS can be a good methodological choice if sample size is low” (IJRM, 2021, p. A3).

It is surely now obvious, if it was not already, that applied users should just adopt better analysis methods, of which plenty are easily available. For example, if researchers make an informed (see MoM and this rejoinder) choice to use composites, GSCA(m) with fixed indicator weights would solve all three of the fatal problems of PLS that we point out:

- (1) capitalization on chance (by using sumscores rather than weighted composites);
- (2) measurement error (through the inclusion of the uniqueness terms); and
- (3) non-zero cross-loadings and error correlations (via the additional parameter matrices).

Even using disattenuated regression with sum scores can only solve (a) and (b), while PLSc only solves (b). Again, PLS loses hands-down to readily available alternatives (Choi and Hwang, 2020; Hwang *et al.*, 2017, 2021).

We would also like to make the important point that on too many occasions, authors appear to choose an analysis method using a small set of criteria:

- how many papers in the journals they read are advocating the method and using it;
- how important or well-known are the advocates versus the critics; and
- how powerful is the rhetorical argumentation for and against the method.

None of these criteria is meaningful when it comes to methodological choices. Rather than rely on precedence for using a method in applied management journals, or second-hand advocacy and applied papers, researchers should consult *methodological* journals as well, to understand more thoroughly any method they wish to use. If a method has clear evidence pointing against its use, the onus is on the researcher (and the reviewer) to understand the limits of that evidence, and not to take counter-claims by obviously partial advocates at face value.

In MoM, we provided a simple tool that can be used to justify the use of PLS or any other complex method for constructing composites: the CEI. The CEI is a simple, method-agnostic tool that researchers can use to provide direct comparisons between different composite methods (see Figure 6 of MoM); in brief, it assesses the degree of correlation between composites constructed by various techniques. It is not intended to privilege one method over another, but simply to show where competing methods of constructing composites make a difference, and where they do not. In cases where the CEI shows that there is no substantive difference between different ways of constructing a composite (i.e. the inter-composite correlations are very close to unity), the simplest method (unit-weighted composites) should be preferred. However, when the CEI shows there *are* differences, the onus is on the researcher to explain why the method they wish to use is beneficial. For example, researchers could support their choices through careful consideration of the simulation-based evidence of different composite methods or by explaining why the specific weights for the indicators make sense considering the indicator wording and the underlying theory.

Every one of the claims we make in MoM is well established in the methodological literature. In fact, it is increasingly obvious that the edifice of PLS is fracturing, and even those who in the past have strongly advocated for PLS, are now either explicitly advocating against it, or instead advancing other methods as being more capable. For example, we cited the recent work of Sarstedt and colleagues as recommending the use of GSCA for modeling composites (Cho *et al.*, 2022). Further, Jörg Henseler (who has been in the past among the most vocal advocates of PLS) has recently stated explicitly that PLS is not suitable for models based on reflectively measured variables (Henseler and Schubert, this issue), and argued against PLS, and the content of much of its advocacy literature, even more strongly in his recent book (Henseler, 2021, p. 96).

In conclusion, surely it is an absolute minimum standard of scientific integrity that we understand the tools we use to draw conclusions about the world we are studying. It is obvious that when considering the use of PLS, the best-case scenario is that researchers are either unaware of or do not understand the clear methodological evidence pointing against its use. The objective of MoM was to remedy this situation by showing simple examples that anyone can replicate with their own data so that no longer can marketing researchers reasonably claim either lack of awareness or understanding. The worst-case scenario is that PLS continues to be used and promoted by advocates despite the methodological flaws clearly demonstrated here and in prior works. Sadly, we can offer no remedy for that.

Notes

1. To be more precise, we presented this as a choice between latent variables and scale scores that can be either linear composites or non-linear functions of the observed data. But, because scale scores are nearly always calculated as linear composites, in practice, the choice is between latent variables and composites.
2. Interestingly, in recent personal conversations, one of the present authors heard stories about occasions where researchers have tried to use OLS and sumscores, and reviewers have pushed back with the criticism that such methods are “too simple,” and thus, the authors should use PLS. Such a situation is almost akin to a “simplicity tax” on research. Of course, it should not need saying that simpler methods should actually be *preferred* when more complex results can offer no meaningful advantage. We hope that the material in MoM and this rejoinder provides enough material for researchers (and editors) to rebut such ill-informed criticism, and perhaps to convince reviewers to stop making it.

References

- Aguirre-Urreta, M.I. and Rönkkö, M. (2018), “Statistical inference with PLS using bootstrap confidence intervals”, *MIS Quarterly*, Vol. 42 No. 3, pp. 1001-1020.
- Asparouhov, T. and Muthén, B. (2021), “Advances in Bayesian model fit evaluation for structural equation models”, *Structural Equation Modeling: A Multidisciplinary Journal*, Vol. 28 No. 1, pp. 1-14.
- Asparouhov, T., Muthén, B. and Morin, A.J.S. (2015), “Bayesian structural equation modeling with cross-loadings and residual covariances: comments on Stromeyer et al”, *Journal of Management*, Vol. 41 No. 6, pp. 1561-1577.
- Casella, G. and Berger, R.L. (2001), *Statistical Inference*, 2nd ed., Duxbury, Pacific Grove, CA.
- Casella, G. and Berger, R.L. (2002), *Statistical Inference*, 2nd ed., Thomson Learning, Pacific Grove, CA.
- Cho, G., Sarstedt, M. and Hwang, H. (2022), “A comparative evaluation of factor- and component-based structural equation modelling approaches under (in)correct construct representations”, *British Journal of Mathematical and Statistical Psychology*, Vol. 75 No. 2, pp. 220-251.

-
- Cho, G., Hwang, H., Sarstedt, M. and Ringle, C.M. (2020), "Cutoff criteria for overall model fit indexes in generalized structured component analysis", *Journal of Marketing Analytics*, Vol. 8 No. 4, pp. 189-202.
- Choi, J.Y. and Hwang, H. (2020), "Bayesian generalized structured component analysis", *British Journal of Mathematical and Statistical Psychology*, Vol. 73 No. 2, pp. 347-373.
- Deng, L. and Yuan, K.-H. (2023), "Which method is more powerful in testing the relationship of theoretical constructs? A meta comparison of structural equation modeling and path analysis with weighted composites", *Behavior Research Methods*, Vol. 55 No. 3, pp. 1460-1479, doi: [10.3758/s13428-022-01838-z](https://doi.org/10.3758/s13428-022-01838-z).
- Dijkstra, T.K. (1981), "Latent variables in linear stochastic models: reflections on 'maximum likelihood' and 'partial least squares' methods", Doctoral dissertation, Rijksuniversiteit te Groningen.
- Dijkstra, T.K. (2017), "A perfect match between a model and a mode", in Latan, H. and Noonan, R. (Eds), *Partial Least Squares Path Modeling: Basic Concepts, Methodological Issues and Applications*, Springer International Publishing, pp. 55-80, doi: [10.1007/978-3-319-64069-3_4](https://doi.org/10.1007/978-3-319-64069-3_4).
- Evermann, J. and Tate, M. (2010), "Testing models or fitting models? Identifying model misspecification in PLS", *ICIS 2010 Proceedings*, available at: http://aisel.aisnet.org/icis2010_submissions/21
- Gonzalez, R. and Griffin, D. (2001), "Testing parameters in structural equation modeling: every 'one' matters", *Psychological Methods*, Vol. 6 No. 3, pp. 258-269.
- Goodhue, D.L., Lewis, W. and Thompson, R. (2006), "PLS, small sample size, and statistical power in MIS research", in Sprague, R. Jr. (Ed.), *Proceedings of the 39th HI International Conference on System Sciences*, IEEE Computer Society Press.
- Goodhue, D.L., Lewis, W. and Thompson, R. (2015), "PLS pluses and minuses in path estimation accuracy", *AMCIS 2015 Proceedings*, available at: <http://aisel.aisnet.org/amcis2015/ISPhil/GeneralPresentations/3>
- Hair, J.F. and Sarstedt, M. (2019), "Factors versus composites: guidelines for choosing the right structural equation modeling method", *Project Management Journal*, Vol. 50 No. 6, pp. 619-624, doi: [10.1177/8756972819882132](https://doi.org/10.1177/8756972819882132).
- Hair, J.F., Hult, G.T.M., Ringle, C.M. and Sarstedt, M. (2014), *A Primer on Partial Least Squares Structural Equations Modeling (PLS-SEM)*, SAGE Publications, Los Angeles, CA.
- Hair, J., Hult, G.T.M., Ringle, C.M. and Sarstedt, M. (2022), *A Primer on Partial Least Squares Structural Equation Modelling*, 3rd ed., SAGE Publications, Los Angeles, CA.
- Hair, J.F., Hult, G.T.M., Ringle, C.M., Sarstedt, M., Danks, N.P. and Ray, S. (2021), *Partial Least Squares Structural Equation Modeling (PLS-SEM) Using R: A Workbook*, Springer International Publishing, doi: [10.1007/978-3-030-80519-7](https://doi.org/10.1007/978-3-030-80519-7).
- Harman, H.H. (1976), *Modern Factor Analysis*, University of Chicago Press, Chicago, IL.
- Henseler, J. (2021), *Composite-Based Structural Equation Modeling: Analyzing Latent and Emergent Variables*, The Guilford Press, New York, NY.
- Henseler, J., Dijkstra, T.K., Sarstedt, M., Ringle, C.M., Diamantopoulos, A., Straub, D.W., Ketchen, D.J., Hair, J.F., Hult, G.T.M. and Calantone, R.J. (2014), "Common beliefs and reality about PLS: comments on Rönkkö and Evermann (2013)", *Organizational Research Methods*, Vol. 17 No. 2, pp. 182-209, doi: [10.1177/1094428114526928](https://doi.org/10.1177/1094428114526928).
- Henseler, J., Hubona, G. and Ray, P.A. (2016), "Using PLS path modeling in new technology research: updated guidelines", *Industrial Management and Data Systems*, Vol. 116 No. 1, pp. 2-20.
- Hwang, H., Takane, Y. and Jung, K. (2017), "Generalized structured component analysis with uniqueness terms for accommodating measurement error", *Frontiers in Psychology*, Vol. 8, p. 2137.
- Hwang, H., Cho, G., Jung, K., Falk, C.F., Flake, J.K., Jin, M.J. and Lee, S.H. (2021), "An approach to structural equation modeling with both factors and components: integrated generalized structured component analysis", *Psychological Methods*, Vol. 26 No. 3, pp. 273-294.

-
- IJRM (2021), "Announcement: winner of 2021 Jan-benedict Steenkamp award for long-term impact", *International Journal of Research in Marketing*, Vol. 38 No. 2, pp. A2-A3, doi: [10.1016/j.ijresmar.2021.06.002](https://doi.org/10.1016/j.ijresmar.2021.06.002).
- Kelley, K. and Preacher, K.J. (2012), "On effect size", *Psychological Methods*, Vol. 17 No. 2, pp. 137-152.
- Klopp, E. and Klößner, S. (2021), "The impact of scaling methods on the properties and interpretation of parameter estimates in structural equation models with latent variables", *Structural Equation Modeling: A Multidisciplinary Journal*, Vol. 28 No. 2, pp. 182-206.
- Klößner, S. and Klopp, E. (2019), "Explaining constraint interaction: how to interpret estimated model parameters under alternative scaling methods", *Structural Equation Modeling: A Multidisciplinary Journal*, Vol. 26 No. 1, pp. 143-155.
- Little, T.D., Slegers, D.W. and Card, N.A. (2006), "A non-arbitrary method of identifying and scaling latent variables in SEM and MACS models", *Structural Equation Modeling: A Multidisciplinary Journal*, Vol. 13 No. 1, pp. 59-72.
- McIntosh, C.N., Edwards, J.R. and Antonakis, J. (2014), "Reflections on partial least squares path modelling", *Organizational Research Methods*, Vol. 17 No. 2, pp. 210-251, doi: [10.1177/1094428114529165](https://doi.org/10.1177/1094428114529165).
- Markus, K.A. and Borsboom, D. (2013), *Frontiers in Test Validity Theory: Measurement, Causation and Meaning*, Psychology Press.
- Marsh, H.W., Guo, J., Dicke, T., Parker, P.D. and Craven, R.G. (2020), "Confirmatory factor analysis (CFA), exploratory structural equation modeling (ESEM), and Set-ESEM: optimal balance between goodness of fit and parsimony", *Multivariate Behavioral Research*, Vol. 55 No. 1, pp. 102-119.
- Muthén, B. and Asparouhov, T. (2012), "Bayesian structural equation modeling: a more flexible representation of substantive theory", *Psychological Methods*, Vol. 17 No. 3, pp. 313-335.
- Reinartz, W.J., Haenlein, M. and Henseler, J. (2009), "An empirical comparison of the efficacy of covariance-based and variance-based SEM", *International Journal of Research in Marketing*, Vol. 26 No. 4, pp. 332-344.
- Rigdon, E.E. (2016), "Choosing PLS path modeling as analytical method in European management research: a realist perspective", *European Management Journal*, Vol. 34 No. 6, pp. 598-605, doi: [10.1016/j.emj.2016.05.006](https://doi.org/10.1016/j.emj.2016.05.006).
- Rönkkö, M. (2014), "The effects of chance correlations on partial least squares path modeling", *Organizational Research Methods*, Vol. 17 No. 2, pp. 164-181.
- Rönkkö, M. (2020), "Introduction to matrixpls", available at: <https://cran.r-project.org/web/packages/matrixpls/vignettes/matrixpls-intro.pdf>
- Rönkkö, M. and Cho, E. (2022), "An updated guideline for assessing discriminant validity", *Organizational Research Methods*, Vol. 25 No. 1, pp. 6-14, doi: [10.1177/1094428120968614](https://doi.org/10.1177/1094428120968614).
- Rönkkö, M. and Evermann, J. (2013), "A critical examination of common beliefs about partial least squares path modeling", *Organizational Research Methods*, Vol. 16 No. 3, pp. 425-448.
- Rönkkö, M. and Ylitalo, J. (2010), "Construct validity in partial least squares path modeling", *ICIS 2010 Proceedings*, available at: http://aisel.aisnet.org/icis2010_submissions/155
- Rönkkö, M., McIntosh, C.N. and Aguirre-Urreta, M.I. (2016), "Improvements to PLS: remaining problems and simple solutions", *Unpublished Working Paper*, available at: <http://urn.fi/URN:NBN:fi:aalto-201603051463>
- Rönkkö, M., McIntosh, C.N. and Antonakis, J. (2015), "On the adoption of partial least squares in psychological research: caveat emptor", *Personality and Individual Differences*, Vol. 87, pp. 76-84.
- Sarstedt, M., Hair, J.F., Ringle, C.M., Thiele, K.O. and Gudergan, S.P. (2016), "Estimation issues with PLS and CBSEM: where the bias lies!", *Journal of Business Research*, Vol. 69 No. 10, pp. 3998-4010, doi: [10.1016/j.jbusres.2016.06.007](https://doi.org/10.1016/j.jbusres.2016.06.007).

-
- Schuberth, F. (2021), "Confirmatory composite analysis using partial least squares: setting the record straight", *Review of Managerial Science*, Vol. 15 No. 5, pp. 1311-1345, doi: [10.1007/s11846-020-00405-0](https://doi.org/10.1007/s11846-020-00405-0).
- Schuberth, F., Rosseel, Y., Rönkkö, M., Trinchera, L. and Henseler, J. (2022), "Relationships between latent variables are neither arbitrary nor equivalent to relationships between proxies: a comment on Yuan and Deng (2021)", *Working Paper*.
- Schweizer, K. and Troche, S. (2019), "The EV scaling method for variances of latent variables", *Methodology*, Vol. 15 No. 4, pp. 175-184.
- Steiger, J.H. (2002), "When constraints interact: a caution about reference variables, identification constraints, and scale dependencies in structural equation modeling", *Psychological Methods*, Vol. 7 No. 2, pp. 210-227.
- Voorhees, C.M., Brady, M.K., Calantone, R. and Ramirez, E. (2016), "Discriminant validity testing in marketing: an analysis, causes for concern, and proposed remedies", *Journal of the Academy of Marketing Science*, Vol. 44 No. 1, pp. 119-134, doi: [10.1007/s11747-015-0455-4](https://doi.org/10.1007/s11747-015-0455-4).
- Wasserstein, R.L. and Lazar, N.A. (2016), "The ASA's statement on p-values: context, process, and purpose", *The American Statistician*, Vol. 70 No. 2, pp. 129-133.
- Yuan, K.H. and Deng, L. (2021), "Equivalence of partial-least-squares SEM and the methods of factor-score regression", *Structural Equation Modeling: A Multidisciplinary Journal*, Vol. 28 No. 4, pp. 557-571.
- Yuan, K.H. and Fang, Y. (2021), "Which method delivers greater signal-to-noise ratio: structural equation modeling or regression analysis with composite scores?", *Under Review*.
- Yuan, K.H., Wen, Y. and Tang, J. (2020), "Regression analysis with latent variables by partial least squares and four other composite scores: consistency, bias and correction", *Structural Equation Modeling: A Multidisciplinary Journal*, Vol. 27 No. 3, pp. 333-350.

Further reading

- Rönkkö, M., Lee, N., Evermann, J., McIntosh, C.N. and Antonakis, J. (2022), "Marketing or methodology? Exposing the fallacies of PLS with simple demonstrations", *European Journal of Marketing*.

Corresponding author

Mikko Rönkkö can be contacted at: mikko.ronkko@jyu.fi

For instructions on how to order reprints of this article, please visit our website:

www.emeraldgrouppublishing.com/licensing/reprints.htm

Or contact us for further details: permissions@emeraldinsight.com